The background of the image is a detailed aerial satellite photograph of a forested landscape. A large, winding river or stream cuts through the center of the image, its path marked by a lighter green color. To the left of the river, there is a small, developed area with a grid-like street pattern, likely a town or city. The surrounding terrain is dominated by dense green forests, with various shades of green indicating different types of vegetation or terrain. The overall image has a high resolution and captures the intricate details of the natural environment and human-made structures.

# THE SAR HANDBOOK

---

Comprehensive Methodologies for Forest  
Monitoring and Biomass Estimation

First edition. Published electronically April 2019.

DOI: 10.25966/nr2c-s697

This publication may be reproduced in whole or in part and in any form for educational or non-profit purposes without special permission from the copyright holder, provided acknowledgement of the source is made. No use of this publication may be made for resale or for any other commercial purpose.

**DISCLAIMER:**

The views expressed in this publication are not necessarily those of the agencies cooperating in this project. Mention of a commercial company or product in this report does not imply endorsement by SERVIR, SilvaCarbon, NASA, or USAID. The use of information from this publication concerning proprietary products for advertising or publicity is not permitted. Trademark names and symbols are used in an editorial fashion with no intention of infringement on trademark or copyright laws.

Figures and maps in this handbook contain material from: Google Maps data: Imagery © 2018 Landsat/Copernicus, DigitalGlobe, Mapdata © 2018 Google. Sentinel-1 data: Earth Big Data, LLC 2018, contains modified Copernicus Sentinel data 2014-2018, processed by ESA. ALOS data: Earth Big Data, LLC 2018; includes Material © JAXA/METI 2007-2018

SERVIR Global Science Coordination Office  
National Space Science and Technology Center  
320 Sparkman Drive, Huntsville, AL 35805

<http://www.SERVIRglobal.net>



**COVER IMAGE:** The Mekong Delta in southern Vietnam, as seen by Sentinel-1 G-band SAR on November 26, 2018. (Jeremy Nicoll, Alaska Satellite Facility, contains modified Copernicus Sentinel data 2018, processed by ESA).

**SERVIR GLOBAL**

**SilvaCarbon**



University of  
Massachusetts  
Amherst



EARTH BIG DATA  
Where Solutions Begin.

THE UNIVERSITY OF  
ALABAMA IN HUNTSVILLE



CILSS  
RCMRD

adpc

ICIMOD

Africa Ixmucane Flores-Anderson · Kelsey E. Herndon · Rajesh Bahadur Thapa · Emil Cherrington  
Editors

# The Synthetic Aperture Radar (SAR) Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation

Preface by Juliann Aukema & Sylvia Wilson  
Foreword by Daniel Irwin

# CONTENTS

<b>Acknowledgements .....</b>	5
<b>Preface .....</b>	8
<b>Foreword .....</b>	10
<b>Chapter 1</b>	
Introduction and Rationale .....	13
<b>Chapter 2</b>	
Spaceborne Synthetic Aperture Radar: Principles, Data Access, and Basic Processing Techniques .....	21
Appendix A: SAR Image Processing Routines - Training Module .....	44
<b>Chapter 3</b>	
Using SAR Data for Mapping Deforestation and Forest Degradation .....	65
APPENDIX B: Getting Started Using SAR for Forest Monitoring - Training Module .....	80
<b>Chapter 4</b>	
Forest Stand Height Estimation .....	173
APPENDIX C: Estimating Forest Stand Height Using L-band SAR - Training Module .....	186
<b>Chapter 5</b>	
SAR Methods for Mapping and Monitoring Forest Biomass .....	207
APPENDIX D: Mapping Forest Biomass with Radar Remote Sensing - Training Module .....	247
<b>Chapter 6</b>	
Radar Remote Sensing of Mangrove Forests .....	255
APPENDIX E: Mapping and Monitoring Mangrove Forests with Radar Remote Sensing - Training Module .....	266
<b>Chapter 7</b>	
Sampling Designs for SAR-Assisted Forest Biomass Surveys .....	281
APPENDIX F: Sampling Design for Forest Biomass Surveys - Training Module .....	290
<b>Chapter 8</b>	
Perspectives on the Future Application of SAR in Forest and Environmental Monitoring .....	297
<b>About the Editors .....</b>	306

# ACKNOWLEDGEMENTS

The development and completion of this Handbook would not have been possible without the direct support and collaboration of the SERVIR-Global family, including SERVIR Hubs from West Africa, Eastern & Southern Africa, Hindu Kush Himalaya, and Lower Mekong, the team at the SERVIR Science Coordination Office, NASA scientists and the world-renowned SAR experts authoring this book.

The editors of this SAR Handbook immensely appreciate the interagency support received, which made possible to conduct needs assessments, gather key attendees at workshops and training events, and develop relevant, hands-on training tutorials.

Below, we would like to recognize not only the authors of each of the chapters, but also all the team members involved in reviewing, editing and directing content for the theory-focused and respective training tutorial sections. This collaborative process made it possible to generate applied content that we expect to be a game-changer for the remote sensing community wanting to use SAR data for forest monitoring and biomass estimation.

## Authors, collaborating authors and reviewers:

### Chapter 1

Authors: Africa Ixmucane Flores-Anderson, Kelsey E. Herndon, Emil Cherrington, Rajesh Bahadur Thapa, Leah Kucera, Nguyen Hanh Quyen, Phoebe Oduor, Anastasia Wahome, Karis Tenneson, Bako Mamane, David Saah, Farrukh Chishtie, Ashutosh Limaye

### Chapter 2

Author: Franz Meyer  
Reviewers: Africa I. Flores-Anderson, Emil Cherrington, Rajesh Bahadur Thapa, Kelsey E. Herndon, Faisal Mueen Qamer, Stella Ofori-Ampofo, Helen Baldwin  
Additional figure editing: Leah Kucera

### Chapter 3

Author: Josef Kellndorfer  
Reviewers: Kelsey E. Herndon, Bhoj Raj Ghimire, Rebekke Muench, Bako Mamane, Africa I. Flores-Anderson, Emil Cherrington, Rajesh Bahadur Thapa  
Additional figure editing: Leah Kucera

### Chapter 4

Author: Paul Siqueira  
Reviewers: Batuhan Osmanoglu, Nguyen Hanh Quyen, Amanda Markert, Africa I. Flores-Anderson, Emil Cherrington, Rajesh Bahadur Thapa, Kelsey E. Herndon  
Additional figure editing: Leah Kucera

### Chapter 5

Author: Sasan Saatchi  
Reviewers: Rajesh Bahadur Thapa, Emil Cherrington, Kelsey E. Herndon, Walter Lee Ellenburg, Sunil Thapa, Africa I. Flores-Anderson  
Additional figure editing: Leah Kucera

### Chapter 6

Author: Marc Simard  
Reviewers: Emil Cherrington, Kelsey E. Herndon, Kenneth Mubea, Edward Ouko, Africa I. Flores-Anderson, Rajesh Bahadur Thapa  
Additional figure editing: Leah Kucera

### Chapter 7

Author: Hans Andersen  
Reviewers: Emily C. Adams, Robinson Mugo, Africa I. Flores-Anderson, Emil Cherrington, Rajesh Bahadur Thapa, Kelsey E. Herndon  
Additional figure editing: Leah Kucera

### Chapter 8

Authors: Emil Cherrington, Africa I. Flores-Anderson, Rajesh Bahadur Thapa, Kelsey E. Herndon, Anastasia Wahome, Phoebe Oduor, Kenneth Mubea, Edward Ouko, Nguyen Hanh Quyen, David Saah, Bako Mamane, Paul Bartel, Foster Mensah, Stella Ofori-Ampofo, Eric Anderson, Emily C. Adams, Rebekke Muench, Amanda Markert, Andrea Nicolau, Walter Lee Ellenburg, Betzy Hernandez Sandoval, Ashutosh Limaye, Robert Griffin, Daniel Irwin  
Additional figure editing: Leah Kucera

## Graphic design and editing:

Copy editing: Troy Farsoun, Jared Austin, Kay Glover  
Art direction and layout: Leah Kucera

# ACKNOWLEDGEMENTS

## Meetings & workshops:

Multiple meetings and international workshops were carried out as part of this effort. The initial scoping meeting helped to identify the topics covered in the Handbook, and we would like to acknowledge and thank all the participants that helped to define the content for trainings and Handbook chapters:

### *Scoping Meeting (February 21-22, 2017, Huntsville, AL)*

Rajesh Bahadur Thapa (ICIMOD, Nepal), Kabir Uddin (ICIMOD, Nepal), Anastasia Wahome (RCMRD, Kenya), Phoebe Odour (RCMRD, Kenya), Bakor Mamane (AGRHYMET, Niger), Kadidia Yero (AGRHYMET, Niger), Paul Bartel (AGRHYMET, Niger), Nguyen Hanh Quyen (ADPC, Thailand), David Saah (Spatial Informatics Group, USA), Sassan Saatchi (NASA JPL, USA), Alessandro Baccini (Woods Hole Research Center, USA), Bruce Chapman (NASA JPL, USA), Paul Siqueira (University of Massachusetts, Amherst, USA), Josef Kellndorfer (Earth Big Data, USA), Hans Andersen (USFS, USA), Sasha Gottlieb (USFS, USA), Sylvia Wilson (USGS, USA), Johan Pontus Olofsson (Boston University, USA), Gustavo Galindo (IDEAM, Colombia), Evan Notman (USAID, USA), Dan Irwin (SERVIR SCO), Ashutosh Limaye (SERVIR SCO), Eric Anderson (SERVIR SCO), Krishna Vadrevu (SERVIR SCO), Raymond French (SERVIR SCO), Emil Cherrington (SERVIR SCO), Lee Ellenburg (SERVIR SCO), Francisco Delgado (SERVIR SCO), Billy Ashmall (SERVIR SCO), Lance Gilliland (SERVIR SCO), Bill Crosson (SERVIR SCO), Emily Adams (SERVIR SCO), Rebekke Muench (SERVIR SCO), Kelsey Herndon (SERVIR SCO)

Organizers: Africa Flores, Emil Cherrington, Krishna Vadrevu, Paula Link, Emily Adams

We would also like to acknowledge all the SAR experts that provided the international trainings, the SERVIR-Hub counterparts that led coordination of these training events, the colleagues that led follow-up workshops at their home institutions, as well as the colleagues that tested scripts, reviewed, and/or developed the Handbook tutorials.

### *Workshop: SAR basics and forest degradation; Deforestation detection (Regions: West Africa, Hindu Kush Himalaya)*

Subject Matter Experts: Franz Meyer, Josef Kellndorfer  
SERVIR Hub POCS: Bakor Mamane, Rajesh Bahadur Thapa  
Follow-on workshop trainers: Rebekke Muench, Begum Rushi

Training Tutorial Reviewers: Kelsey E. Herndon (Meyer)  
Rebekke Muench, Andrea Nicolau (Kellndorfer)

### *Workshop: Forest stand height (Region: Mekong)*

Subject Matter Expert: Paul Siqueira  
SERVIR Hub POC: Nguyen Hanh Quyen  
Follow-on workshop trainer: Krishna Vadrevu  
Training Tutorial Reviewers: Helen Baldwin, Sarva Pulla, Yang Lei

### *Workshop: Mangrove and sampling design (Region: Eastern & Southern Africa)*

Subject Matter Experts: Marc Simard, Hans Andersen  
SERVIR Hub POC: Phoebe Odour  
Follow-on workshop trainers: Emily C. Adams, Kenneth Mubea  
Training Tutorial Reviewers: Africa Flores-Anderson (Anderson), Katherine Stratton (Simard)

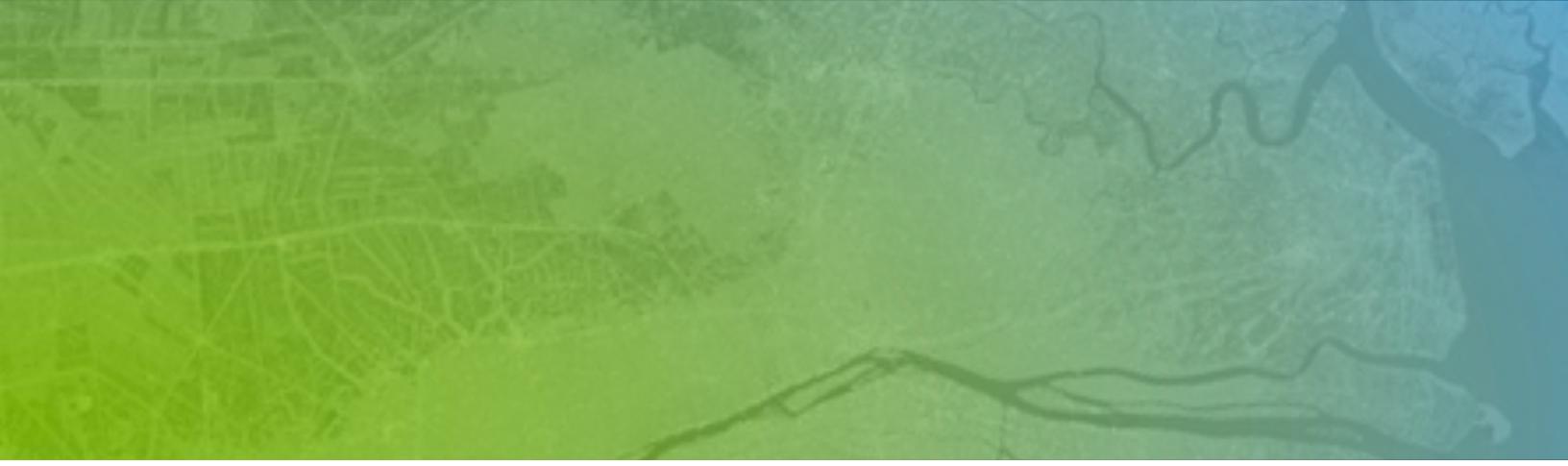
### *Workshop: Biomass (Region: Hindu Kush Himalaya)*

Subject Matter Expert: Sassan Saatchi  
SERVIR Hub POC: Rajesh Bahadur Thapa  
Follow-on workshop trainer: Emil Cherrington  
Training Tutorial Reviewer: Rajesh Bahadur Thapa

A final close-out meeting took place to bring the SAR experts, scientists from SERVIR-Hubs and relevant agencies together to address the main deliverables and provide additional content for chapters and trainings. We would like to acknowledge the meeting participants for their key contributions:

### *Close-Out Meeting (November 5-6, 2018, Huntsville, AL)*

Franz Meyer (ASF, USA), Josef Kellndorfer (Earth Big Data, USA), Paul Siqueira (University of Massachusetts, Amherst, USA), Sasan Saatchi (NASA JPL, USA), Marc Simard (NASA JPL, USA), Hans Andersen (USFS, USA), Rajesh Bahadur Thapa (ICIMOD, Nepal), Kenneth Mubea (RCMRD, Kenya), Anastasia Wahome (RCMRD, Kenya), Sasha Gottlieb (USFS, USA), Gustavo Galindo (IDEAM, Colombia), Eric Anderson (SERVIR SCO), Kelsey Herndon (SERVIR SCO, USA), Africa Flores-Anderson (SERVIR SCO, USA), Leah Kucera (SERVIR SCO, USA), Emil Cherrington (SERVIR SCO, USA), Emily Adams



(SERVIR SCO, USA), Rebekke Muench (SERVIR SCO, USA), Ashutosh Limaye (SERVIR SCO, USA), Dan Irwin (SERVIR SCO, USA), Andrew Molthan (NASA SPoRT, USA), Lori Schultz (NASA SPoRT, USA), Amanda Weigel (SERVIR SCO, USA), Andrea Nicolau (SERVIR SCO, USA), Krishna Vadrevu (SERVIR SCO, USA), Raymond French (SERVIR SCO, USA), Esayas Gebremichael (NASA SPoRT, USA), Katherine Strattman (SERVIR SCO, USA).

Meeting organizers: Africa Flores-Anderson, Kelsey Herndon, Emil Cherrington, Leah Kucera

**Overall funding support:**

SilvaCarbon, NASA SERVIR, NASA Cooperative Agreement with the University of Alabama in Huntsville, NNM11AA01A.

We would like to thank and acknowledge SilvaCarbon colleagues Sylvia Wilson (USGS), Juliann Aukema (USAID), and Sasha Gottlieb (USFS) for directly supporting this effort and for their leadership of SilvaCarbon activities to strengthen capacity worldwide for forest and landscape monitoring. We would also like to present special thanks to USAID colleagues Pete Epanchin, Jennifer Frankel-Reed and Kevin Coffey for enabling and encouraging SERVIR-SilvaCarbon collaboration. A heartfelt thank you to Evan Notman from GFOI for planting the seed that was eventually transformed into this international capacity building effort.

We extend our deep appreciation to the Chief of the Earth Science Branch at NASA's Marshall Space Flight Center, Gary Jedlovec, for his support of the SERVIR program. We will also like to present special thanks to the University of Alabama in Huntsville, specifically John Christy and Laurie Collins for supporting meetings in Huntsville, AL, and for providing additional resources. A wholehearted thank you to Robert Griffin for his guidance and

valuable help to ensure this Handbook and complementary materials were accomplished successfully.

**Additional support:**

SERVIR Hub Institutions (RCMRD, AGRHYMET, ICIWOD, ADPC), and regional USAID Missions.

Our profound thanks to Paula Link and Kathleen Cutting from SERVIR-SCO for the invaluable support provided from the beginning to the end of this SAR Handbook initiative, supporting workshops, meeting preparations, and material development and printing. We also extend our sincere appreciation to Marta Chell and Claire Little for opening to us the beta version of the NASA DOIMS web portal and working to help generate DOIs.

We would also like to thank and acknowledge NASA SERVIR leadership, specifically NASA Applied Sciences: Capacity Building Program Manager, Nancy Searby, SERVIR Global Program Manager, Daniel Irwin, SERVIR Global Chief Scientist, Ashutosh Limaye, and SERVIR Project Manager, Tony Kim, for nurturing an environment of brotherhood and sisterhood across the SERVIR Global network, and for their vision to build geospatial capacity worldwide. Thanks go to former SERVIR Project Manager, Raymond French (NASA MSFC) for the work that ultimately enabled engaging with all the SAR experts authoring the chapters for this Handbook. We deeply thank the contributions of SERVIR Associate Chief Scientist, Eric Anderson, for his guidance and for additional reviews.

We would also like to show our gratitude to Natasha Stavros and Batu Osmanoglu, NISAR Deputy Program Applications Co-Leads, for their interest in making this content available to the broader NISAR community.

# PREFACE

**O**n behalf of both the U.S. Agency for International Development (USAID) and the U.S. Geological Survey (USGS), and specifically on behalf of the SilvaCarbon initiative, we are proud to share with you the following Handbook, the product of a collaboration between SilvaCarbon and the SERVIR program. Established in 2010, SilvaCarbon represents the US contribution to the Global Forest Observation Initiative (GFOI), itself a collaborative effort supporting countries in using Earth observation data for monitoring forests. SilvaCarbon's implementing agencies include USAID, the U.S. Department of State, the U.S. Forest Service (USFS), the USGS, the U.S. Environmental Protection Agency (EPA), the National Aeronautics and Space Administration (NASA), the National Oceanic and Atmospheric Administration (NOAA), and the Smithsonian Institution. This Handbook also represents an important contribution from a number of U.S.-based experts in Synthetic Aperture Radar (SAR), as well as experts from SERVIR's global network of hubs.

The motivation for this Handbook is to translate knowledge gained from decades of research in SAR into practical guidance to countries on how SAR can be used for different aspects of forest monitoring, reporting, and verification (MRV) for REDD+. There has been growing interest in applying this technology to land cover mapping and monitoring in the tropics, where seasonal and permanent cloud cover make detecting deforestation and forest degradation very challenging.

Radar data historically was known for being costly and complicated to use. However, with new datasets becoming available and open source, such as SENTINEL -1 from the European Space Agency (ESA), the fusion of optical and radar data becomes an option for sustainable and replicable methods. The drawback is the lack of historical radar data to include in historical baselines. However, the sooner SAR data is included in National Forest Monitoring Systems, the sooner it will be considered historical data in the future.

One of the GFOI's focus activities is providing 'Methods & Guidance' documentation to support countries' forest monitoring activities. This Handbook thus fits into that context as an important contribution to methods & guidance, especially since the body of available datasets and tools has been growing. One anticipated resource is the joint U.S. / India NISAR satellite mission whose launch is expected in the next few years.

The handbook walks you through the principles of SAR data applications from the beginning, starting from how to access the data and perform basic processing techniques. It describes how to use SAR data to map deforestation and forest degradation and how to estimate forest height. It also provides guidance on the best methods for using SAR to map and monitor forest biomass. It includes a chapter that exemplifies the use of radar for mapping mangrove forests. It concludes with the important issue of choosing a sampling design while using SAR data for biomass estimation.



The handbook was first conceptualized as an outcome from a workshop hosted by SERVIR. This workshop brought together scientists, program managers, and country practitioners to identify challenges on using SAR data, gaps where SAR data is not available, and potential areas where SAR could fill the gap in forest monitoring for remote sensing data. Thanks to the efforts of the SERVIR team, especially Africa Flores for managing the overall initiative, the scientists involved in the development of the handbook, and the SilvaCarbon team, this handbook offers a set of tools and operational methods that will streamline efforts to assist countries to build robust, transparent, replicable and verifiable Monitoring, Reporting and Verification Systems.

We therefore invite you to take advantage of this important resource, and feel free to provide us with feedback on how the Handbook can be improved, as we hope that this will evolve into a truly living document. We also take this opportunity to recognize the contributions of the subject matter experts who drafted the bulk of the Handbook, our counterparts from the respective SERVIR hubs across the globe, and our partners at NASA's SERVIR Science Coordination Office. Thank you.



  
**Juliann Aukema**  
SilvaCarbon Coordinator, USAID



  
**Sylvia Wilson**  
SilvaCarbon Coordinator, USGS

# FOREWORD

Two years ago, a group of scientists and practitioners representing a dozen countries across Africa, Asia, and the Americas identified a pressing need and opportunity for the applied Earth observations and international development communities. Global and national commitments to sustainable landscape management—including forests, mangroves, and the biomass they store and CO<sub>2</sub> they capture—has challenged scientists and resource managers to develop and implement new, accurate, and cost effective monitoring and reporting systems. Field measurements combined with satellite remote sensing techniques have provided industry-standard inputs into monitoring, reporting, and verification systems. In the last decade, critical access to satellite data has skyrocketed, thanks largely to public releases of over 40 years of Landsat data from the NASA and USGS, along with the European Space Agency's (ESA) free and open data policies under the Copernicus Sentinel series. However, data access alone does not guarantee appropriate use. Tools and training are important steps in ensuring adequate capacity at individual and institutional levels.

This Handbook represents a joint contribution from the U.S. government-led SilvaCarbon initiative, and the joint NASA-USAID SERVIR program, to support global capacity building endeavors as called for by the Global Forest Observations Initiative (GFOI). SERVIR's global network of international technical centers of excellence, known as "SERVIR Hubs", played a crucial role in defining needs and initial expansion of Synthetic Aperture Radar (SAR) capacity. SERVIR Hubs have deep knowledge of existing national and regional capacities in remote sensing for forestry and biomass monitoring, which articulated the critical gaps addressed in this Handbook.

A common challenge that the applied remote sensing community faces in forestry and landscape monitoring are clouds. For years, SAR promised all-weather, day-and-night capability, but at a steep cost. Until the launch of the Sentinel-1 series by ESA's Copernicus Program and the release of archived ALOS-1 imagery by JAXA, SAR data were effectively inaccessible and inappropriate for national and regional level forestry and biomass monitoring. The forthcoming NASA and Indian Space Research Organization (ISRO) SAR mission, NISAR, will only add to the free access of SAR data.



The series of chapters in this handbook are authored by leading global experts in SAR remote sensing fundamentals and applications in this field, and co-developed with professionals who thrive at the transition of research to applications for societal benefit. Through careful testing and curation, these materials are meant to complement existing national, regional, and global methods in forestry and biomass estimation. We are proud to share this as a multilateral contribution to improve the use of free satellite data toward better monitoring and management of our terrestrial environments.



**Daniel Irwin**  
*SERVIR Global Program Manager, NASA*



*The chapter authors would like to thank Sylvia Wilson, Juliann Aukema, Sasha Gottlieb and Evan Notman for supporting the SERVIR-SilvaCarbon initiative through SilvaCarbon and GFOI. Thanks go to SERVIR Hub representatives, particularly Land Cover & Land Use Change thematic leads, Science and Data leads, Chiefs of Party and Directors of Institutions hosting SERVIR Hubs, for their investments in this SAR capacity building effort. The development of this SAR Handbook has been truly a community effort that involved the global SERVIR network, including stakeholder institutions, SAR experts and interagency partners in the forestry community. We would also like to thank all of the SAR experts authoring this book; Franz Meyer, Josef Kellndorfer, Sassan Saatchi, Marc Simard, Paul Siqueira, and Hans Andersen for their valuable contributions and for the development and delivery of hands-on trainings in SERVIR Hub regions.*

# CHAPTER 1

## Introduction and Rationale

Africa Ixmucane Flores-Anderson<sup>1,2</sup>, Kelsey Herndon<sup>1,2</sup>, Emil Cherrington<sup>1,2</sup>, Rajesh Thapa<sup>3</sup>, Leah Kucera<sup>1,2</sup>, Nguyen Hanh Quyen<sup>5</sup>, Phoebe Odour<sup>4</sup>, Anastasia Wahome<sup>4</sup>, Karis Tenneson<sup>6</sup>, Bako Mamane<sup>7</sup>, David Saah<sup>6,8</sup>, Farrukh Chishtie<sup>5</sup>, Ashutosh Limaye<sup>1</sup>

### ABSTRACT

This Synthetic Aperture Radar (SAR) handbook of applied methods for forest monitoring and biomass estimation has been developed by SERVIR in collaboration with SilvaCarbon to address pressing needs in the development of operational forest monitoring services. Despite the existence of SAR technology with all-weather capability for over 30 years, the applied use of this technology for operational purposes has proven difficult. This handbook seeks to provide understandable, easy-to-assimilate technical material to remote sensing specialists that may not have expertise on SAR but are interested in leveraging SAR technology in the forestry sector.

This introductory chapter explains the needs of regional stakeholders that initiated the development of this SAR handbook and the generation of applied training materials. It also explains the primary objectives of this handbook. To generate this applied content on a topic that is usually addressed from a research point of view, the authors followed a unique approach that involved the global SERVIR network. This process ensured that the content covered in this handbook actually addresses the needs of users attempting to apply cutting-edge scientific SAR processing and analysis methods. Intended users of this handbook include, but are not limited to forest and environmental managers and local scientists already working with satellite remote sensing datasets for forest monitoring.

### 1.1 Background

As highlighted by the Global Forest Observation Initiative (GFOI) in their Methods and Guidance Document (MGD) (GFOI 2017), SAR datasets have proven useful as a source of activity data for forest Monitoring, Reporting, and Verification (MRV) systems at national and regional levels. SAR datasets can provide substantial aid to MRV systems, particularly in areas with persistent cloud cover, due to the sensor's all-weather capability (Reich et al. 2016).

The MGD also highlights the potential of SAR datasets with higher spatial resolution (<5 m) to characterize forest canopy. SAR captures different target parameters than optical sensors, therefore providing unique information that complements standard optical remote sensing methods. A helpful analogy would be that while the energy captured by optical sensors of a green leaf relates to its amount of chlorophyll (or "greenness"), the amount of microwave energy (part of the electromagnetic spectrum used in SAR) scattered by the leaf would be proportional to its size,

shape, and water content (Woodhouse 2006). SAR indeed has a strong sensitivity to forest structure and biomass (Saatchi 2015).

GFOI's MGD also documents the limitations of optical data such as Landsat and MODIS to estimate forest biomass and to detect early regrowth of secondary vegetation, including a limited ability to expose small disturbances such as removal of individual trees (GFOI 2017). The main premise and recommendation of GFOI's MGD is to combine both remote sensing technologies, optical and SAR. However, the main limitation faced over and over by users is how to start ingesting SAR when its assimilation and analysis is rather difficult, considered highly sophisticated, and previously limited only to experienced professionals.

To clearly define the need for SAR technology for forest monitoring systems, a needs assessment was carried out within the SERVIR global network. SERVIR is a joint initiative between NASA and the U.S. Agency for International Development (USAID) that fosters the use of Earth observations to assess environmental conditions to improve decision-making actions.

Then, in collaboration with SilvaCarbon, a plan was designed and implemented to effectively address needs and knowledge gaps of remote sensing specialists working in MRV systems to use SAR. SilvaCarbon is an interagency technical cooperation program of the U.S. government to enhance the capacity of selected tropical countries in measuring, monitoring, and reporting on carbon in their forests and other lands. In addition, SilvaCarbon is the U.S. primary contributor to GFOI, where their activities focus on capacity building.

### 1.2 Limitations and Opportunities of Applying SAR Technologies

Multiple research efforts investigate empirical relationships between SAR backscatter and biophysical forest properties, particularly aboveground biomass (Woodhouse 2006, Reich et al. 2016). Yet there are a number of limitations that are well summarized by GFOI's MGD in using SAR for biomass estimation:

- Depending on wavelength, rapid saturation of

<sup>1</sup> NASA Marshall Space Flight Center / SERVIR Science Coordination Office, 2 University of Alabama in Huntsville, 3 International Centre for Integrated Mountain Development (ICIMOD) / SERVIR-Hindu Kush Himalaya, 4 Regional Centre for Mapping of Resources for Development (RCMRD) / SERVIR-Eastern & Southern Africa, 5 Asian Disaster Preparedness Center (ADPC) / SERVIR-Mekong, 6 Spatial Informatics Group (SIG) / SERVIR-Mekong, 7 Agro-meteorology, Hydrology, and Meteorology regional center (AGRHYMET) / SERVIR-West Africa, 8 University of San Francisco

the signal at low aboveground biomass stock (~50–100 t C/ha) (Gibbs et al. 2007)

- Increased errors due to terrain
- Rainfall and soil moisture effects
- Localized algorithm development focused on a single biome or mono-species stands
- Lack of consistency in estimates as a function of sensor parameters

Of particular concern is the limited transferability of algorithms within and between different forest structural types (GFOI 2017). **Chapter 5** of this handbook addresses this particular topic in depth. Though these limitations may confine some applicability of SAR technology, this handbook addresses them effectively and provides users with practical information to facilitate the appropriate use of SAR data for accurate results. It explains state-of-the-art methods not only with theory but also with hands-on exercises.

The use of SAR datasets to accurately monitor processes of deforestation, land and forest degradation, and secondary forest regrowth (Hoekman & Quinonez 2000) has produced promising results that should not be overlooked. **Chapter 3** provides practical information on change detection and forest degradation analysis. All chapters provide key information that addresses the listed challenges in working with SAR data. Practical workflows are also included to provide SAR analysis tools specific for stated user needs and forest monitoring applications.

With publicly available C-band data from the Sentinel-1 mission (Malenovský et al. 2012, Berger et al. 2012) since 2014, the outlook of SAR operational use has changed significantly, and in the most promising way. This, in addition to the forthcoming availability of public L-band data from NISAR and TanDem-L, completely changes the game in terms of potential operational use of SAR datasets. Previously designated Research and Development (R&D) topics by GFOI (2013) now use SAR for forestry applications such as change detection within forest land, near-real time forest change indicators, and forest stratification.

This previous R&D has enabled current and operational use of SAR for these applications. This handbook strives to provide technical materials to facilitate these new resources into operational use.

### 1.2.1 FROM NEEDS TO PRACTICE

Though extremely useful and educational, the most complete literature available on SAR is constrained mainly to textbooks (Woodhouse 2006, Shimada 2018) and does not focus on an applied approach needed by technicians to start processing SAR datasets. In addition, there are a number of studies and peer-reviewed publications (Moreira et al. 2013, Hoekman & Quirion 2000) that add to the wealth of knowledge, but also lack applied content that facilitates uptake of SAR technology. This handbook represents a comprehensive resource for using SAR datasets for forestry and biomass applications, and also augments recent efforts to generate reference documentation for interpreting SAR datasets (CEOS 2018), given its applied focus that includes processing workflows for specific forestry applications.

The content of this handbook has been driven by the needs of the community (see **Sec. 1.3**), and as such, is strongly focused on forest biomass estimation. However, it also covers a range of topics from basic preprocessing to change detection, including mangrove monitoring, Forest Stand Height (FSH), and sampling design for uncertainty estimation. Given that current state-of-the-art methods use SAR data for forest biomass estimation as well as the ongoing research on how to better understand forest backscatter signals (Brolly & Woodhouse 2012), this handbook clearly describes the limitations and advantages of using SAR for specific forest applications. It also conveys in practical and understandable terms the main considerations and concepts technical users should be aware of when analyzing SAR data. It is the authors' intent to generate applied knowledge by leveraging the wealth of research knowledge that has been gathered over the past 30 years.

SERVIR and SilvaCarbon are committed to strengthening the technical capacity of users working on forest monitoring applications, and as new SAR missions are launched and new technologies emerge, it becomes crucial to join efforts to achieve this goal.

## 1.3 Needs Assessment

### 1.3.1 WHO IS SERVIR?

SERVIR operates through established technical organizations (called hubs) with unique sets of capabilities, including political buy-in from member

countries, technical expertise in remote sensing, geographic information systems, and database management, along with strong relationships with stakeholders. Science within SERVIR is coordinated through the Science Coordination Office (SCO) at NASA Marshall Space Flight Center. All SERVIR activities and projects are user-driven and involve stakeholders who play a key role in service development and uptake. SERVIR's network of hubs includes:

- Eastern and Southern Africa, with the regional hub at the Regional Center for Mapping of Resources for Development (RCMRD) in Nairobi, Kenya;
- Hindu-Kush Himalaya, with the regional hub at the International Center for Integrated Mountain Development (ICIMOD) in Kathmandu, Nepal;
- Lower Mekong River, with the regional hub at the Asian Disaster Preparedness Center (ADPC) in Bangkok, Thailand;
- West Africa with the regional hub at Centre Regional de Formation et d'Application en Agrométéorologie et Hydrologie Opérationnelle (AGRHYMET) in Niamey, Niger; and
- Amazonia, with the regional hub based at the International Center for Tropical Agriculture (CIAT) in Cali, Colombia. Because this is the newest hub in the SERVIR network and the agreement to start activities did not take effect until December 2018, it was not included in the development of this material.

SERVIR hubs are comprised of recognized experts in their respective regions for satellite remote sensing and have developed multiple applications related to SERVIR's thematic areas, including food security and agriculture, water and water-related disasters, weather and climate, and land cover/land use. Forest monitoring is one of the common threads among the services SERVIR hubs provide in the land cover/land use change thematic area. SERVIR's stakeholders have identified forest biomass estimation using remote sensing as a key need.

Measurements of forest biomass can advance the understanding of forests in the global carbon cycle (Pan et al. 2011). In addition, forest biomass estimates are useful to support the implementation of Reducing Emissions from Deforestation and forest Degradation

(REDD+) initiatives in different regions of the world.

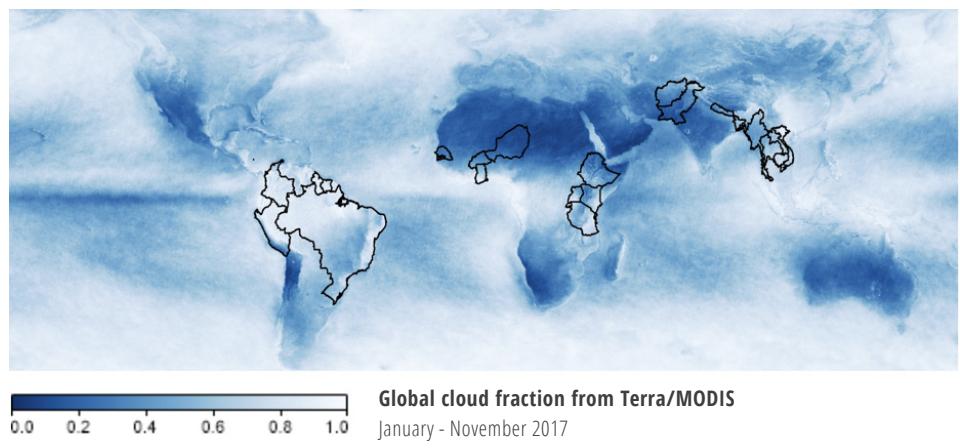
Demand to enhance forest monitoring systems by incorporating additional satellite datasets and to derive forest biomass has increased due to newly available satellite datasets that incorporate SAR technology (De Sy et al. 2012). Currently, there are free and open historical archives (2006–2011) of the Advanced Land Observing Satellite-1 (ALOS-1) SAR datasets, and Sentinel-1 has been providing free and open SAR imagery since 2014. In addition, new forthcoming missions such as NISAR, Biomass, and TanDEM-L will also provide free and open SAR data.

The recent widespread availability of SAR data coupled with the need to overcome cloud cover in most SERVIR regions highlight the need for understanding how these new and free SAR satellite resources might be leveraged, especially in the context of open source and freely available software for land cover and use applications (Brovelli et al. 2018). **Figure 1.1** shows cloud coverage over the five SERVIR regions, including the most recent SERVIR-Amazonia hub.

The SilvaCarbon program provides targeted technical support to countries in the process of developing and implementing national forest and landscape monitoring systems. SilvaCarbon leverages state-of-the-art science and technology to advance the generation and use of improved information related to forest and terrestrial carbon. It is within this context that SERVIR and SilvaCarbon joined efforts to generate state-of-the-art technical materials that will make meaningful contributions to support these countries' efforts to measure, monitor, and report on carbon in their forests.

Although the potential of SAR data is well recognized, the current level of expertise in operational use of SAR data for LCLUC applications, forest mapping/monitoring, and (more specifically) biomass estimation is limited in SERVIR regions. This limitation not only applies to scientists in the SERVIR network, but is also observed worldwide (Reich et al. 2016).

SERVIR hubs have advanced expertise in using optical remote sensing for land cover mapping, which is reflected in the multiple LCLUC projects executed over the years (<http://catalogue.servirglobal.net>). SERVIR hubs and their partners have requested support to increase their capacities to process and analyze SAR



**Figure 1.1** Cloud coverage for year 2017 over SERVIR regions

data for forest monitoring and biomass estimation applications.

The SERVIR SCO performed a needs assessment across the SERVIR hubs to collect information on current capacity, main services, and requests on land cover/land use change applications. This needs assessment inspired an effort to work with SAR data for forest applications. The authors recognize that their focus in the SERVIR network does not represent a thorough global needs assessment; however, it like-

ly provides a good representation of the needs facing remote sensing specialists around the globe in forest monitoring applications (**Table 1.1**).

In addition, due to the authors' collaboration with SilvaCarbon, the focus of this handbook is to contribute to and to be in alignment with their objectives, which include enhancing the capacity of countries for forest MRV systems. Through this collaboration, the authors aim to address user needs through the development of a distinctive product that could serve

SERVIR REGION & HUB	PRIMARY STAKEHOLDERS	CURRENT RS TECHNOLOGY USED FOR MRV SYSTEMS	RELEVANT SAR CAPABILITIES	POTENTIAL APPLICATIONS
<b>West Africa</b> AGHYMET	Ghana Forestry Commission, University of Ghana	Monitoring permanent plots for AGHYMET (GCCA Project, starting support for countries to establish national forest MRV systems	Matching ancillary data with the RS (SAR) within a pilot site, upscaling to cover all of the West African sub-region.	Integration of SAR to support MRV systems Estimation of forest biomass in West Africa
<b>Eastern &amp; Southern Africa</b> RCMRD	Kenya Forest Service, Kenya Water Towers Agency, Kenya Marine and Fisheries Institute, Kenya REDD+, CIFOR, other agencies	Freely-available optical RS used; complicated by persistent cloud cover	Gain knowledge to start processing SAR data in general. This knowledge will go a long way in supporting the countries as well as in the provision of data that can be used by the relevant authorities in decision making.	Forest biomass estimation. Forest monitoring in cloudy regions Support for REDD+ MRV
<b>Hindu Kush-Himalaya</b> ICIMOD	FRTC - Nepal, DFPS - Bhutan	Optical RS Object-based image analysis Optical RS for MRV dev	Strength on capturing biomass estimates Gain knowledge to process SAR data and develop biomass estimation models	Forest biomass estimation to support national communications to UNFCCC for assessing carbon fluxes Accurate forest biomass estimation
<b>Mekong</b> ADPC	Forest Inventory and Planning Institute of Viet Nam	For Activity data: using medium resolution optical satellite images (Landsat and Sentinel 2)  For Emission factors: using National systematic sample plots	Basic SAR processing (Sentinel 1) and application in mapping to recognized forest/nonforest	Using SAR to improve volume-based estimation for forest quality, monitoring forest degradation  Applied SAR for operation on forest change detection, forest plantation monitoring  Forest biomass estimation

GCCA: Global Climate Change Alliance  
FRTC: Forest Research and Training Centre  
DFPS: Department of Forests and Park Services

**Table 1.1** List of needs to apply SAR technology in SERVIR regions as identified by each SERVIR Hub during the needs assessment process.

to lower the barrier and create wise users of SAR technology. Therefore, building from that expertise, the authors have focused on capacity building.

Training SERVIR hubs and their partners on the use of SAR data will further strengthen their remote sensing capabilities. Researchers in the SERVIR network represent a link between national agencies, user groups and the global user/producer community, and NASA scientists; thus, we envision that engagement with SERVIR network researchers on SAR data capacity building and training activities will maximize benefits, in addition to complementing their current optical remote sensing methods with SAR methodologies.

## 1.4 Objectives

The main objective of this handbook is to provide practical guidance on the application of SAR technology for forest monitoring and biomass estimation. It addresses a gap that would otherwise exist on how to process SAR imagery for practical forest applications that can benefit from incorporating SAR technology. More specific objectives include:

- Disseminate practical knowledge on using SAR imagery for forest mapping/monitoring and biomass estimation
- Generate applied theoretical and hands-on materials that will enhance operational uptake of SAR technology for forest monitoring and biomass estimation
- Support SilvaCarbon objectives in developing good practices and approaches for using SAR data for MRV systems

## 1.5 SERVIR Approach

To meet the goals described above, SERVIR followed the next steps:

1. Scoping workshop with SAR experts and SERVIR user community
2. Define main topics for training and subsequent handbook
3. Recruit subject matter experts on SAR applications (SAR experts) to provide training and develop applied documentation
4. “Train the trainer” approach:

- Generate descriptive training material using open source software
- Hold sequential training events based on regional needs
- Trainees repeat training at their centers
- Review and update training materials per feedback collected at training events

Rather than creating another product—such as a local or regional biomass product for a given region that later becomes outdated and which nobody can replicate—the authors chose to develop a strong capacity building approach, enabling and strengthening technical capacities of users to create such a product. This entailed work with world-renowned SAR experts that are also the authors of individual chapters in this handbook. These experts also created training tutorials that are included in the appendices of this handbook and on the handbook webpage.

To identify the main technical topics this handbook addresses, the authors conducted a scoping meeting in February 2017, where SAR experts and technical experts on forest applications from SERVIR regions came together. This scoping meeting served to select topics for technical training that became the main topics addressed in this handbook. Hence, as portrayed in **Table 1.2**, the trainers of these international workshops are the authors of the main technical chapters in this handbook (**Chapters 2–7**). The content they cover in their chapters and training tutorials have been tested at hands-on workshops and reviewed by remote sensing specialists of the SERVIR global network and other

SAR specialists to ensure functionality and usability.

Main deliverables include:

- **SAR Handbook of Applied Methods for Forest Monitoring and Biomass Estimation**—The [SAR Handbook](#) consists of eight chapters that include theory and background on a wide range of topics related to monitoring forests with SAR, as well as appendices that include step-by-step guides to applying the theory to practice. The authors envision this resource being used to develop forest applications and for capacity-building efforts. Chapters can be used together as a relatively complete and cohesive source of information on monitoring forests with SAR, or each chapter can stand alone as a resource on specific topics.
- **Hands-on training materials**—The SAR Handbook website hosts the materials necessary to complete the training described in the appendices. It includes PDFs of the handbook and step-by-step instructions, PowerPoint presentations to complement the training, scripts to process SAR data, and ancillary datasets, such as lidar and in-situ measurements.
- **One-pagers**—The editors have identified the need for quick reference guides for various SAR topics, since no such reference exists for using SAR to monitor forests. The editors developed a number of one-page documents that provide clear and easy-to-read summaries of important SAR concepts discussed more in-depth within

SAR EXPERT	TRAINING	HOSTING HUB(S)	LOCATION & DATES
<b>Franz Meyer</b> Univ. of Alaska Fairbanks	SAR Basics, Forest Degradation & Deforestation	1) West Africa 2) Hindu Kush-Himalaya	Niamey, Niger (Jan 29 - Feb 2, 2018) Kathmandu, Nepal (Feb 12-16, 2018)
<b>Josef Kellndorfer</b> Earth BigData			
<b>Paul Siquiera</b> Univ. of Mass. Amherst	Forest Stand Height	Mekong	Bangkok, Thailand (Mar 12-16, 2018)
<b>Marc Simard</b> CalTech/NASA JPL	Mangroves, Sampling Design	Eastern & Southern Africa	Nairobi, Kenya (April 16-20, 2018)
<b>Hans Andersen</b> US Forest Service			
<b>Sassan Saatchi</b> CalTech/NASA JPL	Biomass Estimation	Hindu Kush-Himalaya	Kathmandu, Nepal (April 30-May 4)

**Table 1.2** List of global SAR capacity building workshops for international partners in Africa and Asia.

the handbook. These one-pagers include “SAR Data Access and Availability,” “SAR Vegetation Indices,” “SAR Preprocessing Steps,” “SAR for Biomass Estimation,” and “SAR for Forest Stand Height (FSH) Estimation.” Digital versions of these one-pagers can be found on the handbook webpage or in the handbook appendices.

- **Technical videos**—The editors have developed short animated videos to simplify and clearly communicate complex concepts found within the handbook. These include an introduction to SAR concepts (wavelength, penetration depth, polarization, etc.), FSH estimation, and biomass estimation.

## 1.6 What to Expect

This handbook was developed to generate applied knowledge on using SAR for forest applications. The content has been generated by world-renowned experts on the topic and vetted, tested, and reviewed by a community of applied remote sensing users. It covers basic concepts to understand how SAR technology works and identifies some of the best practices and approaches to estimating forest change, biomass, and stand height; to mapping mangroves and estimating their biomass; and to sampling design for uncertainty estimation of biomass maps.

All the training tutorials use open source software and programming languages to process and analyze SAR datasets. This was a requirement that was fulfilled by the SAR experts that generated the materials.

This handbook is comprised of eight chapters in total, including this introductory **Chapter 1** and the following:

- **Chapter 2**, Spaceborne Synthetic Aperture Radar – Principles, Data Access, and Basic Processing Techniques. Author: Franz Meyer—This first technical chapter explains basic concepts to start using SAR datasets. It covers basic preprocessing and the peculiarities of SAR imagery, which enables an understanding of how these datasets are interpreted. This chapter also provides a comprehensive inventory of past, current, and planned SAR sensors.
- **Chapter 3**, Use of SAR Data for Mapping Deforestation and Forest Degradation. Author: Josef Kellndorfer—This chapter focuses on the applications of SAR imagery for forest change detection. It discusses how SAR backscatter changes due to sensor and target parameters, with an emphasis on forest targets. It also explains an approach for time series analysis for forest change detection.
- **Chapter 4**, Forest Stand Height. Author: Paul Siqueira—This chapter discusses the estimation of Forest Stand Height (FSH) through the use of spaceborne SAR, especially at L-band repeat-pass Interferometric SAR (InSAR). It covers the theory and software, and provides examples for the use of repeat-pass InSAR for FSH estimation.
- **Chapter 5**, SAR Methods for Mapping and Monitoring Forest Biomass. Author: Sassan Saatchi—This chapter provides a summary of the methodologies and techniques for estimating forest aboveground biomass. The content covers state-of-the-art SAR remote sensing approaches for characterizing vegetation structure and biomass estimation, and provides resources for future developments in the technology and emergency methodologies.
- **Chapter 6**, Radar Remote Sensing of Mangrove Forests. Author: Marc Simard—This chapter addresses the use of SAR imagery to monitor changes in the mangrove forest extent. The state-of-the-art radar remote sensing techniques to measure and monitor mangrove forest structure are also covered in this chapter.
- **Chapter 7**, Sampling Designs for SAR-Assisted Forest Biomass. Author: Hans Andersen—This chapter discusses sampling design and statistical modeling/estimation frameworks to provide a sound, statistically rigorous assessment of the uncertainty of forest biomass maps. It provides examples for efficiently using expensive field plot data and more extensive use of less expensive, remotely sensed information.
- **Chapter 8**, Perspectives on the Future Application of SAR in Forest and Environmental Monitoring. Authors: Emil Cherrington et al.—This chapter discusses future and emerging applications of SAR for forest and environmental monitoring, and also reflects on how SAR is currently being used across a range of applications. This chapter discusses how this is expected to change due to the growing public availability of SAR data and platforms to process and analyze radar data. It also discusses how SERVIR regional hubs are applying SAR technology, and how the SERVIR global network can be important resource centers in support of SilvaCarbon and GFOI to aid in articulating and addressing new environmental monitoring challenges.

## 1.7 References

- Brovelli, M. A., Minghini, M., Molinari, M. E., Wu, H., Zheng, X., & Chen, J. (2018). Capacity building for high-resolution land cover intercomparison and validation: What is available and what is needed. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(4W8), 15–22. <https://doi.org/10.5194/isprs-archives-XLII-4-W8-15-2018>
- Berger, M., Moreno, J., Johannessen, J.A., Levelt, P.F., Hanssen, R.F., 2012. ESA's sentinel missions in support of Earth system science. *Remote Sensing of Environment* 120, 84-90.
- Brolly, M., Woodhouse, I.H., 2012. A "Matchstick Model" of microwave backscatter from a forest. *Ecological Modelling* 237–238, 74–87. <https://doi.org/10.1016/j.ecolmodel.2012.04.014>
- CEOS. (2018). *A Layman's Interpretation Guide to L-band and C-band Synthetic Aperture Radar data* (Vol. Version 2). [http://ceos.org/document\\_management/SEO/DataCube/Laymans\\_SAR\\_Interpretation\\_Guide\\_2.0.pdf](http://ceos.org/document_management/SEO/DataCube/Laymans_SAR_Interpretation_Guide_2.0.pdf)
- De Sy, V., Herold, M., Achard, F., Asner, G. P., Held, A., Kellndorfer, J., & Verbesselt, J. 2012. Synergies of multiple remote sensing data sources for REDD+ monitoring. *Current Opinion in Environmental Sustainability*, 4(6), 696-706.
- Gibbs, H.K., Brown, S., Niles, J.O., Foley, J.A., 2007. Monitoring and estimating tropical forest carbon stocks: making REDD a reality. *Environmental Research Letters* 2, 045023. <https://doi.org/10.1088/1748-9326/2/4/045023>
- GFOI (2013): *Review of Priority Research & Development Topics: R&D related to the use of Remote Sensing in National Forest Monitoring*. Pub. GEO, Switzerland, 2013 ISBN 978-92-990047-5-3.
- Global Forest Observation Initiative (GFOI), 2017. *Integration of remote-sensing and ground-based observations for estimation of emissions and removals of greenhouse gases in forests*. Online document, visited on Dec 2018. <http://www.fao.org/gfoi/resources/documents/detail/en/c/1068378/>
- Hoekman, D.H., Quiriones, M.J., 2000. Land cover type and biomass classification using AirSAR data for evaluation of monitoring scenarios in the Colombian Amazon. *IEEE Transactions on Geoscience and Remote Sensing* 38, 685–696. <https://doi.org/10.1109/36.841998>
- Malenovský, Z., Rott, H., Cihlar, J., Schaepman, M.E., García-Santos, G., Fernandes, R., and Michael, B., 2012. Sentinels for science: Potential of Sentinel-1,-2, and-3 missions for scientific observations of ocean, cryosphere, and land. *Remote Sensing of Environment* 120, 91-101.
- Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I., Papathanassiou, K.P., 2013. A tutorial on synthetic aperture radar. *IEEE Geoscience and Remote Sensing Magazine* 1, 6–43. <https://doi.org/10.1109/MGRS.2013.2248301>
- Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., ... & Ciais, P. 2011. A large and persistent carbon sink in the world's forests. *Science*, 1201609.
- Reiche, J., Lucas, R., Mitchell, A.L., Verbesselt, J., Hoekman, D.H., Haarpaintner, J., Kellndorfer, J.M., et al., 2016. Combining satellite data for better tropical forest monitoring. *Nature Climate Change* 6, no. 2, 120.
- Saatchi, S., 2015. *Mapping tropical forest biomass: Synthesis of ground and remote sensing inventory*. The High Carbon Stock Science Study 2015.
- Shimada, M., 2018. *Imaging from Spaceborne and Airborne SARs, Calibration, and Applications*, SAR Remote Sensing. Taylor & Francis Group.
- Woodhouse, I. (2006). *Introduction to Microwave Remote Sensing*. Boca Raton: CRC Press, <https://doi.org/10.1201/9781315272573>





**DR. FRANZ MEYER'S** research focuses on the theory and applications of Synthetic Aperture Radar (SAR) remote sensing data, including the development of SAR processing technology, the analysis of error sources in SAR and InSAR data, and the use of SAR in disaster monitoring. Dr Meyer also works as the Chief Scientist of the Alaska Satellite Facility, NASA's Prime Data Center for SAR data where he is responsible for interactions with the growing SAR user community, the generation of new value added products, and the development of SAR education and training resources. He is a member of the NASA Science team for the US/Indian NISAR satellite and a regular NASA PI for projects focused on SAR research and SAR capacity building.

*I want to thank Dr. Olaniyi A Ajadi, Dr. Heming Liao, and Dr. Wenyu Gong (all UAF) for providing input on the development of SAR processing algorithms described in this chapter. All Sentinel-1 SAR data shown above were provided by the Copernicus program and processed by the European Space Agency. All SAR data were accessed through the services of the NASA Alaska Satellite Facility SAR Distributed Active Archive Center (DAAC). I owe gratitude to David B McAlpin (UAF) for proof-reading early drafts of the manuscripts, and to Africa Flores, Kelsey Herndon, and Emil Cherrington (all SERVIR) for their valuable input and editorial support. Finally, I want to thank Bako Mamane (AGRHYMED Regional Centre), Rajesh Thapa (ICIMOD), and all the attendees of our SAR training workshops in Niger and Nepal for their hospitality and for their interest in the rapidly evolving field of radar remote sensing.*

# CHAPTER 2

## Spaceborne Synthetic Aperture Radar: Principles, Data Access, and Basic Processing Techniques

Franz Meyer, Associate Professor for Radar Remote Sensing, UAF & Chief Scientist of Alaska Satellite Facility

### ABSTRACT

This chapter provides background information and hands-on processing exercises on the main concepts of Synthetic Aperture Radar (SAR) remote sensing. After a short introduction on the peculiarities of the SAR image acquisition process, the remainder of this chapter is dedicated to supporting the reader in interpreting the often unfamiliar-looking SAR imagery. It describes how the appearance of a SAR image is influenced by sensor parameters (such as signal polarization and wavelength) as well as environmental factors (such as soil moisture and surface roughness). A comprehensive list of past, current, and planned SAR sensors is included to provide the reader with an overview of available SAR datasets. For each of these sensors, the main imaging properties are described and their most relevant applications listed. An explanation of SAR data types and product levels with their main uses and information on means of data access concludes the narrative part of this chapter and serves as a lead-in to a set of hands-on data processing techniques. These techniques use public domain software tools to walk the reader through some of the most relevant SAR image processing routines, including geocoding and radiometric terrain correction, interferometric SAR processing, and change detection.

## 2.1 On the Concepts of Imaging Radars

### 2.1.1 A WORD ABOUT HISTORY

The invention of RAdio Detection And Ranging, or radar, as a concept for detecting and localizing objects in a three-dimensional space dates back to the turn of the 20<sup>th</sup> century and is typically credited either to the German inventor and entrepreneur Christian Huelsmeyer—who proposed the so-called “Telemobiloskop” as an active microwave-based system for detecting distant metallic objects (Vollmar 1960)—or to the British engineer Robert Watson-Watt, who in June of 1935 successfully demonstrated an object detection and ranging system that was capable of accurately locating airborne objects up to a distance of about 30 km (Watson-Watt 1946). Once invented, radar technology developed rapidly during the World War II era, motivated mostly by air defense and over-the-horizon surveillance considerations. By the early 1940s, radars had become small enough to be implemented on airplanes, expanding the application

realm of radar systems into a range of new fields, including the growing discipline of Earth observation. This chapter discusses the application of imaging radar sensors to this discipline.

### 2.1.2 SIDE-LOOKING AIRBORNE RADARS

The allure of using radar systems for imaging purposes mostly stems from the all-weather and all-day capabilities that can be provided by this sensor type. These capabilities are advantageous for many surveillance applications, allowing for regular mapping of areas affected by heavy cloud cover, persistent rain, or extended darkness. Additionally, radar signals interact differently with the surface than most other sensing systems, providing interesting new information about the observed environment.

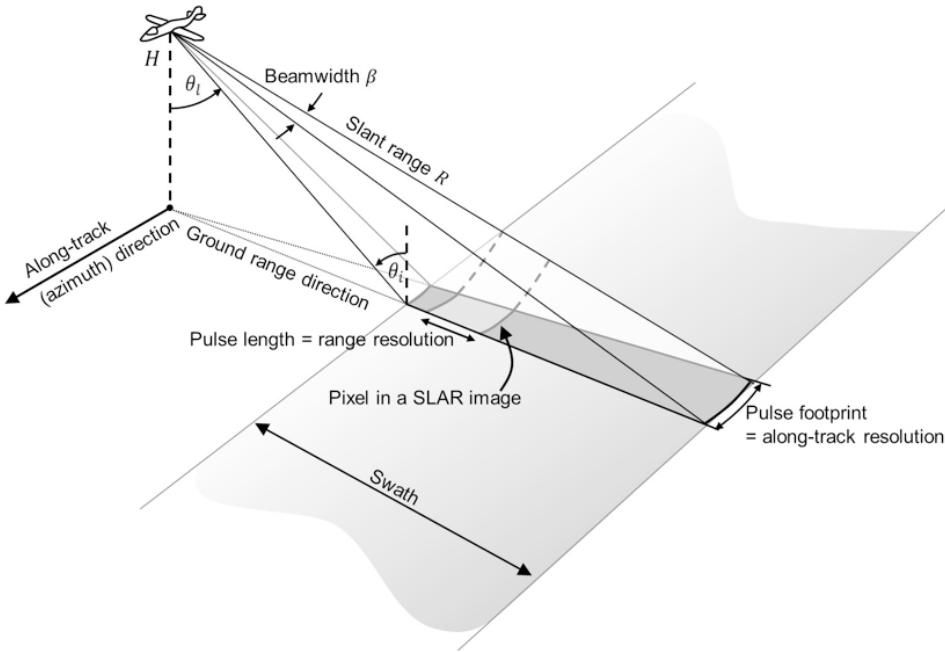
With the development of Side-Looking Airborne Radar (SLAR) systems in the 1950s, the first airborne radar systems with reliable imaging performance became available. The observation configuration of a SLAR system is shown in **Figure 2.1** and consists of a radar sensor mounted on an airborne (or spaceborne) platform that, in this simplified exam-

ple, is moving along a straight path at altitude  $H$ . Unlike most optical imaging systems, which point their sensors towards nadir, the antenna of a SLAR (and any other imaging radar) system is pointed away from nadir by a so-called look angle  $\theta$ , such that it illuminates a continuous swath on the ground as the aircraft moves along..

While flying along its track, the radar system is transmitting a sequence of short microwave pulses of pulse length  $\tau_p$ , each of which illuminates an instantaneous area on the ground that is usually referred to as the antenna footprint (see darker gray area in **Fig. 2.1**). The size  $S$  of this instantaneous footprint in either the range or along-track (azimuth) direction is largely defined by the relationship between system wavelength  $\lambda$  and the side length of the antenna  $L$  (defining the antenna’s beamwidth through  $\beta = \lambda/L$ ) along this direction, as well as by the distance of the radar sensor from the ground  $R$ :

$$S \approx \frac{\lambda}{L} R = \beta \cdot R \text{ [m]} \quad (2.1)$$

To form a two-dimensional image, the echoes received from the ground are sorted by their arrival



**Figure 2.1** Observation geometry of a SLAR imager. The radar flies along a straight line at altitude  $H$  and observes Earth at an oblique look angle  $\theta_l$ . Instead of the look angle, sometimes the incidence angle  $\theta_i = (90^\circ - \theta_l)$  is annotated. The size of the illuminated footprint is defined by the antenna beamwidth  $\beta$  and the distance between satellite and ground  $R$ . Note that the radar beam is wide in range direction but narrow in azimuth. The generation of an image is facilitated by the forward motion of the airborne platform.

time in both range and azimuth direction. In range direction, echoes from the ground arrive progressively later from the near-range to the far-range edge of the swath. Objects at different ranges can be distinguished if their range separation is larger than half the transmitted pulse length. Hence, the range resolution of a SLAR system is defined by

$$\rho_R = \frac{c \cdot \tau}{2} [\text{m}] , \quad (2.2)$$

with  $c$  corresponding to the speed of light. The variable  $\rho_R$  in Eq. (2.2) is usually referred to as the slant range resolution of a SLAR system as it describes a SLAR's ability to distinguish objects at different (slant) distances from the radar (see "slant range direction" in Fig. 2.1). While the slant range parameter  $\rho_R$  is useful for many system design questions, remote sensing is often more interested in the ground range resolution  $\rho_G$ , which describes the ability to discriminate objects that are situated on the ground and is calculated from  $\rho_R$  via the local incidence angle  $\theta_i$ :

$$\rho_G = \frac{\rho_R}{\sin(\theta_i)} [\text{m}] . \quad (2.3)$$

Eq. (2.3) shows that the ground range resolution

$\rho_G$  is not constant across the swath and actually improves with distance from nadir (due to the increase of  $\theta_i$ ). This is opposite to the behavior of most optical systems for which the ground resolution degrades with increasing  $\theta_i$ .

In the along-track (or azimuth) direction, the ground is scanned by the movement of the radar along its track. In the case of SLAR systems, the azimuth resolution  $\rho_{Az}$  (the ability to discriminate objects in azimuth direction) is defined by the width of the antenna footprint in azimuth  $S_{Az}$ , which, in turn, is limited by the side length  $L_{Az}$  of the antenna in this direction. Hence, following Eq. (2.1), the azimuth resolution corresponds to

$$\rho_{Az} = S_{Az} \approx \frac{\lambda}{L_{Az}} R = \beta_{Az} \cdot R [\text{m}] . \quad (2.4)$$

Eq. (2.4) indicates that the azimuth resolution  $\rho_{Az}$  is linearly degrading with increasing distance between the sensor and the ground  $R$ . This has two important implications for SLAR systems: first, as  $R$  changes from the near-range to the far-range edge of the swath, the azimuth resolution of a SLAR is not

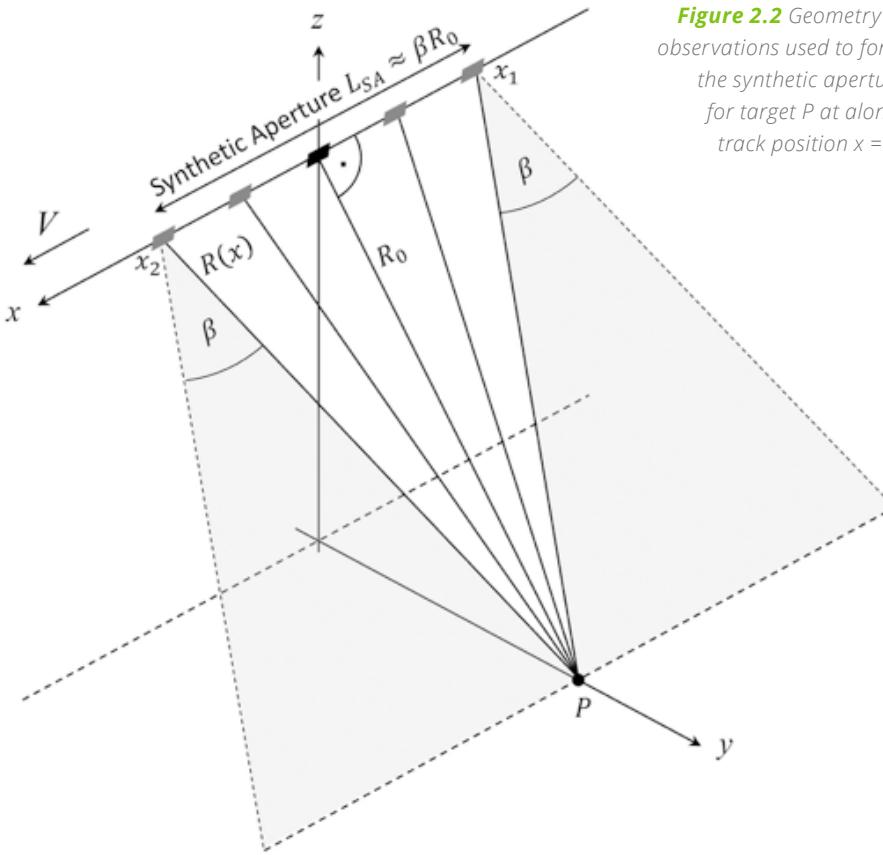
constant across range. Second, and more importantly, the dependence of  $\rho_{Az}$  on the distance to the ground  $R$  makes the application of SLAR systems on high-altitude or even spaceborne platforms highly impractical. To illustrate this point, assume a C-band SLAR system operating at  $\lambda = 0.03[\text{m}]$  and utilizing an antenna of  $L = 3[\text{m}]$  length. If operated from an aircraft at  $H = 3000[\text{m}]$  altitude and observing at a look angle of  $\theta_l = 30^\circ$ , this system will achieve an acceptable azimuth resolution of  $\rho_{Az} = 0.01 \cdot 3000 \cdot 2 = 60[\text{m}]$ . However, if the same system is operated from a spaceborne platform at  $H = 800[\text{km}]$ ,  $\rho_{Az}$  will degrade to  $\rho_{Az} = 16[\text{km}]$ , which is below the required system performance for most Earth observation applications.

A straightforward approach for keeping the system's azimuth resolution at an acceptable level even for spaceborne applications is to increase the length of the antenna used by the system until a desired value for  $\rho_{Az}$  is reached. Simple mathematics show, however, that this solution is not practical. Using the numbers from the previous example we find that an unreasonable antenna length of about  $L = 800[\text{m}]$  would be needed to achieve a  $\rho_{Az} = 60[\text{m}]$  resolution from space. An elegant and more practical solution for the azimuth resolution issue—the synthetic aperture principle—was developed in 1952 and will be introduced in Section 2.1.3.

Despite their resolution limitations, SLAR systems remain popular for many ground-based and airborne applications. This continued popularity is largely due to the simplicity of SLAR systems in both their system design and data processing demands.

### 2.1.3 SYNTHETIC APERTURE RADAR SENSORS

In 1952, Carl Wiley, an engineer with the Goodyear Aircraft Cooperation, made an essential discovery that provided a solution to the azimuth resolution problem plaguing existing SLAR technology (see Sec. 2.1.2). In technical terms, he observed that a one-to-one correspondence exists between the along-track coordinate (relative to a transmitted radar beam) of a reflecting object and the instantaneous Doppler shift of the signal reflected back to the radar by that object. He further postulated that a frequency analysis of the recorded signals could en-



**Figure 2.2** Geometry of observations used to form the synthetic aperture for target  $P$  at along-track position  $x = 0$ .

point  $x_2$  (last time  $P$  is observed). Once the radar data are acquired, a postprocessing approach is applied to combine all acquisitions between  $x_1$  and  $x_2$  and into a single dataset that looks like it was acquired with a much longer antenna. This longer (virtual) antenna is typically called the “synthetic aperture,” as it was synthesized from a number of acquisitions with shorter antennas. The length  $L_{SA}$  of this synthetic aperture can be calculated via

$$L_{SA} = \frac{\lambda}{V} \cdot R_0 \approx \beta \cdot R_0$$

and is equivalent to the footprint  $S$  illuminated by the (shorter) real antenna installed on the spacecraft (see Eq. (2.1)).

The dataset resulting from the aperture synthesis process is typically referred to as a SAR image and has much higher resolution than SLAR images acquired from the same distance. An example of a SAR image acquired by the European Space Agency’s (ESA’s) ERS-2 sensor is shown in **Figure 2.3**.

The aperture synthesis concept is the basis of all modern radar systems even though various modifications of the basic imaging concept are currently used to maximize either image resolution (Spotlight concept: Eineder et al. 2009, Lanari et al. 2001, Mittermayer et al. 1999) or image coverage (ScanSAR: Bamler and Eineder 1996, Bamler and Holzner 2004, Monti Guarnieri and Prati 1996). Modern spaceborne SAR sensors typically achieve ground resolutions between roughly 0.5 and 20 m, depending on their specific design. Recent developments in antenna design and image processing

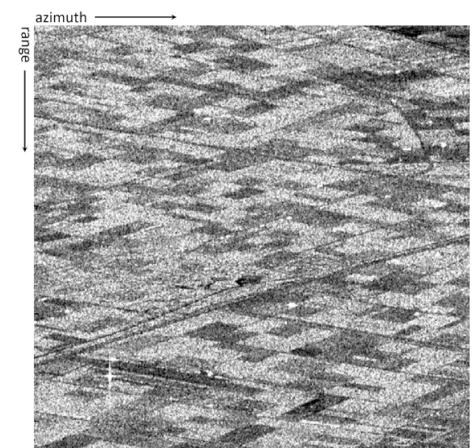
able a finer along-track resolution than that achieved by conventional SLAR systems. Wiley’s discovery—which was originally published under the name Doppler beam-sharpening but is often referred to as aperture synthesis—is the key element behind all modern high-resolution imaging radar systems. There is a large body of literature on the mathematical details of Wiley’s aperture synthesis solution to radar imaging. Readers interested in more technical information should refer to the excellent summary by Cumming and Wong (2005).

The following conceptual explanations of Wiley’s invention provide a good summary. The aperture synthesis principle essentially allows one to create (or “synthesize”) a much longer effective antenna (the so-called synthetic aperture) from a sequence of acquisitions made with a shorter antenna as it moves along its flight line. As antenna length is intrinsically linked to the resolution capabilities of a radar system (we know this from Eq. (2.1)), the much longer antenna synthesized by Wiley’s principle allows high-resolution imaging even from spaceborne platforms using

antenna hardware of a manageable size.

A simplified conceptual illustration of Wiley’s concept is shown in **Figure 2.2**. There, a radar antenna (indicated by a gray rectangle) of reasonably short length is moving at a velocity  $V$  along its flight path from the right to the left. While moving, it is constantly transmitting short radar pulses and receiving echoes returned from objects on the ground. Each radar pulse illuminates an instantaneous footprint of size  $S$  on the Earth surface. For spaceborne applications, the limited length  $L$  of the radar antenna (Eq. (2.1)) results in instantaneous footprints that typically measure several kilometers in size, resulting in the typical resolution limitation that plagues SLAR systems.

To apply Wiley’s aperture synthesis concept, we have to first ensure that an object  $P$  on the Earth surface is imaged by many consecutive radar pulses as the antenna beam sweeps across the ground. This requirement is indicated in **Figure 2.2** by several antenna positions that illuminate object  $P$  as the sensor moves from point  $x_1$  (first time object  $P$  is seen) to



**Figure 2.3** Example of a spaceborne SAR dataset acquired by ESA’s C-band sensor ERS-2.

techniques have made high-resolution imaging across large image swaths possible. These developments rely on the concept of digital beamforming (Gebert et al. 2009, Krieger & Moreira 2003, Younis et al. 2003) and have spawned new imaging modes such as Terrain Observation with Progressive Scans SAR (TOPSAR) (De Zan & Monti Guarnieri 2006) and SweepSAR (Freeman et al. 2009). For technical details on these techniques, please see the literature cited.

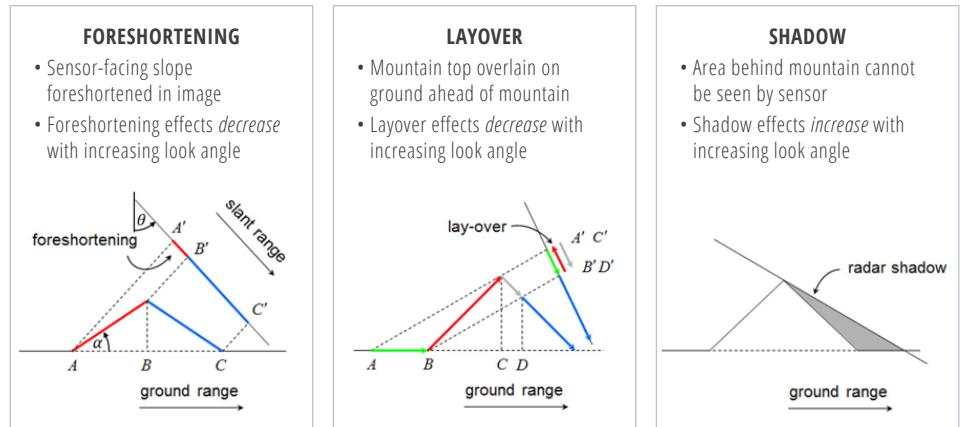
#### 2.1.4 GEOMETRIC PROPERTIES OF SAR DATA

Due to the oblique observation geometry inherent to all imaging radar systems, surface slopes and similar terrain features lead to geometric distortions in data acquired by SAR systems. The most relevant of these distortions are foreshortening, layover, and shadow. The origins and main characteristics of these distortions are summarized in **Figure 2.4**.

In side-looking viewing geometries, sensor-facing slopes appear foreshortened such that a symmetric mountain would appear in the radar image as if “leaning” towards the sensor. The geometric background of foreshortening is shown in **Figure 2.4(a)**, showing that the slope between points A and B will get foreshortened into the image area A'B'. The amount of foreshortening depends both on the system’s look angle  $\theta$  and on the slope angle  $\alpha$ , and reaches its maximum if  $\theta \rightarrow \alpha$ . In areas where  $\theta < \alpha$  (e.g., in areas of steep slopes combined with steep incidence angles), foreshortening turns into layover (see **Fig. 2.4(b)**). In layover situations, the tops of mountains are imaged ahead of their base (see projections of points B and C in **Fig. 2.4(b)**) and backscatter from mountain slopes will overlay with image information at closer and farther image ranges (see green, red, and gray areas in **Fig. 2.4(b)**). Both foreshortening and layover can be reduced if the look angle  $\theta$  is increased; however, larger  $\theta$  will produce more image shadow (**Fig. 2.4(c)**). Hence, topography-related image distortions cannot be entirely removed, and image acquisitions from more than one vantage point may be necessary to jointly minimize all three imaging effects.

#### 2.1.5 RADIOMETRIC PROPERTIES – THE SPECKLE EFFECT

Besides these geometric distortions, SAR images additionally are characterized by a somewhat grainy



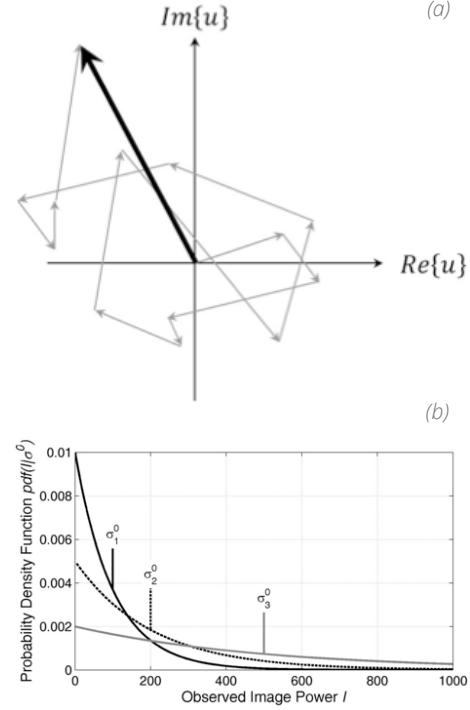
**Figure 2.4** Main geometric distortions on SAR images with their dependence on acquisition geometry: (a) foreshortening, (b) layover, and (c) shadow.

appearance that resembles “salt and pepper” noise. This noise-like pattern can be seen in **Figure 2.3** and is usually referred to as “speckle.” The speckle effect is inherent to all narrow-banded coherent imaging systems and is a result of interference from the many scattering echoes within a resolution cell.

In a medium-resolution SAR image, the scattering response from one resolution cell (of about  $10 \times 10$  [m] in size) is the coherent sum of thousands of individual scattering events, as shown in **Figure 2.5(a)**. Imagine the SAR system is imaging a homogeneous surface, such as a smooth meadow, and assume that the individual scattering events within one resolution cell (gray arrows in **Fig. 2.5(a)**) are all about equally strong. Due to their different positions within the resolution cell, the phase of the individual scatterers will vary randomly, such that the scattering response from one pixel is the summation of thousands of random vectors (black arrow in **Fig. 2.5(a)**). As the arrangement of scatterers in different resolution cells is not identical even for homogeneous targets, both the amplitude and phase of the summation vector (black arrow) will vary randomly from pixel to pixel, resulting in the typical grainy signature shown previously in **Figure 2.3**. If the number of individual scattering events is large, the distribution of intensities in a SAR image follows an exponential distribution of the form

$$pdf(I|\sigma^0) = \frac{1}{\sigma^0} \exp\left\{-\frac{I}{\sigma^0}\right\}, \quad (2.5)$$

where  $I = Re\{u\}^2 + Im\{u\}^2$  is the image intensity in



**Figure 2.5** (a) Speckle originating from the coherent summation of many individual scattering events within a resolution cell and (b) shape of the speckle pdf for images areas with different normalized radar cross sections  $\sigma^0$ .

a pixel. The distribution in Eq. (2.5) is often called speckle distribution and is a valid description for the noise patterns observed for homogeneous targets in medium-resolution SAR images.

Eq. (2.5) shows that the shape of the speckle distribution depends on the (true) normalized radar cross section  $\sigma^0$  of the observed target, such that

brighter image patches will show more intense noise. Here,  $\sigma^0$  describes the percentage of incoming radar energy that is scattered back to the sensor by an object on the ground. It is a normalized version of Eq. (2.6) discussed in the next section. The dependence of the speckle statistics on  $\sigma^0$  is visualized in **Figure 2.5(b)**, where the speckle distribution  $pdf(I|\sigma^0)$  is plotted for three different  $\sigma^0$  values. It can be seen that the speckle distribution becomes wider with increasing  $\sigma^0$  and starts to approximate a uniform distribution for very high  $\sigma^0$ .

Speckle noise is distinguished from most other noise sources, which are often constant throughout the image, by its dependence on image brightness. Multiplicative noise such as speckle is difficult to treat, as the true radar cross section  $\sigma^0$  of the target needs to be known to correctly model  $pdf(I|\sigma^0)$ . Hence, throughout the last decade, a lot of effort has been dedicated to the development of effective speckle filters, resulting in a wealth of different filtering methods. While the most relevant/well known of these filters are listed in **Table 2.1**, readers interested in this topic are referred to specialized literature such as Bruniquel & Lopes 1997, Ferretti et al. 2011, Huang et al. 2009, Lee et al. 1991, Lee et al. 1994, Lopez-Martinez & Pottier 2007, Novak & Burl 1990, and Sveinsson & Benediktsson 2003.

## 2.2 How SAR Images the World

SARs transmit microwave signals at an oblique angle and measure the backscattered (in the direction of the sensor) portion of this signal in order to analyze features on the surface. Mathematically, this (calibrated) measurement is described using the term Radar Cross Section (RCS)  $\sigma$ , which is defined as the ratio between the incident and received signal intensity:

$$\sigma = \frac{I_{\text{received}}}{I_{\text{incident}}} 4\pi R^2 \left[ \text{m}^2 \right]. \quad (2.6)$$

The RCS recorded by a SAR for a specific surface feature is not always straightforward to interpret, as it is influenced both by a range of scene characteristics as well as by the parameters of the imaging sensor.

The most important scene parameters driving RCS are surface roughness  $h_{\text{rough}}$  and the dielectric properties of the imaged object quantified by its complex relative dielectric constant  $\epsilon$ . While  $h_{\text{rough}}$  describes how much of the scattered radar energy is directed back to the sensor, the dielectric properties guide

whether or not (and how deep) signals may penetrate into the scattering surface. The fact that both of these parameters are a function of sensor wavelength (and to some degree signal polarization) explains why the characteristics of the sensor play a role when attempting to interpret the measured signature of real-life objects in a SAR image.

### 2.2.1 DIELECTRIC PROPERTIES AND PENETRATION DEPTH OF RADAR SIGNALS

The dielectric properties of a medium govern how a microwave signal of wavelength  $\lambda$  interacts with a scattering medium such as the Earth's surface or a vegetation canopy. These properties dictate how much of the incoming radiation scatters at the surface, how much signal penetrates into the medium, and how much of the energy gets lost to the medium through absorption. While a detailed explanation of microwave scattering processes is beyond the scope of this chapter, information is provided on how these processes change with sensor wavelength. This will provide the reader with the required background to interpret differences in the appearance of observed data from different SAR instruments. For a more detailed discussion on the interactions of microwaves with media, please refer to the excellent introducto-

Speckle Filters	Description	Related Publication(s)
<b>Change-preserving multi-temporal Speckle filter</b>	Filter for stacks of SAR images; reduces speckle while preserving changes in the time series (e.g., related to deforestation)	Quegan and Yu, 2001
<b>Lee filter</b>	Standard deviation-based (sigma) filter, filtering data based on statistics calculated from the data. Unlike a Gaussian or boxcar filter, the Lee filter and other similar sigma filters preserve image sharpness and detail while suppressing noise.	Lee, 1980
<b>Enhanced Lee filter</b>	The enhanced Lee filter is an adaptation of the Lee filter. Each pixel is put into one of three classes, which are treated as follows: <b>Homogeneous</b> : The pixel value is replaced by the average of the filter window. <b>Heterogeneous</b> : The pixel value is replaced by a weighted average. <b>Point target</b> : The pixel value is not changed.	Lopes et al., 1990
<b>Frost and enhanced Frost filters</b>	The Frost filter is an exponentially damped circularly symmetric filter that uses local statistics. The Enhanced Frost filter is an adaptation of the Frost filter. It classifies and filters pixels according to the logic explained in the row above.	Frost et al., 1982; Lopes et al., 1990
<b>Non-local means filters</b>	The basic idea behind non-local means filters is to provide an estimate of the clean image via a proper averaging of similar pixels or patches, found in the image. Essentially, the algorithm searches for image patches that resemble the area around the pixel to be filtered. Using some similarity criterion, these patches are found and averaged together to de-noise the image without losing resolution.	Buades et al., 2005; Chen et al., 2014; Di Martino et al., 2016; Martino et al., 2015

**Table 2.1** Summary of most relevant speckle filters with their properties and related publications.

ry book on microwave remote sensing by Iain Woodhouse (2006).

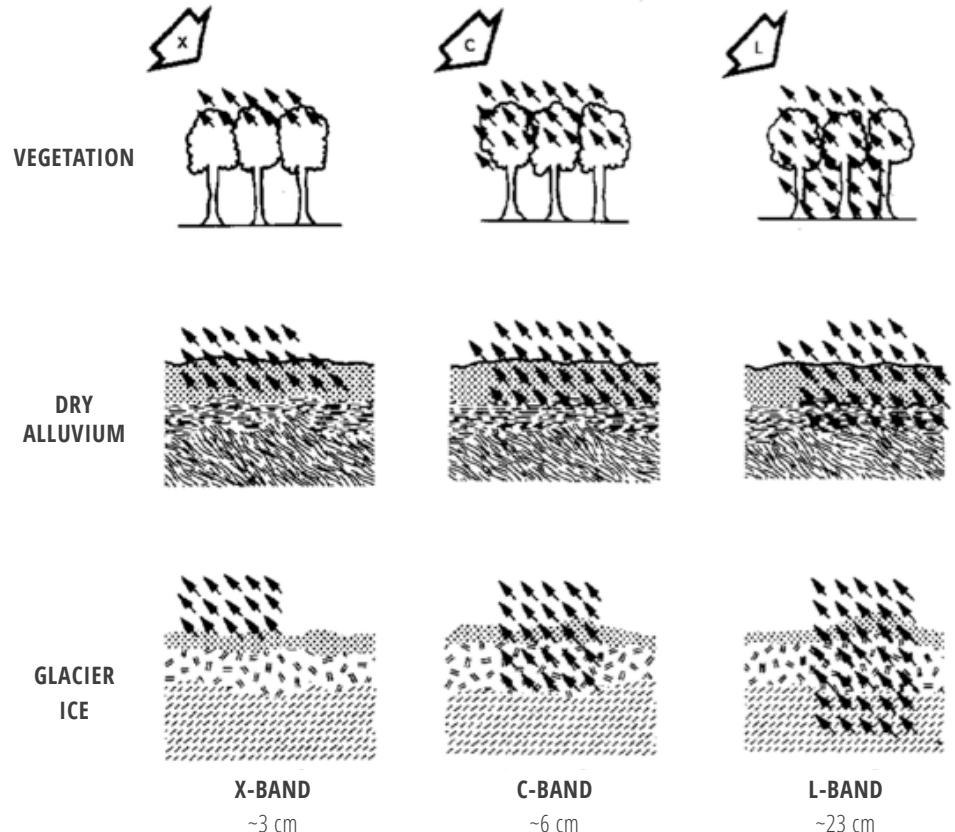
**Figure 2.6** provides a conceptual overview of the influence of sensor wavelength  $\lambda$  on signal penetration into a variety of surface types. The radar signals penetrate deeper as sensor wavelength increases. This is related to the dependence of the dielectric constant  $\epsilon_r$  on the incident wavelength, allowing for higher penetration at L-band than at C- or X-bands. For vegetated areas, this implies that X-band SAR sensors mostly scatter at the tops of tree canopies, while C- and L-band signals penetrate increasingly deeper into the vegetation volume. Hence, if vegetation parameters (e.g., vegetation structure, biomass, etc.) are to be characterized using SAR, longer wavelength systems should be used (see **Table 2.3** to identify sensors operating at longer wavelengths). Similarly, users interested in mapping inundation under forest canopies should select longer wavelength SAR sensors as their main data source.

In addition to sensor wavelength, the penetration depth of a SAR signal into a vegetation canopy is also influenced by the density of this canopy. For example, while C-band SAR data may “see” the ground underneath sparse boreal forests, C-band signals will not be able to fully penetrate the denser and layered canopy structure of rainforests.

The rule of increasing penetration with increasing sensor wavelength also holds true for bare surfaces such as alluvium soils or glacier ice; X-band signals scatter close to the surface, while C- and L-band data penetrate progressively deeper into the medium. To quantify penetration depths  $\delta_p$  into bare surfaces, information about the dielectric properties  $\epsilon_r$  of the medium is needed. If information on  $\epsilon_r$  is available,  $\delta_p$  can be approximated by

$$\delta_p \approx \lambda \sqrt{\epsilon_r' / (2\pi\epsilon_r'')} , \quad (2.7)$$

where  $\epsilon_r'$  is the real component and  $\epsilon_r''$  is the imaginary component of the complex relative dielectric constant. In addition to soil density and sensor wavelength,  $\epsilon_r'$  and  $\epsilon_r''$  are strongly dependent on the moisture content of the medium. **Figure 2.7(a)** shows an example of the dependence of dielectric properties on moisture content for loam soils com-



**Figure 2.6** SAR signal penetration by sensor wavelength  $\lambda$ .

posed of a mix of sand, silt, and clay ingredients. The dielectric properties are plotted as a function of soil moisture for several sensor wavelengths. It can be seen that both  $\epsilon_r'$  and  $\epsilon_r''$  increase with soil moisture, leading to a reduction of penetration depth according to Eq. (2.7). Also,  $\epsilon_r'$  and  $\epsilon_r''$  depend on sensor frequency  $f = c/\lambda$ . With increasing frequency (decreasing wavelength),  $\epsilon_r'$  reduces and  $\epsilon_r''$  increases such that penetration depth  $\delta_p$  is significantly larger for low-frequency (long wavelength) SARs. A plot of the dependence of penetration depth  $\delta_p$  on sensor wavelength  $\lambda$  is shown in **Figure 2.7(b)**. Penetration depth is approximated according to Eq. (2.7) for the soil type shown in **Figure 2.7(a)** and assuming a volumetric soil moisture of 0.35. A near-linear increase of penetration depth with increasing sensor wavelength can be observed.

## 2.2.2 SURFACE ROUGHNESS

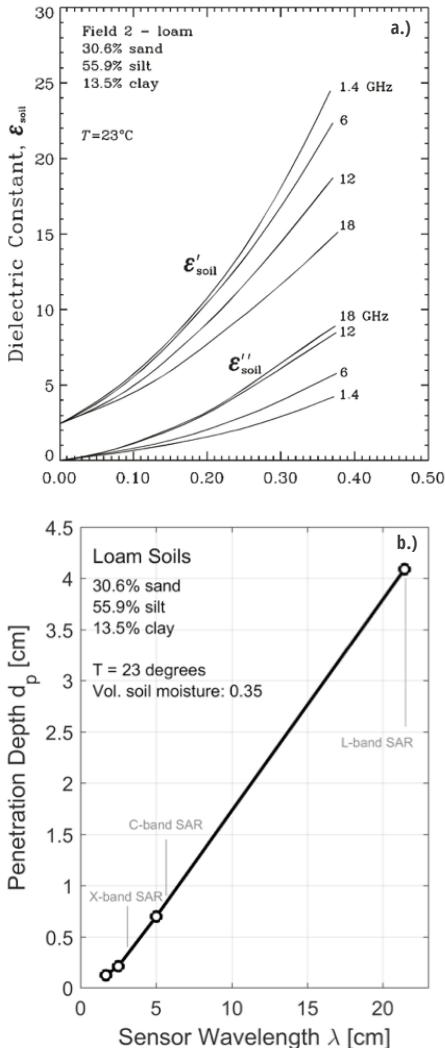
With few exceptions (dry snow, dry sandy soils), most bare or low-vegetation surfaces allow very little

penetration for microwave radiation (**Fig. 2.7(b)**) such that surface scattering dominates the measured radar response. In these cases, the roughness of the scattering surface is the main driver defining the observed RCS in a SAR scene.

For narrow-band imaging systems like SAR, whether a surface appears rough or not can only be decided with the observing sensor wavelength in mind. If the scale of roughness of a randomly rough surface is characterized using the standard deviation of the height deviation  $h$  from some mean height  $\bar{h}$  of the surface, then the question of how large  $h$  has to be for a surface to appear rough to an observing SAR system can be answered. According to the Fraunhofer criterion, a surface is defined as rough if the height deviations exceed the value  $h_{\text{rough}}$ , which is determined by Eq. (2.8):

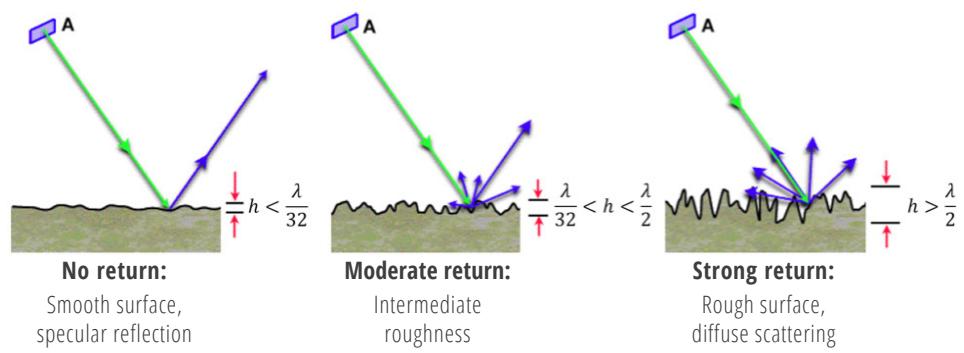
$$h_{\text{rough}} > \lambda / ((32 \cdot \cos\theta)) . \quad (2.8)$$

Note that the relationship in Eq. (2.8) depends on the signal wavelength  $\lambda$  and indicates that a surface with fixed height variations  $h$  may qualify as rough in



**Figure 2.7** (a) Relationship between soil moisture and dielectric constant and (b) dependence of penetration depth  $\delta_p$  on sensor wavelength  $\lambda$  for a fixed soil moisture.

X-band but possibly not in C- or L-bands. This concept of wavelength-dependent roughness is visualized in **Figure 2.8**, which shows increasing roughness conditions from left to right and identifies the transition from smooth (**Fig. 2.8(a)**) to intermediately rough (**Fig. 2.8(b)**) to rough surfaces (**Fig. 2.8(c)**) in accordance with the Fraunhofer criterion in Eq. (2.8). It can be seen that the amount of backscatter increases (length of blue arrows pointing toward the sensor) as roughness increases such that rough surfaces (at wavelength  $\lambda$ ) have higher RCS than intermediately rough or smooth surfaces. The wavelength dependence also means that a surface will look increasingly darker as wavelength increases from X-band ( $\lambda = 3.1 \text{ cm}$ ) through C-band ( $\lambda = 5.66 \text{ cm}$ ) to L-band ( $\lambda = 24 \text{ cm}$ ).



**Figure 2.8** Conceptual sketch of the dependence of surface roughness on the sensor wavelength  $\lambda$ : (a) smooth, (b) intermediate, and (c) rough.

As SAR is an active instrument with its own source of illumination, it is one of the few sensing instruments that allows one to fully control (and fully exploit) the polarization of the signal on both the transmit and the receive paths. Polarization describes the orientation of the plane of oscillation of a propagating signal. In linearly polarized systems, the orientation of this plane of oscillation is constant along the propagation path of the electromagnetic wave. In other systems, such as circular or elliptically polarized SARs, the orientation of the oscillation plane changes, describing geometric shapes such as ellipses or circles.

### 2.2.3 THE INFLUENCE OF SIGNAL POLARIZATION

The majority of today's SAR sensors are linearly polarized and transmit horizontally and/or vertically polarized wave forms. Many of the heritage SAR satellites carry single-polarized sensors, which support only one linear polarization. These sensors predominantly operate in HH- (horizontal polarization on transmit; horizontal polarization on receive) or VV-polarization (vertical transmit; vertical receive), while single-polarized sensors transmitting one linear polarization and receiving the other (e.g., HV (horizontal transmit; vertical receive)) are rare in practice.

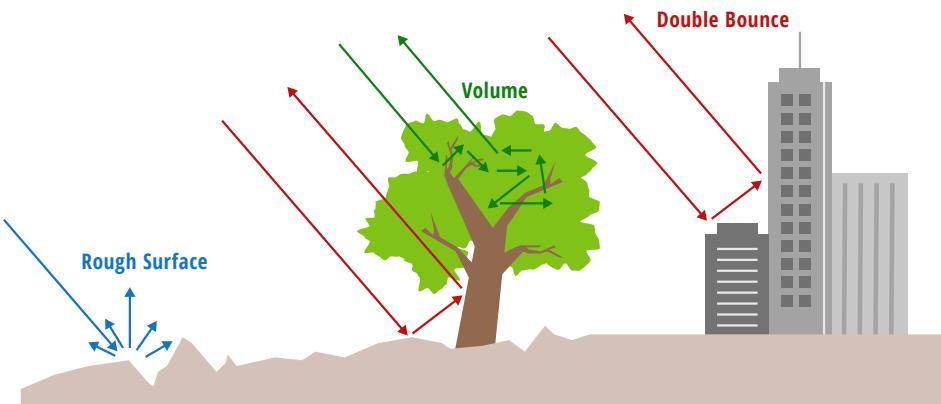
More recent sensors provide either dual-polarization or quad-polarization capabilities. In the latter,

the sensor alternates between transmitting H- and V-polarized waveforms and receiving both H and V simultaneously, providing HH-, HV-, VH-, and VV-polarized imagery.

Knowing the polarization from which a SAR image was acquired is important, as signals at different polarizations interact differently with objects on the ground, affecting the recorded radar brightness in a specific polarization channel. While the details of polarimetric scattering are beyond the scope of this chapter, the following paragraph provides rules of thumb that should aid in the interpretation of polarimetric SAR data.

For simplicity, it is assumed that a natural scene can be described as a combination of three types of scatterers: (1) rough surface scatterers, (2) double-bounce scatterers, and (3) volume scatterers. The nature of these scattering types is illustrated in **Figure 2.9**. The category of surface scatterers (shown in blue in **Fig. 2.9**) is made up of low-vegetation fields and bare soils, as well as roads and other paved surfaces. Double-bounce scatterers (red in **Fig. 2.9**) include buildings, tree trunks, light poles, and other vertical structures that deflect an initial first forward reflection back to the sensor. Finally, vegetation canopies belong to the category of volume scatterers (green in **Fig. 2.9**) as the signals bounce multiple times as they propagate through the vegetation structure.

It turns out that these scattering types do not contribute to all polarimetric channels equally. Instead, each polarimetric channel "prefers" certain scattering types such that the scattering power  $|S|$  in the indi-



**Figure 2.9** Schematic sketch of the three main scattering types considered for SAR data.

vidual polarimetric channels follows the following general scheme shown in **Table 2.2**.

#### RELATIVE SCATTERING STRENGTH BY POLARIZATION:

<b>Rough Surface Scattering</b>	$ S_{\text{W}}  >  S_{\text{HH}}  >  S_{\text{HV}}  \text{ or }  S_{\text{VH}} $
<b>Double Bounce Scattering</b>	$ S_{\text{HH}}  >  S_{\text{W}}  >  S_{\text{HV}}  \text{ or }  S_{\text{VH}} $
<b>Volume Scattering</b>	Main source of $ S_{\text{HV}} $ and $ S_{\text{VH}} $

**Table 2.2** Relative scattering strength by polarization

These general rules should help when comparing the RCS in different polarimetric channels. They can be applied to perform an automatic classification of scattering types if data with all relevant polarizations (i.e., quad-polarization data) are available. For more information on polarimetric SAR and polarimetric SAR data analysis, see Pottier & Lee 2009 and Van Zyl 2011.

An example of the information contained in quad-polarization SAR data is shown in **Figure 2.10**. There, the polarimetric scattering power of  $|S_{\text{HH}}|$ ,  $|S_{\text{WV}}|$ , and  $|S_{\text{HV}}|$  are presented in **Figures 2.10(a), (b), and (c)**, respectively, for an ALOS PALSAR scene over Niamey, Niger. According to the rule above, strong scattering in  $|S_{\text{HH}}|$  indicates a predominance of double-bounce scattering (e.g., stemmy vegetation, manmade structures), while strong  $|S_{\text{WV}}|$  relates to rough surface scattering (e.g., bare ground, water), and spatial variations in  $|S_{\text{HV}}|$  indicate the distribution of volume scatterers (e.g., vegetation and high-penetration soil types such as sand or other dry porous soils) across the scene. To enhance the visibility of differences between the channels, the HH, WV, and

HV information is often combined into a single RGB image, with  $|S_{\text{HH}}|$  in red,  $|S_{\text{WV}}|$  assigned to blue, and  $|S_{\text{HV}}|$  in green. Such an RGB image composite for the scene over Niamey is shown in **Figure 2.10(d)**. Extensive red areas can be seen in some urban districts (buildings) and some agricultural zones (stemmy vegetation). A patch of green can be seen to the south of Niamey, presumably relating to higher penetration sandy soils and the volumetric scattering on inclusion within the sand body. Most other areas have a tinge of blue, indicating bare soils.

## 2.3 Historic, Current, and Future SAR Sensors

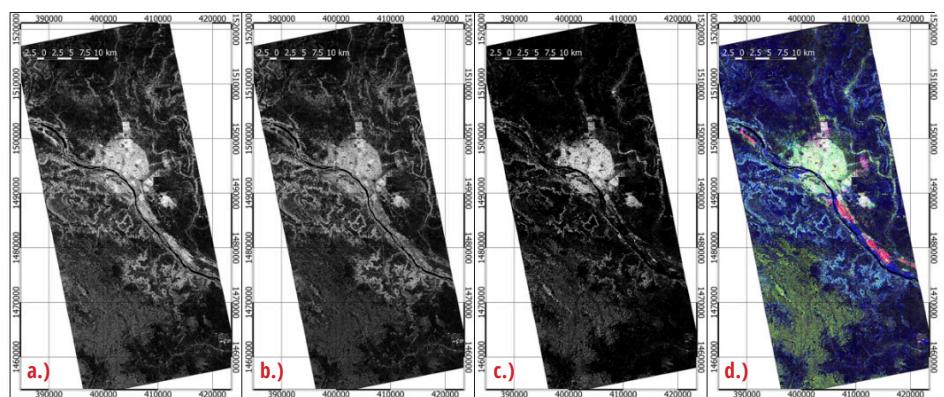
Amazingly, spaceborne SAR sensors have been around for more than 40 years. The first SAR was

launched on June 28, 1978, on board NASA's Seasat satellite, a spaceborne platform aimed at monitoring oceanographic phenomena. As part of its sensor suite, Seasat carried an HH-polarized L-band SAR that was mounted at a fixed angle to observe global surface wave fields and polar sea ice conditions. Even though Seasat's SAR operated for only 106 days (a short circuit in the satellite's electrical system occurred on October 10, 1979), the mission was deemed an extensive success, demonstrating a SAR capability both ocean and land surface observation (Fu & Holt 1982).

Since the days of Seasat, SAR remote sensing has come a long way. Starting with ERS-1 in 1991, several SAR sensors with ever-improving imaging characteristics have been launched by an international community of satellite providers, collectively ensuring continuous coverage of the Earth with SAR data. Unfortunately, this international constellation of SAR systems comes with a downside. The SAR satellites launched by the various agencies vary widely in their sensor configurations such that data from different sensors are not always directly comparable (see **Sec. 2.2**). **Section 2.3.1** outlines the main differences between different sensors in order to assist new users in choosing the correct SAR data for an intended application.

### 2.3.1 SAR SENSOR WAVELENGTHS

SAR sensors transmit energy in one of the micro-



**Figure 2.10** Fully-polarimetric L-band SAR scenes from the ALOS PALSAR sensor over Niamey, Niger: (a)  $|S_{\text{HH}}|$ , (b)  $|S_{\text{WV}}|$ , and (c)  $|S_{\text{HV}}|$  scattering powers. An RGB color combination of these channels is shown in (d).

wave frequency bands shown in **Table 2.3**. Roughly speaking, radar systems use frequencies from 1 to 90 GHz, a spectral range that is subdivided into the frequency bands shown in the first column of **Table 2.3**. These bands were initially defined according to the different equipment needed to generate and detect signals at these particular wavelengths, but now, they can be understood as the equivalent of colors in the visual range. As microwave remote sensing was developed largely during World War II, a rather cryptic naming convention was used to disguise the meaning of microwave bands from the enemy. Unfortunately, this letter-based naming scheme (Ka-band to P-band) was never modified and may lead to confusion among new users of SAR.

Not all of the microwave bands shown in **Table 2.3** are used for SAR remote sensing. While some experimental airborne Ka- and Ku-band SAR systems exist, civilian spaceborne sensors have been exclusively using the lower frequency bands ranging from X- to P-band (blue shaded region in **Table 2.3**).

As explained in **Section 2.2**, the wavelength of a SAR sensor is intrinsically linked to the penetration

capabilities of the transmitted microwave signal, such that longer wavelength signals (e.g., signals at L- and P-band) penetrate deeper into vegetation canopies and soils. Hence, the applications supported by a SAR sensor depend on the SAR frequency band used.

**Table 2.3** summarizes typical applications of SAR as a function of frequency band. It shows that sensors at X-band are predominantly used for urban and infrastructure monitoring. Due to the higher resolution capabilities of X-band radars, sensors at this frequency find broad application in surveillance and tracking and are also often used in the monitoring of industry installations. Due to the limited penetration into vegetation covers, X-band is rarely used for characterizing forest canopies for monitoring activity underneath vegetation.

With the predominate number of legacy systems operating at this frequency range, C-band sensors have been the workhorse of SAR monitoring over the last 30 years. With moderate- to high-resolution capabilities and increased vegetation penetration, C-band data can be seen as a good compromise between X-band and the longer wavelength L-band sensor classes. Compared to X-band SARs, C-band sensors typically allow

for wider swath imaging, lending themselves to regional- and global-scale applications. While C-band has improved canopy penetration capabilities, its signals will typically not penetrate all the way through a vegetation layer. Especially in regions with denser vegetation, C-band is of limited use for analyzing activity underneath canopy layers.

While S-band SAR sensors were rarely used in Earth observations in the past, this frequency will have increased usage in the near future. NovaSAR-S, an S-band SAR sensor, was launched in September 2018, and while access to NovaSAR-S data may be limited, it will provide some medium-resolution SAR data to explore the performance of S-band data for applications such as hazard monitoring, crop monitoring, forest monitoring (temperate and rainforests), as well as land-use mapping. More interesting to most users will be the upcoming NASA ISRO SAR satellite, NISAR. In addition to an L-band radar, NISAR will carry a fully polarimetric S-band SAR. While NISAR's S-band coverage will likely not be global, all data will be freely and openly available to the SAR science and applications community.

While most of the historic SAR systems operated

BAND	FREQUENCY	WAVELENGTH	TYPICAL APPLICATION
Ka	27 – 40 GHz	1.1 – 0.8 cm	Rarely used for SAR (airport surveillance)
K	18 – 27 GHz	1.7 – 1.1 cm	Rarely used ( $H_2O$ absorption)
Ku	12 – 18 GHz	2.4 – 1.7 cm	Rarely used for SAR (satellite altimetry)
X	8 – 12 GHz	3.8 – 2.4 cm	High-resolution SAR (urban monitoring; ice and snow, little penetration into vegetation cover; fast coherence decay in vegetated areas)
C	4 – 8 GHz	7.5 – 3.8 cm	SAR workhorse (global mapping; change detection; monitoring of areas with low to moderate vegetation; improved penetration; higher coherence); ice, ocean, maritime navigation
S	2 – 4 GHz	15 – 7.5 cm	Little but increasing use for SAR-based Earth observation; agriculture monitoring (NISAR will carry an S-band channel; expands C-band applications to higher vegetation density)
L	1 – 2 GHz	30 – 15 cm	Medium resolution SAR (Geophysical monitoring; biomass and vegetation mapping; high penetration; InSAR)
P	0.3 – 1 GHz	100 – 30 cm	Biomass. First P-band spaceborne SAR will be launched ~2020; vegetation mapping and assessment. Experimental SAR.

**Table 2.3** Designation of microwave bands. Spaceborne SARs typically operate in the frequency bands shaded in green. Note: This table uses standard terminology common to the radar community. This nomenclature is not identical to ones used by other disciplines. For instance, P-band is often referred to as UHF band. Also note that the actual frequencies allocated for radar use by the International Telecommunications Union are narrower bands within these broad classifications.

in C-band, the family of future SAR sensors is largely focused on the L-band frequency range. While L-band SARs do not provide the high-resolution capabilities of shorter wavelength SARs, their ability to penetrate vegetation holds a number of advantages for Earth observation. With a higher likelihood of seeing the ground, L-band SARs are useful for mapping activity underneath canopies such as flooding. Due to the high penetration into vegetation covers, L-band SAR also lends itself well to characterizing canopy structure, especially in denser forests. Finally, the higher canopy penetration is also advantageous for users of Interferometric SAR (InSAR), achieving higher interferometric coherence (see [Sec. 2.6.2.](#)) and better deformation tracking capabilities.

P-band SAR sensors are currently under development. Spaceborne applications at this frequency are hampered by ionospheric distortions, and only recent developments in ionospheric correction (Belcher 2008, Belcher and Rogers 2009, Gomba et al. 2016, Jehle et al. 2010, Jehle et al. 2009, Kim et al. 2011, Meyer et al. 2006, Meyer & Nicoll 2008a, Meyer 2011, Meyer et al. 2016, Meyer & Nicoll 2008b, Pi et al. 2012) have allowed spaceborne P-band SAR missions to go forward. The first spaceborne P-band SAR—ESA's Biomass mission—is planned to launch in 2021 and will focus on mapping the status and the dynamics of Earth's forests, as represented by the distribution of forest biomass and its changes.

### 2.3.2 A SUMMARY OF RELEVANT SAR PLATFORMS WITH THEIR PROPERTIES

A list of the most relevant past, current, and future SAR platforms is provided in [Table 2.4](#). The sensors are sorted by their period of performance. For each instrument, the sensor wavelength, supported polarization modes, resolution and size of image products, repeat cycle, and means of data access are listed. This quick guide may be useful in selecting appropriate sensors for a specific application.

## 2.4 SAR Data Types and Their Applications

[Table 2.4](#) showcases the diversity of SAR sensors that have been launched since the beginning of the

spaceborne SAR era in 1979. While the deep, multitemporal archive provided by these sensors is of tremendous value for users interested in long-term Earth observation, SAR data products from these various platforms are, unfortunately, plagued by inconsistent naming conventions and come in a range of data types and formats, which can cause confusion even for more senior users of SAR. The following sections attempt to summarize and categorize the various data types and nomenclatures used by different data providers to provide guidance to users new to this tremendously useful Earth observation asset. For every data type, typical naming conventions are listed and appropriate open source software tools are introduced. Also summarized are the main applications associated with a specific data type. A concise summary of all information provided can also be found in [Table 2.5](#).

The variety of data types provided by a SAR system are related to the diverse flavors of information that are captured in every SAR acquisition. In every pixel, a SAR provides measurements of signal amplitude, phase, and polarization, all of which are related to different physical quantities of the observed ground. As extracting and utilizing these different information layers is often not straightforward—and as amplitude, phase, and polarization information is often relevant to different user communities—SAR data providers have decided to offer their imagery up in a range of different processing levels, each progressively simplified and tailored to emphasize different components of the SAR information space.

### 2.4.1 SAR RAW DATA

**General Description:** As the purest of all SAR processing levels, RAW data corresponds to the decoded but otherwise unfocused (i.e., Wiley's aperture synthesis processing has not yet been applied; [Sec. 2.1.3](#)) raw observables made by a SAR sensor. Unlike optical sensors, visualizing raw SAR data does not provide much useful information about the scene. Only after aperture synthesis processing is the RAW data transformed into an interpretable image.

**Applications:** RAW data products are the basis for all higher level SAR processing levels, and as such, RAW is an essential data type in every SAR data

archive. Outside of the user community interested in SAR data processing, however, RAW products find very little use. Interestingly, while RAW data are an essential product for every SAR sensor, not every satellite operator has decided to make his RAW data products available to the community. For some sensors, satellite data security laws prohibit the publication of RAW data products. Mostly, however, sensor providers elect to hide RAW data to retain proprietary information about their SAR processing routines.

**Naming Convention:** RAW products are categorized as processing Level 0 data, a processing level typically abbreviated as L0. An exception to this abbreviation exists for data from the ALOS PALSAR sensor, which uses L1.0 when referring to their RAW data products.

**Open Source Software Tools:** There are a number of open source software tools that can be used to read and manipulate (focus) RAW SAR data products. These include the following:

- **InSAR Scientific Computing Environment (ISCE)**—Developed by Jet Propulsion Laboratory (JPL)/Stanford/Caltech. More information and download: <https://winsar.unavco.org/isce.html>.
- **GMTSAR**—Developed by Scripps Institution of Oceanography. More information and download: <http://topex.ucsd.edu/gmtsar/>.
- **Repeat Orbit Interferometry Package (ROI\_PAC)**—Developed by JPL/Caltech. More information and download: [https://winsar.unavco.org/portal/wiki/ROI\\_PAC](https://winsar.unavco.org/portal/wiki/ROI_PAC).
- **Delft Object-oriented Radar Interferometric Software (DORIS)**—Developed by Delft University of Technology. More information and download: <http://doris.tudelft.nl/>.

Note that most of these tools are focused on the SAR expert community and therefore require a considerable amount of expertise to use correctly. Furthermore, these tools predominantly reside on Linux or UNIX operating systems and use command-line methods as the means of user interaction.

### 2.4.2 SINGLE LOOK COMPLEX IMAGE

**General Description:** Single Look Complex (SLC) images are fully focused SAR data that are

SENSOR	LIFETIME	WAVELENGTH/ FREQUENCY	POLARIZATION	RESOLUTION	FRAME SIZE	REPEAT CYCLE	ACCESS
<b>Seasat</b>	1978	L-band $\lambda = 24.6\text{cm}$	HH	Az: 25m Rg: 25m	100km	-	Free & open
<b>ERS-1</b>	1991-2001	C-band $\lambda = 05.6\text{cm}$	W	Az: 6-30m Rg: 26m	100km	35 days	Restrained
<b>JERS-1</b>	1995-1998	L-band $\lambda = 24.6\text{cm}$	HH	Az: 18m Rg: 18m	75km	44 days	Restrained
<b>ERS-2</b>	1995-2011	C-band $\lambda = 05.6\text{cm}$	W	Az: 6-30m Rg: 26m	100km	35 days	Restrained
<b>ENVISAT</b>	2002-2012	C-band $\lambda = 05.6\text{cm}$	HH, W, W/HH, HH/HV, W/VH	Az: 28m Rg: 28m	100km	35 days	Restrained
<b>ALOS-1</b>	2006-2011	L-band $\lambda = 24.6\text{cm}$	FBS: HH, W FBD: HH/HV, HH/VH PLR: HH/HV /VH /W ScanSAR: HH, W	FBS: 10x10m FBD: 20x10m PLR: 30x10m ScanSAR: 100m	FBS: 70km FBD: 70km PLR: 30km ScanSAR: 250-350km	46 days	Free & open
<b>Radarsat-1</b>	1995-2013	C-band $\lambda = 05.6\text{cm}$	HH	Standard: 25x28m Fine: 9x9m Wide1: 35x28m Wide2: 35x28m ScanSAR: 50x50-100x100m	Standard: 100km Fine: 45km Wide1: 165km Wide2: 150km ScanSAR: 305-510km	24 days	1995-2008: Restrained 2008-2013: Commercial
<b>TerraSAR-X</b> <b>TanDEM-X</b>	2007-2010-	X-band $\lambda = 03.5\text{cm}$	Single: HH, W Dual: HH/W, HH/HV, W/VH Twin: HH/W, HH/VH, W/VH	Spotlight: 0.2x1.0-1.7x3.5m Stripmap: 3x3m ScanSAR: 18-40m	Spotlight: 3-10km Stripmap: 50x30km ScanSAR: 150x100-200x200km	11 days	Application-dependent; restricted scientific, commercial
<b>Radarsat-2</b>	2007-	C-band $\lambda = 05.6\text{cm}$	Single: HH, W, HV, VH Dual: HH/HV, WV/VH Quad: HH/HV/VH/WV	Spotlight: ~1.5m Stripmap: ~3x3-25x25m ScanSAR: 35x35-100x100m	Spotlight: 18x8km Stripmap: 20-170m ScanSAR: 300x300- 500x500km	24 days	Commercial
<b>COSMO-SkyMed</b>	2007-	X-band $\lambda = 03.5\text{cm}$	Single: HH, W, HV, VH Dual: HH/HV, HH/WV, W/VH	Spotlight: ≤1m Stripmap: 3-15m ScanSAR: 30-100m	Spotlight: 10x10km Stripmap: 40x40km ScanSAR: 100x100 - 200x200km	Satellite: 16 days Constellation: ~hrs	Commercial; limited proposal-based scientific
<b>ALOS-2</b> <b>PALSAR-2</b>	2014-	L-band $\lambda = 24.6\text{cm}$	Single: HH, W, HV, VH Dual: HH/HV, WV/VH Quad: HH/HV/VH/WV	Spotlight: 1x3m Stripmap: 3-10m ScanSAR: 25-100m	Spotlight: 25x25km Stripmap: 55x70-70x70km ScanSAR: 355x355km	14 days	Commercial; limited proposal-based scientific
<b>Sentinel-1</b>	2014-	C-band $\lambda = 05.6\text{cm}$	Single: HH, W Dual: HH/HV, W/VH	Stripmap: 5x5m Interferometric Wide Swath (IW): 5x20m Extra Wide Swath (EW): 20-40m	Stripmap: 375km IW: 250km EW: 400km	Satellite: 12 days Constellation: 6 days	Free & open
<b>SAOCOM</b>	2018-	L-band $\lambda = 24.6\text{cm}$	Single: HH, W Dual: HH/HV, W/VH Quad: HH/HV/VH/WV	Stripmap: 10x10m TopSAR: 100x100m	Stripmap: >65km TopSAR: 320km	Satellite: 16 days Constellation: 8 days	TBD
<b>PAZ SAR</b>	2018-	X-band $\lambda = 03.5\text{cm}$	*See TerraSAR/TanDEM-x	*See TerraSAR/TanDEM-x	*See TerraSAR/TanDEM-x	11 days	Commercial
<b>RCM</b>	2019	C-band $\lambda = 05.6\text{cm}$	Single: HH, W, VH, HV Dual: HH/HV, WV/VH, HH/WV Compact Quad	Very high, high, medium, and low-res modes (3-100m)	20x20-500x500km	Satellite: 12 days Constellation: ~hrs	TBD
<b>NISAR</b>	2021	L-band $\lambda = 24.6\text{cm}$	Single: HH, W, VH, HV Dual: HH/HV, WV/VH, HH/WV Quad	3-20m (mode dependent)	250km	12 days	Free & open
<b>BIOMASS</b>	2021	P-band $\lambda = 70.0\text{cm}$	Quad	≤60x50m	160km	17 days	Free & open
<b>TanDEM-L</b>	2023	L-band $\lambda = 24.6\text{cm}$	Single, dual, quad modes	12x12m	350km	Satellite: 16 days Constellation: 8 days	Free & open

**Table 2.4** List of past, current and upcoming spaceborne SAR sensors with their properties.

Sensor	Format	Product Name	Product Files	Processing Level	Open Source Tools	Applications
<b>CURRENT SPACEBORNE SENSORS</b>						
<b>Seasat</b>	HDF5	L1 HDF5 Image	h5, xml, kml, jpg, qc_report	Amplitude	ASF MapReady, QGIS	Visualization; GIS-compatible
	GeoTIFF	L1 GeoTIFF	tif, xml, kml, jpg, qc_report	Geocoded amplitude	QGIS; graphics software	Visualization; GIS-compatible
<b>ERS-1&amp;2 Envisat Radarsat-1 JERS-1</b>	EOS	L0	D,L,P, kml, jpg	Raw	N/A	Production of higher-level products
		L1 Image		Amplitude	ASF MapReady; S1TBX	Visualization, mapping, change detection
<b>ALOS-1</b>	CEOS	L1.0	LED, IMG, VOL, TRL	Raw	N/A	Production of higher-level products
		L1.1 Complex		SLC	SNAP; ROI_PAC; DORIS; PolSARpro; GMTSAR	Interferometry
		L1.5		Amplitude	ASF MapReady; S1TBX; PolSARpro	Visualization, mapping, change detection
<b>TerraSAR-X TanDEM-X</b>	COSAR format	L1 SSC (Single Look Slant Range Complex)		SLC	SNAP; ROI_PAC; DORIS; PolSARpro; GMTSAR	Interferometry
	GeoTIFF	L1 MGD (Multi Look Ground Range Detected)		Amplitude	ASF MapReady; SNAP; PolSARpro	Visualization, mapping, change detection
	GeoTIFF	L1 GEC (Geocoded Ellipsoid Corrected)		Amplitude	ASF MapReady; SNAP; PolSARpro	Visualization, mapping, change detection
	GeoTIFF	L1 EEC (Enhanced ellipsoid corrected)		Amplitude	ASF MapReady; SNAP; PolSARpro	Visualization, mapping, change detection
<b>Radarsat-2</b>	GeoTIFF or NITF 2.1 with XML	L1 SLC		SLC	SNAP; ROI_PAC; DORIS; PolSARpro; GMTSAR	Interferometry
		L1 Ground Range (SGX; SGF; SCN; SCW; SCF; SCS)		Amplitude	SNAP; PolSARpro	Visualization, mapping, change detection
		L1 Geocorrected (SSG; SPG)		Amplitude	SNAP; PolSARpro	Visualization, mapping, change detection
<b>COSMO-SkyMed</b>	HDF5	L0 RAW		Raw		Production of higher-level products
		L1A		SLC	SNAP; ROI_PAC; DORIS; PolSARpro; GMTSAR	Interferometry
		L1B MDG (Multi-look Detected Ground Range)		Amplitude	SNAP; PolSARpro	Visualization, mapping, change detection
		L1C GEC		Amplitude	SNAP; PolSARpro	Visualization, mapping, change detection
		L1D GTC (Geocoded Terrain Corrected)		Amplitude	SNAP; PolSARpro	Visualization, mapping, change detection
		Various higher-level products				
<b>ALOS-2 PALSAR-2</b>		L1.1 SLC		SLC	SNAP; ROI_PAC; DORIS; PolSARpro; GMTSAR	Interferometry
		L1.5 (slant-range detected)		Amplitude	SNAP; PolSARpro	Visualization, mapping, change detection
		L2.1 GTC		Geocoded amplitude	SNAP; PolSARpro	Visualization, mapping, change detection
		L3.1 (Quality corrected L1.5)		Enhanced amplitude	SNAP; PolSARpro	Visualization, mapping, change detection
<b>Sentinel-1</b>	SAFE	L0 raw data	tiff, xml, xsd, kml, html, png, pdf, safe	Raw	N/A	Production of higher-level products
	GeoTIFF	L1 SLC		SLC	S1TBX; ROI_PAC; DORIS; PolSARpro	Interferometry
	GeoTIFF	L1 Detected High-Res Single- & Dual-Pol		Georeferenced Amplitude	ASF MapReady; Google Earth Engine; S1TBX; PolSARpro	Visualization, mapping, change detection
	GeoTIFF	L1 Detected Single- & Dual-Pol		Georeferenced Amplitude	ASF MapReady; Google Earth Engine; S1TBX; PolSARpro	Visualization, mapping, change detection

**Table 2.5** Current and upcoming spaceborne SAR sensors with their properties.

SENSOR	FORMAT	PRODUCT NAME	PRODUCT FILES	PROCESSING LEVEL	OPEN SOURCE TOOLS	APPLICATIONS
<b>RECENT AND FUTURE SPACEBORNE SENSORS</b>						
<b>SAOCOM</b> <b>PAZ SAR</b> <b>RCM</b> <b>NISAR</b> <b>BIOMASS</b> <b>TanDEM-L</b>					Formats and data types yet to be determined	
<b>AIRBORNE SENSORS</b>						
<b>UAVSAR</b> PolSAR	UAVSAR	Ground Projected Complex [full-res; 3x3; 5x5]	grd, ann	Georeferenced Amplitude	ASF MapReady; PolSARpro	Visualization
		Multi-Look Complex	mic, ann	MLC	ASF MapReady; PolSARpro	Polarimetry
		Compressed Stokes Matrix	dat, ann	AIRSAR compressed stokes matrix	ASF MapReady; PolSARpro	Polarimetry
	GeoTIFF	Pauli Decomposition	tif	MLC pol. decomposition	QGIS; graphics software	Visualization, GIS compatible
	KMZ	Google Earth KMZ	kmz	KML compressed	Google Earth	Visualization
<b>UAVSAR</b> InSAR	UAVSAR	Amplitude	amp1, amp2, ann	Amplitude	ASF MapReady; PolSARpro	Visualization
		Ground Projected Amplitude	amp1.grd, amp2.grd, hgt.grd, ann	Georeferenced Amplitude	ASF MapReady; PolSARpro	Visualization
		Interferogram	int, unw, cor, ann	Interferogram	ASF MapReady; PolSARpro	
		Ground Projected Interferogram	cor.grd, hgt.grd, int.grd, unw.grd, ann	Interferogram	ASF MapReady; PolSARpro	
	KMZ	Google Earth KMZ	amp.kmz, cor.kmz, hgt.kmz, int.kmz, osr.kmz, unw.kmz		Google Earth	Visualization

**Table 2.5, continued**

provided at the full native resolution (single look) with both amplitude and phase information stored in each (complex) pixel. SLC products are typically provided in the original slant-range observation geometry (**Fig. 2.1**) and are therefore not geocoded or terrain-corrected. In contrast to most optical sensors, the native resolution of SAR sensors is often significantly different along the azimuth and range image directions. Hence, SLC images often look geometrically distorted when viewed in image processing software. While SLCs usually come with radiometric calibration factors already applied, speckle noise remains unmitigated in these full-resolution products. For polarimetric data, separate SLC products are provided for each polarimetric channel.

**Applications:** The phase information stored in SLC products is an essential prerequisite for InSAR processing (**Sec. 2.6.2**), which is used for mapping surface topography or surface deformation. In addition to its use in InSAR, SLCs are also the basis for

higher level image products such as amplitude images, polarimetric products, and geocoded images. See **Table 2.5** for more information.

**Naming Convention:** In the SAR world, SLC products are categorized as processing Level 1 data, typically abbreviated as L1 or L1 SLC data. An exception to this abbreviation exists for data from the ALOS PALSAR sensor, which uses L1.1 when referring to its SLC products.

**Open Source Software Tools:** SLC data can be read and further processed by a series of open source software tools. These include (but are not limited to) the following:

- All previously named RAW data tools (ISCE, ROI\_PAC, GMTSAR, DORIS)
- **MapReady:** Developed by the Alaska Satellite Facility. More information and download: <https://www.asf.alaska.edu/data-tools/mapready/>.
- **Sentinel Application Platform**

**(SNAP):** Developed by ESA. More information and download: <http://step.esa.int/main/download/>.

While all of these tools are capable of processing SLC products, not all tools work with all sensors. Users should refer to the links above to ensure that their data can be successfully processed with a particular tool choice.

#### 2.4.3 DETECTED (AMPLITUDE) IMAGES

**General Description:** Amplitude products are fully focused SAR images that have been stripped of phase information and are typically multi-looked (spatially averaged) to reduce speckle noise and to create pixels of approximately square size. While useful for a range of mapping and monitoring applications, amplitude products unfortunately come in a variety of geocoding stages. Most legacy SAR systems (e.g., ERS-1/2, Envisat, Radarsat-1, JERS-1, and ALOS PALSAR) provide non-geocoded amplitude products

that are left in the original acquisition geometry, and it is upon the user to geocode these datasets manually. Other sensors (e.g., Sentinel-1) make georeferenced amplitude products available. While these products remain in their native acquisition geometry, information needed to link the image coordinate system to geographic coordinates is stored within the image file. Currently, only the recently reprocessed archive of Seasat (available at the Alaska Satellite Facility) provides data in full geocoded formats.

Note that amplitude products are typically geo-referenced or geocoded to an ellipsoidal approximation of the Earth. This means that image distortions caused by surface topography (see Fig. 2.4) are not corrected in amplitude products.

**Applications:** In their original form, the main applications of amplitude images are limited to visualization and data inspection. Only after an end user applies geocoding and terrain correction steps do these products have relevance in mapping, change detection, hazard monitoring, and other Earth observation disciplines.

**Naming Convention:** Amplitude products belong to the L1 family of products. To distinguish them from SLCs, they are often referred to as L1.5 (ALOS PALSAR) or L1 Detected (Sentinel-1, ERS-1/2, Envisat, Radarsat-1, JERS-1). While data are provided in a range of custom formats, most modern sensors increasingly favor standard formatting such as GeoTIFF or HDF5. For more information, please refer to Table 2.5.

**Open Source Software Tools:** Amplitude products can be read and further processed by all of the software tools mentioned in Section 2.4.2 and Table 2.5.

#### 2.4.4 POLARIMETRIC PRODUCTS

**General Description:** Most SAR sensors provide the different channels (i.e., HH, HV, VH, and VV; see Sec. 2.2.3) of multi-polarization data as separate layers, processed to either an L0, L1 SLC, or L1 Detected product. There are, however, some exceptions to this general approach. The NASA JPL-run airborne remote sensing system UAVSAR offers two product types (the Compressed Stokes Matrix and Pauli Decomposition products) that are true po-

larimetric products. The Compressed Stokes Matrix captures information about the polarization state of the measured polarimetric signal, while the Pauli Decomposition provides information on the polarimetric scattering properties of an observed surface. Polarimetric products are also planned for upcoming SAR missions NISAR and TanDEM-L.

**Applications:** Polarimetric data are useful for studying the structure of the observed surface and performing unsupervised image classifications. Polarimetric products have been used extensively in agriculture monitoring (crop classification, soil moisture extraction, and crop assessment) (Alemohammad et al. 2016, Jagdhuber et al. 2013, Liu et al. 2013, Quegan et al. 2003, Xie et al. 2015), oceanography (surface currents and wind field retrieval) (Hooper et al. 2015, Latini et al. 2016, Migliaccio & Nunziata 2014), forestry (forest monitoring, classification, and tree height estimation) (Banqué et al. 2016, Mitchard et al. 2011, Shimada et al. 2016, Walker et al. 2010), disaster monitoring (oil spill detection and disaster assessment), and military applications (ship detection and target recognition/classification).

**Naming Convention:** Due to the recent development of standalone polarimetric products, no naming convention has been established thus far.

**Open Source Software Tools:** Polarimetric data can be processed with the following software packages (sorted in ascending order of sophistication of available polarimetric processing):

- **MapReady**—Developed by the Alaska Satellite Facility. More information and download: <https://www.asf.alaska.edu/data-tools/mapready/>.
- **SNAP (Sentinel Application Platform)**—Developed by ESA. More information and download: <http://step.esa.int/main/download>.
- **PolSARpro**—Developed by ESA. More information and download: <https://earth.esa.int/web/polsarpro/home>.

#### 2.4.5 LEVEL 2 AND HIGHER LEVEL PRODUCTS

**General Description:** For the sake of this handbook, Level 2 data are defined as all data products that are projected to the ground, gridded in regular grids, and transformed into physical variables such as

a calibrated radar cross section (e.g., the radiometric terrain-corrected data generated via the processing flow described in Sec. 2.6.1), line-of-sight deformation, or polarimetric decomposition variables.

While Level 2 products are offered by only a few missions at this point in time, an increasing number of future sensors will offer products at advanced processing levels. Currently, operationally produced Level 2 products are only provided for the Seasat and the UAVSAR archive and include geocoded amplitude images (both sensors) as well as polarimetric and interferometric products (UAVSAR; Table 2.4).

Several data formats have been used for Level 2 and higher products from SAR, as no common data format has been established yet for this still uncommon product type. Currently used formats include GeoTIFF, HDF5, and KMZ.

**Applications:** Level 2 SAR data products facilitate a wide range of applications. Calibrated amplitude images find use in a range of fields such as general mapping, land-use classification, change detection, and hazard analysis. Interferometric data may be used in damage mapping, geophysical analyses of surface deformation, and more.

**Naming Convention:** Various.

**Open Source Software Tools:** The geocoded products can be used in most Geographic Information System (GIS) tools. While GeoTIFF products are natively compatible with GIS tools, readers may have to import HDF5 formats with their respective metadata information.

### 2.5 ACCESSING SAR DATA

While an increasing number of satellite systems are operating under a free and open data policy, many legacy sensors and some currently operating higher resolution sensors are still providing data under a restricted or commercial paradigm. The following sections provide information on how, where, and under which conditions data from these different types of sensors can be accessed.

#### 2.5.1 FREE AND OPEN DATA POLICY MISSIONS

Past and current SAR missions that have (at the writing of this chapter) adopted a free and open data

policy include the spaceborne sensors Seasat, ERS-1/2, ALOS-1, and Sentinel-1, as well as the NASA-operated airborne SARs AirSAR, UAVSAR, and AIRMOSS. Upcoming missions also expected to provide data in a free and open manner include the Canadian Radarsat Constellation Mission (RCM), NASA's NISAR, DLR's TanDEM-L, and ESA's Biomass mission. Means of data access for these sensors is briefly described in the following sections. A summary of data access to free and open SAR sensors is provided in **Table 2.6**.

### 2.5.1.1 Accessing Data from the ERS-1/2 and Envisat Missions

With a combined lifetime from 1991 until 2011, the ERS system (composed of ERS-1 and ERS-2) provides unique insights into 20 years of changes on the Earth's surface. Therefore, it remains a relevant data source for those interested in climate change, hazard monitoring, and environmental analysis.

Two means of accessing data from this long-lived legacy SAR system are:

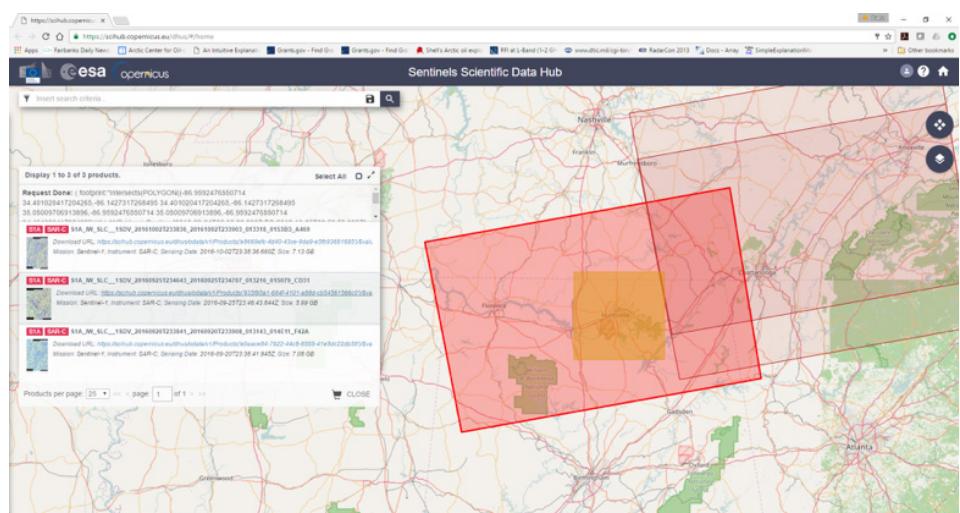
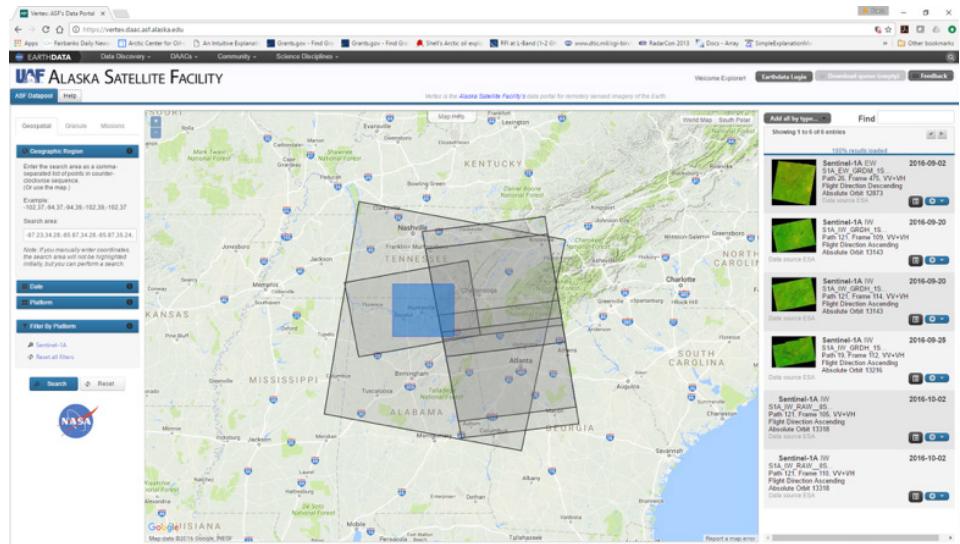
- **ESA Simple Online Catalogue**—The global archive of the ERS and Envisat systems can be searched and ordered via the ESA-maintained [Simple Online Catalogue](#). ESA SOC replaced the EO-L-SA (Earth Observation Link – Stand Alone) browser in early 2019. Once relevant data are identified, images can be downloaded by the user free of charge once reproduction is completed, with the possibility of additional data downloads depending on overall system availability.
- **ASF Vertex**—Alternately, a subset of the available ERS SAR data is available through the services of the NASA Alaska Satellite Facility (ASF) Distributed Active Archive Center (DAAC). Level 0 and L1.5 data over North America ([ASF station mask](#)) are freely and openly available through [ASF's Vertex](#) client for immediate download. See **Figure 2.11** for a view of the ASF Vertex interface. Note that the ASF archive does not include data from the Envisat mission.

### 2.5.1.2 ALOS-1 PALSAR

Data from JAXA's ALOS-1 PALSAR sensor are available through a distributed set of data nodes that

MISSION(S)	REGION	DATA ACCESS	REQUIREMENTS
<b>Seasat</b>	Global	ASF Vertex	One-time registration
<b>ERS-1&amp;2, ENVISAT</b>	Global	<a href="#">ESA Simple Online Catalogue</a>	PI proposal
<b>ALOS-1 PALSAR</b>	ASF Station Mask (ERS only)	ASF Vertex	One-time registration
	Americas/Antarctica	ASF Vertex	One-time registration
	Europe/Africa/Greenland	<a href="#">ALOS PALSAR On-The-Fly</a>	PI proposal
	Asia	<a href="#">AUIG2</a>	PI proposal
<b>Sentinel-1</b>	Australia/Oceania	Unknown	Unknown
	Global	ASF Vertex <a href="#">ESA's Copernicus Open Access Hub</a>	One-time registration
<b>AIRSAR, UAVSAR</b>	Limited extent	ASF Vertex	One-time registration
<b>AirMOSS</b>	Limited extent	<a href="#">ORNL DAAC</a>	One-time registration

**Table 2.6** List of free and open SAR sensors with modes of data access.



**Figure 2.11.** A look at the interfaces of two major SAR data search clients: (a) ASF Vertex client and (b) ESA Copernicus Open Access Hub. Both clients allow for convenient data search via a map interface.

were originally established to make access to ALOS data more effective for end users. Separate data centers are available for the Americas (data located at ASF; free and open access via [ASF Vertex](#)), Europe and Africa (ESA; Principal Investigator (PI) proposal needed; access via [ALOS PALSAR On-The-Fly web interface](#)), Asia (JAXA; PI proposal needed; access via ALOS User Interface Gateway ([AUG2](#))), and Australia/Oceania (Geoscience Australia; data access via JAXA's [AUG2](#)).

In 2015, data from the ALOS-1 PALSAR sensor became [unrestricted](#), enabling all ALOS data nodes to provide data freely and openly to its users. While ASF has fully implemented this data policy—ALOS PALSAR data over the Americas are now freely and openly available through ASF's Vertex client—other data nodes are still working on implementing this unrestricted data policy.

### 2.5.1.3 Sentinel-1

The Sentinel-1 mission is the first of the six Sentinel-dedicated missions operated by the European Copernicus programme. Sentinel-1 is based on a constellation of two SAR satellites to ensure continuity of C-band SAR observations across the globe. Sentinel-1A was launched on April 3, 2014, and the second Sentinel-1 satellite, Sentinel-1B, was launched on April 25, 2016.

The operational nature of Sentinel-1 is a game changer in a number of application domains thanks to the large-scale mapping capability and revisiting frequency of the two identical satellites, together with a high-capacity ground segment that systematically processes, archives, and makes available all the generated data products to users online in a routine operational way (Potin et al. 2016).

The growing global archive of Sentinel-1 is accessible through two freely available search clients:

- **ESA's Copernicus Open Access Hub:** The global archive of the Sentinel-1 SAR constellation can be accessed via [ESA's Copernicus Open Access Hub](#). Requiring only a simple, one-time registration, this hub allows for quick and easy data download via an interactive map interface. In addition to Sentinel-1, the Copernicus Open Access Hub also provides access

MISSION(S)	REGION	DATA ACCESS	REQUIREMENTS
<b>ALOS-2 PALSAR-2</b>	Global	Commercial: <a href="#">PASCO</a> Science: <a href="#">AUG2</a>	<a href="#">Price list</a> <a href="#">Proposal to JAXA</a>
<b>Radarsat-1&amp;2</b>	Global	Commercial: <a href="#">MDA</a>	<a href="#">MDA price list</a>
	North America	Science: <a href="#">ASF Vertex</a>	<a href="#">Proposal to NASA</a>
<b>COSMO-SkyMed</b>	Global	Commercial: <a href="#">e-goes</a>	<a href="#">e-goes price list</a>
		Science: <a href="#">ASI</a>	<a href="#">Proposal to ASI</a>
<b>TerraSAR-X, TanDEM-X</b>	Global	Commercial: <a href="#">Airbus</a>	<a href="#">Airbus price list</a>
		Science (reduced cost): <a href="#">TSX / IDX</a>	<a href="#">Proposal to DLR (TSX / IDX)</a>
		Archived data (free): <a href="#">TSX</a>	<a href="#">Proposal to DLR</a>

**Table 2.7** List of restricted/commercial SAR sensors with modes of data access.

to all other Sentinel missions (at the time of writing, access to Sentinel-1 to Sentinel-3 is possible), making it a convenient one-stop-shop for users interested in multi-sensor Earth observation data. A screenshot of the Copernicus Open Access Hub interface is shown in [Figure 2.11](#).

- **ASF Vertex:** The global Sentinel-1 archive is also available through the previously mentioned [ASF Vertex](#) client ([Fig. 2.11](#)). Similar to the Copernicus Open Access Hub, data can be searched via a convenient map interface. In addition to Sentinel-1, ASF Vertex provides free and open access to other SAR data such as those from the ERS, UAVSAR, AirSAR, and Seasat missions.

- **Google Earth Engine:** In addition to the previous options, geocoded Sentinel-1 Detected (Amplitude) products are now available through [Google Earth Engine](#) (GEE). While GEE does not allow downloading of Sentinel-1 image products, it provides a convenient cloud-based analysis platform within which Sentinel-1 data can be analyzed together with data from optical sensors. Hence, GEE may allow new users of SAR to gain experience with this dataset without requiring local software installs and without having to download large volumes of SAR data.

### 2.5.1.4 NASA's Open Access Airborne SAR Sensors

Data from the NASA airborne SAR sensors AIR-

SAR (C-, L-, and P-bands; 1990–2004) and UAVSAR (L-band; 2008–present) are accessible through the ASF Vertex client ([Fig. 2.11](#)). While covering only limited areas, the versatility and high resolution of these sensors make them interesting for a range of Earth observation disciplines. UAVSAR data are also available at JPL, which is also operating the mission.

To download data directly from JPL, please visit <https://uvsar.jpl.nasa.gov/>.

Currently, AirMOSS data (P-band; 2012–present) are being offered through the NASA Oak Ridge National Laboratory (ORNL) DAAC. To access AirMOSS data from ORNL, please visit [https://daac.ornl.gov/cgi-bin/dataset\\_lister.pl?p=36](https://daac.ornl.gov/cgi-bin/dataset_lister.pl?p=36).

## 2.5.2 RESTRICTED-ACCESS AND COMMERCIAL SAR MISSIONS

Despite their largely commercial nature, there are means to access certain data from the Radarsat-1 and -2, TerraSAR-X, TanDEM-X, COSMO-SkyMed, and ALOS-2 PALSAR-2 missions at low cost. Detailed information on these individual sensors is provided in [Table 2.7](#) and the following paragraphs.

### 2.5.2.1 Radarsat-1 and -2

Radarsat-1 and -2 are Canada's staple SAR sensors which have been continuously providing C-band medium- to high-resolution SAR data since 1995. Since the launch of Radarsat-2 in 2007, most of the combined Radarsat archive has migrated to a commercial data policy with data access fees above \$1,000 CAD per image frame. Information on the data costs for Radarsat-1 and -2 datasets can be found from the com-

mercial distributor [MDA](#). On occasion, low-cost data access is granted to Canadian and European PIs under the [Science and Operational Applications Research for Radarsat-2 \(SOAR\)](#) program after a competitive PI proposal is approved.

Outside of this general agreement, some limited Radarsat-1 data (from the period of 1995–2009) are also available through the NASA ASF DAAC under a restricted data access agreement and can be discovered through its ASF Vertex search client. To access ASF-held Radarsat-1 data, a proposal to NASA is necessary. Once approved, data can be accessed free of charge. More on ASF's Radarsat-1 restricted data use agreement [here](#) at this link.

### 2.5.2.2 ALOS-2 PALSAR-2

While ALOS-2 PALSAR-2 data are distributed commercially by the PASCO Corporation, a limited amount of data is provided for free to the science community. To apply for limited free data access (50 scenes per year), look for regularly released [ALOS Research Announcements](#). A proposal describing the research effort is needed and is reviewed for validity. If approved, free data access to up to 50 scenes per year is granted via the [AUIG2](#) interface. Information about the data costs for commercial ALOS-2 PALSAR-2 data can be found [here](#).

### 2.5.2.3 High-Resolution X-band SAR Data from TerraSAR-X, TanDEM-X, and COSMO-SkyMed

While the high-resolution X-band SAR sensors TerraSAR-X, TanDEM-X, and COSMO-SkyMed provide most of their data under a commercial license, similar to ALOS-2 PALSAR-2, some limited data can be accessed at a low cost (or for free) once a PI proposal is reviewed and approved.

Information on how to access commercial data from the TerraSAR-X and TanDEM-X missions (including pricing information) is available from the [Airbus Defense and Space Company](#) website. For information on how to apply for access to low-cost science use data, see the [TerraSAR-X Science Server](#) or, accordingly, the [TanDEM-X Science Server](#). While proposals to access archived TerraSAR-X data can always be submitted, look for special announcements of opportunities to

apply for access to newly acquired or special mission phase data. Through the TanDEM-X Science Server, users can also apply for segments of the TanDEM-X Digital Elevation Model (DEM) in addition to the SAR images themselves.

Information on commercial access to COSMO-SkyMed data can be retrieved from their commercial vendor, [e-geos](#). Reduced-rate science data access is available regularly through COSMO SkyMed Constellation Data Utilization announcement of opportunities. Please check for upcoming opportunities on the [Italian Space Agency \(ASI\)](#) webpage.

## 2.6 SAR Image Processing Routines – Theory

### 2.6.1 GEOCODING AND RADIOMETRIC TERRAIN CORRECTION

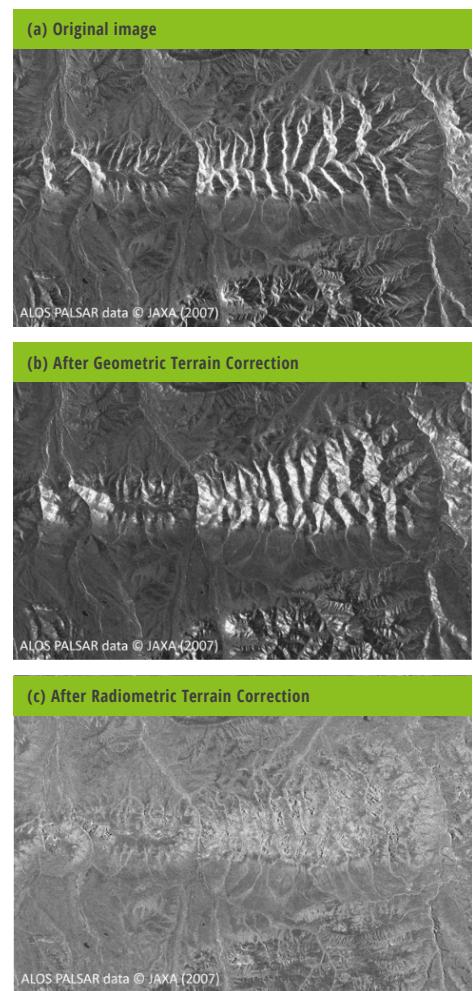
#### 2.6.1.1 Theoretical Background

Due to the side-looking observation geometry, SAR images are subject to geometric and radiometric distortions ([Sec. 2.1.4](#)). In addition to the geometric mislocation of pixels in topographically inclined areas, the oblique angle of the illuminating radar energy adds topographic shading to the true surface RCS, giving the sensor-facing side of hill slopes a radiometrically “overexposed” appearance (see [Figure 2.12\(a\)](#)). Both of these effects hamper the use of SAR for many applications. The radiometric modulations often disguise the true radar reflectance of the observed scene, reducing the applicability of SAR for studying the properties of the surface. Furthermore, geometric and radiometric distortions make the application of SAR for change detection more difficult, as these highly incidence angle-dependent artifacts lead to classification errors if images with different observation geometries are combined. Hence, correction of geometric and radiometric distortions is advisable if SAR data are to be analyzed together with other image data or across datasets with varying incidence angles.

The RCS of a pixel in a calibrated SAR image is composed of:

$$\sigma = \sigma^0(\theta) \cdot A_o(\theta) , \quad (2.9)$$

where  $\sigma^0$  is the (incidence angle-dependent) normal-



**Figure 2.12** Example of geometric (b) and radiometric (c) normalization applied to an ALOS PALSAR image over Alaska (a). The applied corrections enable the use of SAR data in GIS environments (geometric correction step), provide physically correct RCS values for every pixel, and enable unbiased change detection from multiple observation geometries.

ized RCS,  $\theta$ , is the local incidence angle, and  $A_o$  is the surface area covered by a pixel. Following Eq. (2.9), two images acquired from different geometries will differ due to the incidence angle dependence of  $\sigma^0$  and  $A_o$ , even if the observed surface remains unchanged.

Hence, to enable unbiased analysis of SAR images in a GIS and to allow for a joint change detection analysis of SAR amplitude images acquired from different observation geometries, geometric and radiometric distortions in these images need to first be removed. To retrieve the true RCS of the imaged surface  $\sigma^0$  from

the observed radar data  $\sigma$ , the geometry dependence of  $\sigma$  needs to be removed by correcting for  $A_g(\theta)$ . The process of correcting for  $A_g(\theta)$  is called Radiometric Terrain Correction (RTC) (Small 2011). RTC includes both geometric terrain correction (geocoding) and radiometric compensation and is typically performed using the following steps:

- Geometric terrain correction (geocoding) is conducted to remove geometric image distortions. A DEM is needed to correct the location of topographically inclined pixels. In areas between  $\pm 60^\circ$  geographic latitude, the DEM provided by the Shuttle Radar Topography Mission (SRTM) should be sufficient (Gesch et al. 2014).
- Radiometric terrain normalization is performed to remove geometry-dependent radiometric distortions corresponding to a pixel-by-pixel estimation and compensation of  $A_g(\theta)$  using a DEM. The radiometric normalization technique in Small (2011) is applied.

**Figure 2.12** shows an example of the effects of geometric and radiometric normalization. **Figure 2.12(a)** shows an original ALOS PALSAR image over an area near the Denali fault in Alaska. The effects of geometric correction are shown in **Figure 2.12(b)**, and the effects of radiometric normalization are presented in **Figure 2.12(c)**. The normalized data are now largely devoid of geometric influences, reducing radiometric differences between images acquired from different geometries. As a consequence, the RTC-corrected image data show improved performance when combined with other remote sensing datasets and in multi-geometry change detection.

### 2.6.1.2 More Information on Geocoding and RTC Processing

To learn more about the theory behind geocoding and RTC processing please visit Lecture 9 of UAF's [Online Class on Microwave Remote Sensing](#). You can find Lecture 9 in [Class Module 2 "Imaging Radar Systems."](#) To go directly to the slide deck, [click here](#).

## 2.6.2 THEORY OF INTERFEROMETRIC SAR

InSAR processing exploits the difference between the phase signals of repeated SAR acquisitions to analyze the shape and deformation of the Earth's

### Supplemental materials on InSAR

#### The Principles and Applications of Interferometric SAR (InSAR):

Interferometric SAR (InSAR) analyzes phase differences between two or more SAR acquisitions with the goal of measuring surface topography and/or surface deformation. While the quality of derived topographic information depends on the relative observation geometry of the SAR acquisitions used, surface deformation can be measured at a fraction of the signal wavelength and, hence, with millimeter to centimeter accuracy. In this lecture, you will hear about the concepts of InSAR and the general processing approaches to arrive at either surface topography or surface deformation. Limitations of InSAR as well as advanced processing concepts will be covered in future lectures.

**Link:** [https://radar.community.uaf.edu/files/2017/03/Lecture12\\_ConceptsAndGeneralApproachesOfInSAR.pdf](https://radar.community.uaf.edu/files/2017/03/Lecture12_ConceptsAndGeneralApproachesOfInSAR.pdf)

#### Phase Unwrapping & Limitations of Traditional InSAR Methods:

The first part of this lecture will deal with the problem of phase unwrapping. As InSAR phase measurements are initially only available wrapped into the value range, a phase unwrapping process has to be applied to create an unambiguous phase map ready for topography or deformation analysis. You will be introduced to the general process of phase unwrapping and learn about several popular solutions to this problem. In the second part of this lecture, we will look into the main limitations of the traditional two-image InSAR approach. These identified limitations will set us up for future lectures, which will describe advanced processing techniques (e.g., PS- and SBAS InSAR).

**Link:** [https://radar.community.uaf.edu/files/2017/03/Lecture13\\_PhaseUnwrappingandLimitationsofInSAR.pdf](https://radar.community.uaf.edu/files/2017/03/Lecture13_PhaseUnwrappingandLimitationsofInSAR.pdf)

#### The Role of InSAR in Geophysics:

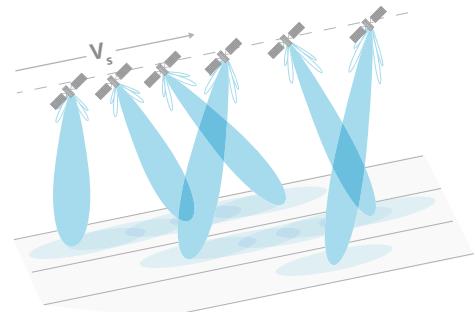
Intrinsically, InSAR is a geodetic discipline, providing accurate measurements of surface deformation. While this is interesting by itself, geoscientists are typically more interested in the geophysical source that causes an observed deformation rather than the deformation itself. Using volcanic activity as an example, this lecture will provide you with some insight on how geophysical parameters can be determined using InSAR measurements in combination with inverse modeling.

**Link:** [https://radar.community.uaf.edu/files/2017/03/Lecture14\\_UsingInSARinGeophysics.pdf](https://radar.community.uaf.edu/files/2017/03/Lecture14_UsingInSARinGeophysics.pdf)

surface. While the principles and processing flows of InSAR will not be described here in detail, is recommended to look through the following material that is available freely and openly online. The lecture materials listed are part of a full-semester, graduate-level class on microwave remote sensing offered by the University of Alaska Fairbanks, listed in the "Supplemental materials on InSAR" callout.

### 2.6.2.1 A Word on Sentinel-1 Interferometric Wide Swath Data

The Interferometric Wide (IW) swath mode is the main acquisition mode over land for Sentinel-1. It acquires data with a 250-km swath at 5-x-20-m spatial resolution (single look). Interferometric wide mode



**Figure 2.13** TOPSAR acquisition principle.

captures three sub-swaths using the TOPSAR acquisition principle. With the TOPSAR technique, in addition to steering the beam in range as in ScansAR, the beam is also electronically steered from backward

to forward in the azimuth direction for each burst, avoiding scalloping and resulting in homogeneous image quality throughout the swath. A schematic of the TOPSAR acquisition principle is shown in [Figure 2.13](#).

The TOPSAR mode replaces the conventional ScanSAR mode, achieving the same coverage and resolution as ScanSAR, but with nearly uniform image quality (in terms of signal-to-noise ratio and distributed target ambiguity ratio).

Interferometric wide SLC products contain one image per sub-swath and one per polarization channel, for a total of three (single-polarization) or six (dual-polarization) images in an IW product. Each sub-swath image consists of a series of bursts, where each burst has been processed as a separate SLC image. The individually focused complex burst images are included, in azimuth-time order, into a single sub-swath image with black-fill demarcation in between.

### 2.6.3 CHANGE DETECTION USING SAR

#### 2.6.3.1 Problem Statement

Detecting changes in land-use/land-cover is one of the most fundamental and common uses of remote sensing image analysis. One of the most rudimentary forms of change detection is the visual comparison of two images by a trained interpreter. With an effective display system large enough to display both images simultaneously and to explore and digitize with a cursor tracking to the same location in both images, this is a quick method that can be used to locally collect valuable GIS-compatible data while streaming the images themselves over a relatively low-bandwidth Internet connection.

In an attempt to automate change detection (and hence make it available for large-scale and more operational implementation), a wealth of digital change detection algorithms have been developed over the last decade that operate on a range of different sensors and are grouped into “supervised” and “unsupervised” categories. While a great many methods for detecting changes from remote sensing data are available in literature, this short introduction is limited to methods that are used in reference to SAR.

#### 2.6.3.2 Summary of SAR-Based Change

### Detection Techniques

Even when limiting research to SAR-based change detection only, the number of algorithms proposed in recent years can seem overwhelming. Hence, instead of providing an extensive summary of all available techniques, this section attempts to categorize techniques to help in choosing the right method for an envisioned application. Methods will be categorized using several indicators such as by the type of input information needed, the required amount of training data, and the amount of processing expertise needed to implement the algorithms.

#### 2.6.3.2.1 Input Data Used for Change Detection

SAR-based change detection techniques can be categorized by the type of SAR information used for change identification. Categories include “amplitude-based methods,” “phase/coherence-based techniques,” and “polarimetric techniques.”

**Amplitude-based methods** focus on the RCS information contained in the data, initially ignoring information coming from phase and polarization. One of the advantages of amplitude-based methods lies in their ability maximize the temporal sampling that can be achieved with SAR-based change detection information. Amplitude information is available for every SAR collection, making every new image useful for change detection. As not all SAR acquisitions allow for the use of phase and/or polarization, amplitude data naturally lead to better temporal sampling. This benefit can be further enhanced if RTC is applied to all images. As RTC processing removes most geometry-dependent distortions from the measured SAR RCS, it allows for combining SAR data acquired from multiple incidence angles, leading to further improvements in temporal sampling. However, a disadvantage of amplitude-based methods relates to its limited sensitivity, which often increases the likelihood of false negatives, in which true changes are erroneously missed in the classification.

**Phase/coherence-based techniques** utilize the fact that significant surface change results in a significant reduction of interferometric coherence, enabling the automatic detection of change via coherence thresholding. Coherence-based techniques are

highly sensitive to change, which interestingly is both the main advantage and disadvantage of this category. On one hand, the high sensitivity is an asset, as it reduces the likelihood for false negatives. On the other, coherent change detection methods tend to have very large false positive rates, where change is vastly overestimated. While methods have been developed to combat these problems, the need for false positive correction makes coherence-based methods appear very complicated and non-straightforward for the uninitiated user. Coherent image pairs are required for these methods to be applicable, which somewhat limits the temporal sampling that can be achieved.

**Polarimetric techniques** are often highly capable, as they can analyze surface changes across several polarimetric channels. This maximizes the likelihood of change detection and allows one to associate those changes with scattering types (e.g., changes associated with double-bounce, roughness, and volume scattering). The latter is especially relevant, as it enables one to ascertain as to whether a change signature is related to vegetation or the ground, enabling change classification. However, the main disadvantage of polarimetric change detection is related to its reliance on multi-polarization data, which are not always available. Furthermore, polarimetric processing theory may be a bit overwhelming to uninitiated readers.

Independent of change detection methods, the proper choice of sensor is essential to optimizing change detection performance. In particular, the choice of sensor wavelength should be appropriate given the surface and vegetation characteristics of an area of interest. If changes underneath vegetation canopies are the target, longer wavelength sensors are preferred. For bare surfaces, shorter wavelengths often have an advantage. This is because shorter wavelength sensors often increase the RCS associated with rough surfaces and provide more dynamic range that can be used for the identification of change.

#### 2.6.3.2.2 Supervised vs. Unsupervised Methods

Change detection can be performed either unsupervised (Bruzzone & Prieto 2000) or supervised (Huo et al. 2010). In unsupervised change detection,

a change map is generated by comparing objects in two images with a similarity metric. The change map then undergoes thresholding to classify each pixel into changed and unchanged classes (Bruzzone & Prieto 2000, Otsu 1979). In supervised change detection, training samples are selected from the available dataset and are used to train a classifier, which is then used to classify an image into changed and unchanged classes (Huo et al. 2010).

Supervised methods are useful, as radar signatures associated with change do not always have to be theoretically understood to be able to detect them. Instead, the impact of a surface change on the signal observed by a SAR is learned using training data, reducing the need for sophisticated modeling. The reliance on reference data, however, is also the main downside of these methods, as training data are sometimes hard to come by and are seldom free

	AMPLITUDE-BASED	PHASE-BASED	POLARIZATION-BASED
Supervised	White, 1991 Gong et al., 2016 Liu et al., 2016 Gong et al., 2017	Gamba et al., 2007 Pulvirenti et al., 2016	Marino and Hajnsek, 2014
Unsupervised	Meyer et al., 2014 Ajadi et al., 2016 Bruzzone and Prieto, 2000 Bazi et al., 2005 Celik, 2010 Bovolo and Bruzzone, 2005	Yun et al., 2015a Yun et al., 2015b Sharma et al., 2017	Akbari et al., 2016

of errors. Some recent supervised algorithms based on amplitude, phase, and polarization data are listed in **Table 2.8**.

Unsupervised methods have the advantage that no reference data are required to arrive at a classification result. Instead, signal models are used to encode the impact of surface change on the observed

**Table 2.8** List of change detection methods categorized by source data and need for reference data.

data. Unsupervised techniques are particularly beneficial in hazard monitoring, where changes are often unanticipated and training data are typically not available in time. Selected recent unsupervised change detection methods are listed in **Table 2.8** as a function of input data type.

## SARbian – A free and open SAR Operating System:

SARbian is an easy-to-use, Linux-based SAR processing virtual machine provided by the group behind the EO-College initiative (<https://eo-college.org>) that comes loaded with a wide range of currently-available, free-and-open SAR processing and GIS software tools. The virtual machine is completely pre-installed, ready for use in research, education, or operational applications. No knowledge of installation steps is needed. Hence, SARbian is a convenient resource for researchers and decision-makers that are looking for a hassle-free start with SAR.

SARbian can be downloaded from <https://eo-college.org/sarbian>, and comes with the following list of software tools:

- SAR Processing Tools: ESA S1TBX; ASF MapReady; pyroSAR
- SAR Polarimetry: PolSARPro
- SAR Interferometry: DORIS; SNAPHU (phase unwrapping); PyRAT
- GIS Tools: GDAL; QGIS; GRASS GIS
- Supporting Tools: A number of Python, R, and Octave resources

## 2.7 References

- Ajadi, O. A., Meyer, F. J., and Webley, P. W., 2016, *Change Detection in Synthetic Aperture Radar Images Using a Multiscale-Driven Approach*: *Remote Sensing*, v. 8, no. 6, p. 482.
- Akbari, V., Anfinsen, S. N., Doulgeris, A. P., Eltoft, T., Moser, G., and Serpico, S. B., 2016, *Polarimetric SAR Change Detection With the Complex Hotelling–Lawley Trace Statistic*: *IEEE Transactions on Geoscience and Remote Sensing*, v. 54, no. 7, p. 3953-3966.
- Alemohammad, S. H., Jagdhuber, T., Moghaddam, M., and Entekhabi, D., *Decomposing soil and vegetation contributions in polarimetric L-and P-band SAR observations*, in *Proceedings Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International* 2016, IEEE, p. 7553-7556.
- Bamler, R., and Eineder, M., 1996, *ScanSAR processing using standard high precision SAR algorithms*: *IEEE Transactions on Geoscience and Remote Sensing*, v. 34, no. 1, p. 212-218.
- Bamler, R., and Holzner, J., 2004, *ScanSAR interferometry for RADARSAT-2 and RADARSAT-3*: *Canadian Journal of Remote Sensing*, v. 30, no. 3, p. 437-447.
- Banqué, X., Lopez-Sánchez, J. M., Monells, D., Ballester, D., Duro, J., and Koudogbo, F., 2016, *Polarimetry-based Land Cover Classification with Sentinel-1 Data*: *Seminarios*, v. 13, p. 07.
- Bazi, Y., Bruzzone, L., and Melgani, F., 2005, *An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images*: *IEEE Transactions on Geoscience and Remote Sensing*, v. 43, no. 4, p. 874-887.
- Belcher, D. P., 2008, *Theoretical Limits on SAR Imposed by the Ionosphere*: *IET Radar Sonar and Navigation*, v. 2, no. 6, p. 435-448.
- Belcher, D. P., and Rogers, N. C., 2009, *Theory and Simulation of Ionospheric Effects on Synthetic Aperture Radar*: *IET Radar Sonar and Navigation*, v. 3, no. 5, p. 541-551.
- Bovolo, F., and Bruzzone, L., 2005, *A detail-preserving scale-driven approach to change detection in multitemporal SAR images*: *IEEE Transactions on Geoscience and Remote Sensing*, v. 43, no. 12, p. 2963-2972.
- Bruniquel, J., and Lopes, A., 1997, *Multi-variate optimal speckle reduction in SAR imagery*: *International Journal of Remote Sensing*, v. 18, no. 3, p. 603-627.
- Bruzzone, L., and Prieto, D. F., 2000, *Automatic analysis of the difference image for unsupervised change detection*: *IEEE Transactions on Geoscience and Remote sensing*, v. 38, no. 3, p. 1171-1182.
- Buades, A., Coll, B., and Morel, J. M., *A non-local algorithm for image denoising*, in *Proceedings 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) 20-25 June 2005, Volume 2*, p. 60-65 vol. 62.
- Celik, T., 2010, *A Bayesian approach to unsupervised multiscale change detection in synthetic aperture radar images*: *Signal Processing*, v. 90, no. 5, p. 1471-1485.
- Chen, S., Hou, J., Zhang, H., and Da, B., 2014, *De-speckling method based on non-local means and coefficient variation of SAR image*: *Electronics Letters*, v. 50, no. 18, p. 1314-1316.
- Cumming, I. G., and Wong, F. H., 2005, *Digital Processing of Synthetic Aperture Radar Data*, Norwood, MA, Artech House, Inc.
- De Zan, F., and Monti Guarnieri, A., 2006, *TOPSAR: Terrain Observation by Progressive Scans*: *IEEE Transactions on Geoscience and Remote Sensing*, v. 44, no. 9, p. 2352-2360.
- Di Martino, G., Di Simone, A., Iodice, A., and Riccio, D., 2016, *Scattering-based nonlocal means SAR despeckling*: *IEEE Transactions on Geoscience and Remote Sensing*, v. 54, no. 6, p. 3574-3588.
- Eineder, M., Adam, N., Bamler, R., Yague-Martinez, N., and Breit, H., 2009, *Spaceborne Spotlight SAR Interferometry With TerraSAR-X*: *IEEE Transactions on Geoscience and Remote Sensing*, v. 47, no. 5, p. 1524-1535.
- Ferretti, A., Fumagalli, A., Novali, F., Prati, C., Rocca, F., and Rucci, A., 2011, *A New Algorithm for Processing Interferometric Data-Stacks: SqueezSAR*: *IEEE Transactions on Geoscience and Remote Sensing*, v. 49, no. 9, p. 3460-3470.
- Freeman, A., Krieger, G., Rosen, P., Younis, M., Johnson, W., Huber, S., Jordan, R., and Moreira, A., *SweepSAR: Beam-forming on receive using a reflector-phased array feed combination for spaceborne SAR*, in *Proceedings IEEE Radar Conference, 2009 4-8 May 2009* 2009, p. 1-9.
- Frost, V. S., Stiles, J. A., Shanmugan, K. S., and Holtzman, J. C., 1982, *A model for radar images and its application to adaptive digital filtering of multiplicative noise*: *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 2, p. 157-166.
- Fu, L.-L., and Holt, B., 1982, *Seasat views oceans and sea ice with synthetic-aperture radar*, California Institute of Technology, Jet Propulsion Laboratory.
- Gamba, P., Dell'Acqua, F., and Trianni, G., 2007, *Rapid damage detection in the Bam area using multitemporal SAR and exploiting ancillary data*: *IEEE Transactions on Geoscience and Remote Sensing*, v. 45, no. 6, p. 1582-1589.
- Gebert, N., Krieger, G., and Moreira, A., 2009, *Digital Beamforming on Receive: Techniques and Optimization Strategies for High-Resolution Wide-Swath SAR Imaging*: *IEEE Transactions on Aerospace and Electronic Systems*, v. 45, no. 2, p. 564-592.
- Gesch, D. B., Oimoen, M. J., and Evans, G. A., 2014, *Accuracy assessment of the US Geological Survey National Elevation Dataset, and comparison with other large-area elevation datasets: SRTM and ASTER*: *US Geological Survey*, 2331-1258.
- Gomba, G., Parizzi, A., De Zan, F., Eineder, M., and Bamler, R., 2016, *Toward operational compensation of ionospheric effects in SAR interferograms: the split-spectrum method*: *IEEE Transactions on Geoscience and Remote Sensing*, v. 54, no. 3, p. 1446-1461.

- Gong, M., Yang, H., and Zhang, P., 2017, Feature learning and change feature classification based on deep learning for ternary change detection in SAR images: *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 129, p. 212-225.
- Gong, M., Zhao, J., Liu, J., Miao, Q., and Jiao, L., 2016, Change detection in synthetic aperture radar images based on deep neural networks: *IEEE transactions on neural networks and learning systems*, v. 27, no. 1, p. 125-138.
- Hooper, B. A., Van Pelt, B., Williams, J., Dugan, J., Yi, M., Piotrowski, C., and Miskey, C., 2015, Airborne spectral polarimeter for ocean wave research: *Journal of Atmospheric and Oceanic Technology*, v. 32, no. 4, p. 805-815.
- Huang, S.-q., Liu, D.-z., Gao, G.-q., and Guo, X.-j., 2009, A novel method for speckle noise reduction and ship target detection in SAR images: *Pattern Recognition*, v. 42, no. 7, p. 1533-1542.
- Huo, C., Zhou, Z., Lu, H., Pan, C., and Chen, K., 2010, Fast object-level change detection for VHR images: *IEEE Geoscience and Remote Sensing Letters*, v. 7, no. 1, p. 118-122.
- Jagdhuber, T., Hajnsek, I., Bronstert, A., and Papathanassiou, K. P., 2013, Soil moisture estimation under low vegetation cover using a multi-angular polarimetric decomposition: *IEEE Transactions on Geoscience and Remote Sensing*, v. 51, no. 4, p. 2201-2215.
- Jehle, M., Frey, O., Small, D., and Meier, E., 2010, Measurement of Ionospheric TEC in Spaceborne SAR Data: *IEEE Transactions on Geoscience and Remote Sensing*, v. 48, no. 6, p. 2460-2468.
- Jehle, M., Ruegg, M., Zuberbuhler, L., Small, D., and Meier, E., 2009, Measurement of Ionospheric Faraday Rotation in Simulated and Real Spaceborne SAR Data: *IEEE Transactions on Geoscience and Remote Sensing*, v. 47, no. 5, p. 1512-1523.
- Kim, J. S., Danklmayer, A., and Papathanassiou, K., Correction of ionospheric distortions in low frequency interferometric SAR data, in *Proceedings Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International* 24-29 July 2011 2011, p. 1505-1508.
- Krieger, G., and Moreira, A., Potential of digital beamforming in bi- and multistatic SAR, in *Proceedings Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. Proceedings. 2003 IEEE International* 21-25 July 2003 2003, Volume 1, p. 527-529 vol.521.
- Lanari, R., Tesauro, M., Sansosti, E., and Fornaro, G., 2001, Spotlight SAR data focusing based on a two-step processing approach: *IEEE Transactions on Geoscience and Remote Sensing*, v. 39, no. 9, p. 1993-2004.
- Latini, D., Del Frate, F., and Jones, C. E., 2016, Multi-frequency and polarimetric quantitative analysis of the Gulf of Mexico oil spill event comparing different SAR systems: *Remote Sensing of Environment*, v. 183, p. 26-42.
- Lee, J.-S., 1980, Digital image enhancement and noise filtering by use of local statistics: *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 2, p. 165-168.
- Lee, J.-S., Grunes, M., and Mango, S. A., 1991, Speckle reduction in multipolarization, multifrequency SAR imagery: *Geoscience and Remote Sensing, IEEE Transactions on*, v. 29, no. 4, p. 535-544.
- Lee, J.-S., Jurkevich, L., Dewaele, P., Wambacq, P., and Oosterlinck, A., 1994, Speckle filtering of synthetic aperture radar images: A review: *Remote Sensing Reviews*, v. 8, no. 4, p. 313-340.
- Liu, C., Shang, J., Vachon, P. W., and McNairn, H., 2013, Multiyear crop monitoring using polarimetric RADARSAT-2 data: *IEEE Transactions on Geoscience and Remote sensing*, v. 51, no. 4, p. 2227-2240.
- Liu, R., Jia, Z., Qin, X., Yang, J., and Kasabov, N., 2016, SAR Image Change Detection Method Based on Pulse-Coupled Neural Network: *Journal of the Indian Society of Remote Sensing*, v. 44, no. 3, p. 443-450.
- Lopes, A., Touzi, R., and Nezry, E., 1990, Adaptive speckle filters and scene heterogeneity: *IEEE transactions on Geoscience and Remote Sensing*, v. 28, no. 6, p. 992-1000.
- Lopez-Martinez, C., and Pottier, E., 2007, On the Extension of Multidimensional Speckle Noise Model from Single-look to Multilook SAR Imagery: *IEEE Transactions on Geoscience and Remote Sensing*, v. 45, no. 2, p. 305-320.
- Marino, A., and Hajnsek, I., 2014, A change detector based on an optimization with polarimetric SAR imagery: *IEEE Transactions on Geoscience and Remote Sensing*, v. 52, no. 8, p. 4781-4798.
- Martino, G. D., Simone, A. D., Iodice, A., Riccio, D., and Ruello, G., Non-local means SAR despeckling based on scattering, in *Proceedings 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 26-31 July 2015* 2015, p. 3172-3174.
- Meyer, F., Bamler, R., Jakowski, N., and Fritz, T., 2006, The Potential of Low-Frequency SAR Systems for Mapping Ionospheric TEC Distributions: *IEEE Geoscience and Remote Sensing Letters*, v. 3, no. 4, p. 560-564.
- Meyer, F., and Nicoll, J., Mapping Ionospheric TEC using Faraday Rotation in Full-Polarimetric L-Band SAR Data, in *Proceedings European Synthetic Aperture Radar Conference EU-SAR'08*, Friedrichshafen, Germany, 2008a, Volume 2, VDE Verlag, p. 23-26.
- Meyer, F. J., 2011, Performance Requirements for Ionospheric Correction of Low-Frequency SAR Data: *IEEE Transactions on Geoscience and Remote Sensing*, v. 49, no. 10, p. 3694-3702.
- Meyer, F. J., Chotoo, K., Chotoo, S. D., Huxtable, B. D., and Carrano, C. S., 2016, The Influence of Equatorial Scintillation on L-Band SAR Image Quality and Phase: *IEEE Transactions on Geoscience and Remote Sensing*, v. 54, no. 2, p. 869-880.
- Meyer, F. J., McAlpin, D. B., Gong, W., Ajadi, O., Arko, S., Webley, P. W., and Dehn, J., 2014, Integrating SAR and derived products into operational volcano monitoring and decision support systems: *Isprs Journal of Photogrammetry and Remote Sensing*, no. 0.

- Meyer, F. J., and Nicoll, J. B., 2008b, *Prediction, Detection, and Correction of Faraday Rotation in Full-Polarimetric L-Band SAR Data*: *IEEE Transactions on Geoscience and Remote Sensing*, v. 46, no. 10, p. 3076-3086.
- Migliaccio, M., and Nunziata, F., 2014, *On the exploitation of polarimetric SAR data to map damping properties of the Deepwater Horizon oil spill*: *International journal of remote sensing*, v. 35, no. 10, p. 3499-3519.
- Mitchard, E. T. A., Saatchi, S. S., Lewis, S. L., Feldpausch, T. R., Woodhouse, I. H., Sonké, B., Rowland, C., and Meir, P., 2011, *Measuring biomass changes due to woody encroachment and deforestation/degradation in a forest-savanna boundary region of central Africa using multi-temporal L-band radar backscatter*: *Remote Sensing of Environment*, v. 115, no. 11, p. 2861-2873.
- Mittermayer, J., Moreira, A., and Loffeld, O., 1999, *Spotlight SAR data processing using the frequency scaling algorithm*: *IEEE Transactions on Geoscience and Remote Sensing*, v. 37, no. 5, p. 2198-2214.
- Monti Guarnieri, A., and Prati, C., 1996, *ScanSAR focusing and interferometry*: *IEEE Transactions on Geoscience and Remote Sensing*, v. 34, no. 4, p. 1029-1038.
- Novak, L. M., and Burl, M. C., 1990, *Optimal speckle reduction in polarimetric SAR imagery*: *Aerospace and Electronic Systems, IEEE Transactions on*, v. 26, no. 2, p. 293-305.
- Otsu, N., 1979, *A threshold selection method from gray-level histograms*: *IEEE transactions on systems, man, and cybernetics*, v. 9, no. 1, p. 62-66.
- Pi, X., Meyer, F. J., Chotoo, K., Freeman, A., Caton, R. G., and Bridgwood, C. T., *Impact of ionospheric Scintillation on Spaceborne SAR Observations Studied Using GNSS*, in *Proceedings ION GNSS2012*, p. 1998-2006.
- Potin, P., Rosich, B., Grimont, P., Miranda, N., Shurmer, I., O'Connell, A., Torres, R., and Krassenburg, M., *Sentinel-1 mission status*, in *Proceedings EUSAR 2016: 11th European Conference on Synthetic Aperture Radar, Proceedings of 2016*, VDE, p. 1-6.
- Pottier, E., and Lee, J. S., 2009, *Polarimetric Radar Imaging: From Basics to Applications*, Boca Raton, FL, CRC Press; Taylor & Francis Group, 438 p.
- Pulvirenti, L., Chini, M., Pierdicca, N., and Boni, G., 2016, *Use of SAR data for detecting floodwater in urban and agricultural areas: The role of the interferometric coherence*: *IEEE Transactions on Geoscience and Remote Sensing*, v. 54, no. 3, p. 1532-1544.
- Quegan, S., Le Toan, T., Skriver, H., Gomez-Dans, J., Gonzalez-Sampedro, M. C., and Hoekman, D. H., *Crop classification with multitemporal polarimetric SAR data*, in *Proceedings Applications of SAR Polarimetry and Polarimetric Interferometry 2003, Volume 529*.
- Quegan, S., and Yu, J. J., 2001, *Filtering of multichannel SAR images*: *IEEE Transactions on Geoscience and Remote Sensing*, v. 39, no. 11, p. 2373-2379.
- Sharma, R. C., Tateishi, R., Hara, K., Nguyen, H. T., Gharechelou, S., and Nguyen, L. V., 2017, *Earthquake damage visualization (EDV) technique for the rapid detection of earthquake-induced damages using SAR data*: *Sensors*, v. 17, no. 2, p. 235.
- Shimada, M., Itoh, T., Motooka, T., Watanabe, M., and Thapa, R., *Generation of the first PALSAR-2 global mosaic 2014/2015 and change detection between 2007 and 2015 using the PALSAR and PALSAR-2*, in *Proceedings Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International 2016*, IEEE, p. 3871-3872.
- Small, D., 2011, *Flattening Gamma: Radiometric Terrain Correction for SAR Imagery*: *IEEE Transactions on Geoscience and Remote Sensing*, v. 49, no. 8, p. 3081-3093.
- Sveinsson, J. R., and Benediktsson, J. A., 2003, *Almost translation invariant wavelet transformations for speckle reduction of SAR images*: *Geoscience and Remote Sensing, IEEE Transactions on*, v. 41, no. 10, p. 2404-2408.
- Van Zyl, J. J., 2011, *Synthetic Aperture Radar Polarimetry*, Hoboken, New Jersey, John Wiley & Sons, Inc, 312 p.
- Vollmar, F., 1960, *Das Telemobiloskop von Christian Hülsmeier, ein früher Vorläufer des Radargeräts*: Deutsches Museum, Abhandlungen und Berichte, v. 28, p. 33-40.
- Walker, W. S., Stickler, C. M., Kellndorfer, J. M., Kirsch, K. M., and Nepstad, D. C., 2010, *Large-area classification and mapping of forest and land cover in the Brazilian Amazon: A comparative analysis of ALOS/PALSAR and Landsat data sources*: *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, v. 3, no. 4, p. 594-604.
- Watson-Watt, R., 1946, *The evolution of radiolocation*: *Journal of the Institution of Electrical Engineers - Part IIIA: Radiolocation*, v. 93, no. 1, p. 11-19.
- White, R., 1991, *Change detection in SAR imagery*: *International Journal of remote sensing*, v. 12, no. 2, p. 339-360.
- Woodhouse, I. H., 2006, *Introduction to Microwave Remote Sensing*, Boca Raton, FL, CRC Press, Taylor & Francis Group, 400 p.
- Xie, L., Zhang, H., Wu, F., Wang, C., and Zhang, B., 2015, *Capability of rice mapping using hybrid polarimetric SAR data*: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, v. 8, no. 8, p. 3812-3822.
- Younis, M., Fischer, C., and Wiesbeck, W., 2003, *Digital beamforming in SAR systems*: *IEEE Transactions on Geoscience and Remote Sensing*, v. 41, no. 7, p. 1735-1739.
- Yun, S.-H., Fielding, E. J., Webb, F. H., and Simons, M., 2015a, *Damage proxy map from interferometric synthetic aperture radar coherence*, Google Patents.
- Yun, S.-H., Hudnut, K., Owen, S., Webb, F., Simons, M., Sacco, P., Gurrola, E., Manipon, G., Liang, C., and Fielding, E., 2015b, *Rapid Damage Mapping for the 2015 M w 7.8 Gorkha Earthquake Using Synthetic Aperture Radar Data from COSMO-SkyMed and ALOS-2 Satellites*: *Seismological Research Letters*, v. 86, no. 6, p. 1549-1556.

# APPENDIX A

## SAR Image Processing Routines

### – Chapter 2 Training Module

#### 1 GEOCODING AND RTC PROCESSING USING ASF MAPREADY

Many of the SAR data holdings in the global archives are available as so-called ground-range projected products. While these products are typically georeferenced, they usually use an ellipsoid as reference surface. Hence, geometric distortions such as foreshortening are not corrected in these products and geolocation errors occur at points that don't lie at the height of the applied reference surface.

This lab is for users who wish to geocode and generate an RTC image from ERS-2 data using MapReady. MapReady is a free software tool distributed by ASF that can be used to correct geometric distortions from SAR data and generate fully geocoded GeoTIFF products ready for use in GIS analyses. In this part of the lab, we will demonstrate the MapReady tool and use it to geocode a ERS-2 scene over Fairbanks, AK.

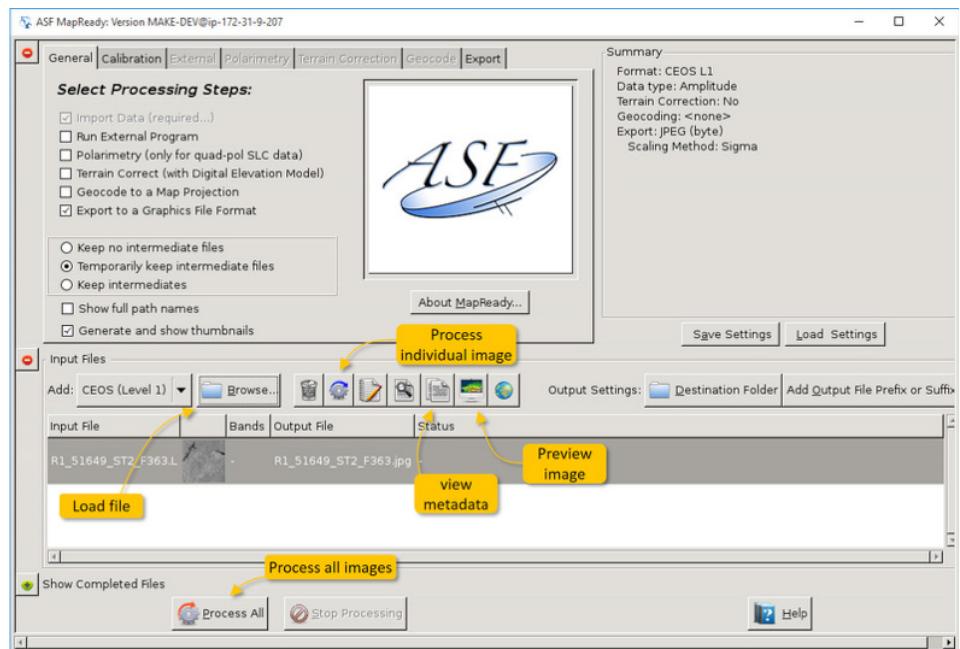
##### 1.1 Starting and Exploring MapReady

MapReady is a free-and-open software tool provided by the Alaska Satellite Facility that provides some basic SAR data processing capabilities such as reading of SAR data formats, reprojection and geocoding, as well as some polarimetric data manipulations. MapReady can be downloaded from <https://www.asf.alaska.edu/data-tools/mapready/>. Installation instructions are provided in the same location. For further information about MapReady functionality, please consult the MapReady user manual ([http://media.asf.alaska.edu/asfmainsite/documents/mapready\\_manual\\_3.1.22.pdf](http://media.asf.alaska.edu/asfmainsite/documents/mapready_manual_3.1.22.pdf)).

To start MapReady, type **mapready** in your command window. You should see the MapReady interface load (**Figure 1.1**).

##### 1.2 Geocoding a ERS-2 SAR Scene over Fairbanks, AK Using MapReady

ERS-2, a G-band ( $\lambda=5.66\text{cm}$ ) SAR operated by the European Space Agency from 1995 to 2011, has



**Figure 1.1** The ASF MapReady user interface

provided a wealth of Earth observation data, much of which can be accessed through the services of ASF. While the depth of the archive provides a large potential value for a range of user communities, the images of the ERS-2 archive are currently not yet available in fully geocoded formats. Hence, being able to geocode ERS-2 images will help unlock this sensor's vast potential in environmental analysis.

The data to be geocoded is ERS-2 granule E2\_80464\_STD\_F163, which was acquired on September 10 of 2010 over the area of Fairbanks, AK.

##### 1.3 Load the Image into MapReady and visualize the content of the Data Set

Here some instructions for loading and exploring the image:

- To load the image, click the Browse button in the Input Files section of the interface. Pick the E2\_80464\_STD\_F163.D file within the E2\_80464\_STD\_F163 folder and click Open.

To visualize the image, click on the icon labeled with "Preview image" as shown in **Figure 1.2**. A viewer will open, displaying the image as well as metadata information. Scroll around the image. Zoom in to evaluate image noise and structure. Also investigate metadata information on the left side of the viewer interface.

##### 1.4 Geocode and Terrain Correct the Image using MapReady

Apply the following settings to geocode and terrain correct your data:

In the "General" Tab:

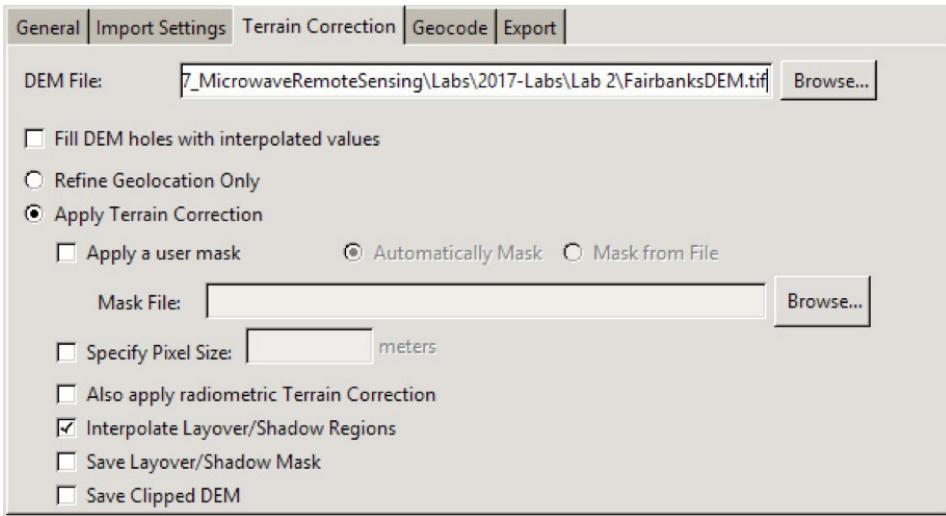
- To terrain correct and geocode the image, activate the Terrain Correct and Geocode to a Map Projection radio buttons in the top element of the interface. The Terrain Correction and Geocode tabs become active.

To separate input data from your processing results, change the Destination Folder settings in the Input Files section of the interface. I recommend the following folder as your destination folder: **/home/ubuntu/SARLabs/SARFocusin-gAndGeocoding/Results**

Navigate to the "Calibration" tab:

This tab allows the chosen calibration procedure to be applied to the data.

- Typically, scientists prefer data in  $\sigma$  projection, which allows relating the brightness in an image pixel to physical quantities. Hence, I suggest picking Sigma as Radiometric projection.



**Figure 1.2** The ASF MapReady “Terrain Correction” Tab

- You can also choose whether to output your image in amplitude or decibel (dB) format. As radar data have enormous dynamic range, converting the pixel values to a dB scale is often recommended.

#### Navigate to the “Terrain Correction” tab:

Please create two geocoded datasets here: (1) a dataset where only geometric terrain correction was applied; and (2) an image where both geometric and radiometric terrain correction was done.

Initially, create the geometrically corrected (GTC) image:

- To Pick a DEM file for terrain correction, click on Browse and pick the file E2\_80464\_STD\_F163\_dem.tif in the Data directory.
- Explore the various options in the geocoding tab. We will discuss those options.

Ensure that the Apply Terrain Correction feature is activated (see **Figure 1.2**).

In a second run (after you complete the rest of the instructions), create the RTC image by:

- Selecting Also apply radiometric Terrain Correction in addition to the previous settings
- Click on Add Output File Prefix or Suffix in the “Input Files” section and add suffix “\_RTC”.

#### Navigate to the “Geocode” tab:

In this tab, you can change geocoding parameters such as the desired projection, the pixel size, and the inter-

polation method. In our case, we will simply accept the default (UTM projection; in default mode, the pixel size is set to half the original image sampling distance).

#### Navigate to the “Export” tab:

This tab allows you to set output formats. Please set the Export format to GeoTIFF and activate the Output data in byte format radio button (see **Figure 1.3**).

Once all of these parameters are set, click on either the Process Individual Image icon (**Figure 1.1**) or the Process All button to start the geocoding and terrain correction process. You can monitor the progress of the procedure in your command window.

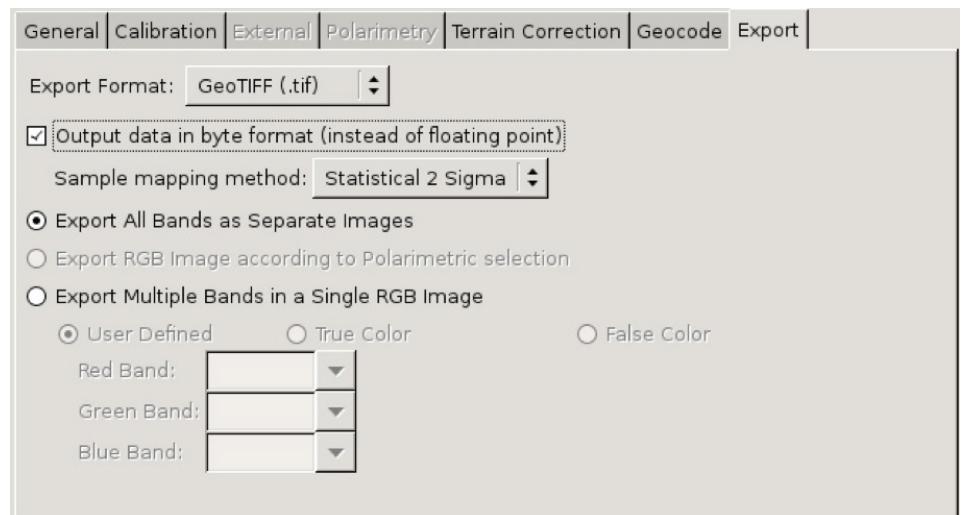
## 1.5 Visualize Geocoded Image in QGIS

Once the geocoding process has completed, you can visualize the product both within and outside of the MapReady tool. To compare the result with map information, we will open the file in QGIS. To do so, run the following command:

```
Qgis /home/ubuntu/SARLabs/
SARFocusingAndGeocoding/
Results/E2_80464_
STD_F163.tif
```

## 2 GEOCODING AND RTC PROCESSING USING SNAP

This lab is for users who wish to generate an RTC image from Sentinel-1 data using easy-to-follow instructions in a graphical user interface (GUI). Specifically, we will use ESA’s Sentinel Application Platform (SNAP) to perform geocoding and RTC processing on Sentinel-1 images over Kathmandu, Nepal. The advantages of the SNAP tool include (1) its graphical user interface, which renders the SNAP tool straightforward to use (compared to other InSAR processing tools); (2) the easy-to-access, free-of-charge, and public domain nature of the SNAP tool; and (3) the fact that SNAP is an integrative multi-sensor toolbox and enables processing data from all Sentinel sensors within one joint processing platform. To install SNAP on your own workstation, please visit <http://step.esa.int/main/download/> for download instructions.



**Figure 1.3** The ASF MapReady “Export” Tab

## 2.1 Sentinel-1 SAR Data Sets Used in this Exercise

The data for this exercise are two Sentinel-1 acquisitions bracketing the devastating 2015 Gorkha earthquake in Nepal, which killed nearly 9,000 people and injured nearly 22,000. It occurred at 11:56 Nepal Standard Time on 25 April, with a magnitude of 8.1Ms and a maximum Mercalli Intensity of VIII (Severe). Its epicenter was east of Gorkha District at Barpak, Gorkha, and its hypocenter was at a depth of approximately 8.2 km (5.1 mi). It was the worst natural disaster to strike Nepal since the 1934 Nepal–Bihar earthquake.

The following data sets, called Ground Range Detected (GRD) images, will be used for this exercise:

- Pre-event image acquired on April 17, 2015:  
S1A\_IW\_GRDH\_1SSV\_20150417T001852\_0150417T001921\_005516\_0070C1\_17AA
- Post-event image acquired on April 29, 2015:  
S1A\_IW\_GRDH\_1SDV\_20150429T001909\_20150429T001934\_005691\_0074DC\_B016

Please download these Sentinel-1 SAR images using ASF's Vertex search engine (<http://vertex.daac.asf.alaska.edu>).

## 2.2 Geocoding and RTC Processing Steps in SNAP

Start SNAP by clicking on the associated desktop icon or by typing in **snap** in your command window.

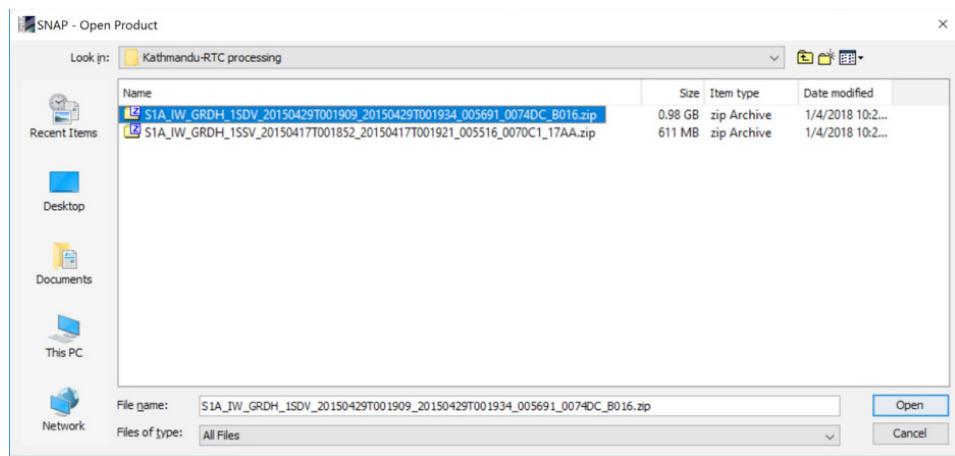
### 2.2.1 Open a SAR image in SNAP

In order to perform geocoding and RTC processing in SNAP, the input products should be one or more GRD or SLC products over your area of interest. While both data types can be processed to RTC images, we are using GRD images in this lab due to their smaller size.

#### Step 1 - open the products:

Use the Open Product button in the top toolbar of the SNAP interface and browse for the location of the Sentinel-1 GRD products (**Figure 1.4**).

Select the \*.zip file containing the post-event image (dated 20150429) and click Open. Press and hold the Ctrl button on the keyboard should you want to select multiple products at a time.



**Figure 1.4** Open Product dialog in SNAP.

#### Step 2 - view the product:

In the Product Explorer (**Figure 1.5**) you will see the opened products. For GRD data, the product band folder will contain two to four layers, depending on whether the data set was acquired in single- or dual-pol (an amplitude and intensity image is provided per polarization). For SLC data, you will find two bands per polarization containing the real (i) and imaginary (q) parts of the complex data.

Note that in Sentinel-1 IW SLC products, you will find three subswaths labeled IW1, IW2, and IW3. Each subswath is for an adjacent acquisition collected by Sentinel-1's TOPS mode. For more information on this mode and on the meaning of the subswath data, please refer to the lab on Interferometric SAR processing in Section 3.

#### Step 3 - view a band:

To view the data, double-click on either the amplitude or intensity band for one of the polarizations (e.g., Intensity\_VV). The image will appear on the right side of

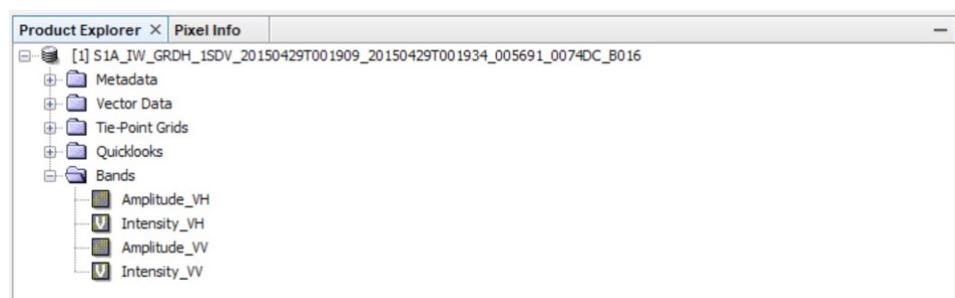
the interface. Zoom in using the mouse wheel and pan by clicking and dragging the left mouse button.

#### 2.2.2 Apply Precise Orbit File

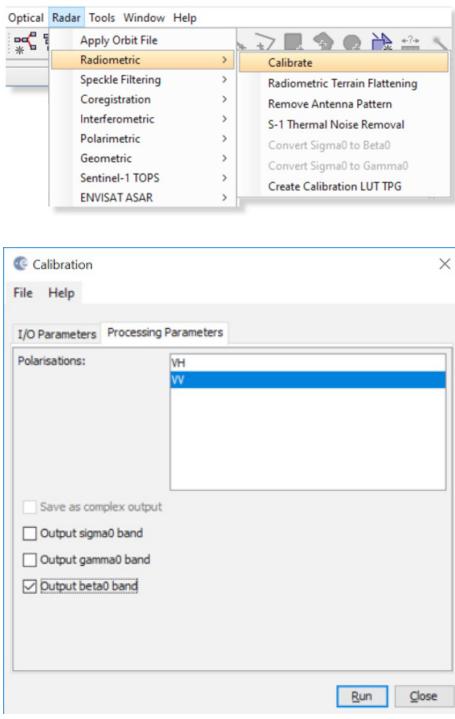
This is an optional step that will maximize the geolocation quality that can be achieved during geocoding. Precise orbit files are issued by the European Space Agency within weeks after the acquisition of a data set. These orbits are not annotated in the image data directly but are rather provided as a separate file. SNAP is able to locate, download, and apply these precise orbit files automatically via the following step:

#### Step 4 – apply orbit file:

To apply precise orbits select Apply Orbit File in the Radar menu of SNAP. A new window will appear (**Figure 1.6**) providing some processing options. Note that the default settings for processing options should work for most applications.



**Figure 1.5** Product Explorer tab within the SNAP user interface.



**Figure 1.6** Calibration interface with relevant options selected.

The only exception to this general rule pertains to the I/OParameters tab where the output directory for the Target Products can be changed from the default to a desired storage location. Click Run to initiate the automatic download and application of files. A window will pop up showing the progress of the processing. Depending on the computing power of your machine, expect one to several minutes of processing time. Once processing is complete, the output from the previous step will appear in the Product Explorer window of the interface (filename ending in “\_Orb”). Single click the file name to select it for the next processing step.

### 2.2.3 Apply Radiometric Calibration

#### Step 5 - calibration:

To correctly apply RTC corrections to the data, the image information needs to be calibrated following the  $\beta^0$  definition. To calibrate to  $\beta^0$ , select the Radiometric

> Calibrate option within SNAP’s Radar menu. In the box that appears, radiometrically calibrate the image to  $\beta^0$  by going into the Processing Parameters tab and selecting the Output beta0 band option (**Figure 1.6**). If dual-pol data are available, you have the choice of processing both polarizations or selecting a subset of polarizations by clicking on the desired channels. Click Run to initiate processing. The defaults place the output into the same directory as the input.

#### 2.2.4 Apply Radiometric Terrain Flattening (RTC Processing)

RTC processing is referred to as “Radiometric Terrain Flattening” in the SNAP tool. This step will remove most of the radiometric distortions from the data that are introduced by surface topography.

#### Step 6 – RTC processing:

To apply RTC processing, first select the output of the previous processing step (extension “\_Orb\_Cal”) in the Product Explorer window. Then, select Radiometric > Radiometric Terrain Flattening from SNAP’s Radar menu (see below). The default settings download a digital elevation model (DEM) matching the geolocation of the scene being corrected, placing the output file into the same directory as the input. Most applications will not require a modification of the default settings. Click Run (~45 minutes or longer, depending on system capability).

Note that an internet connection is necessary for this step as the DEMs necessary for processing are downloaded from an online repository.

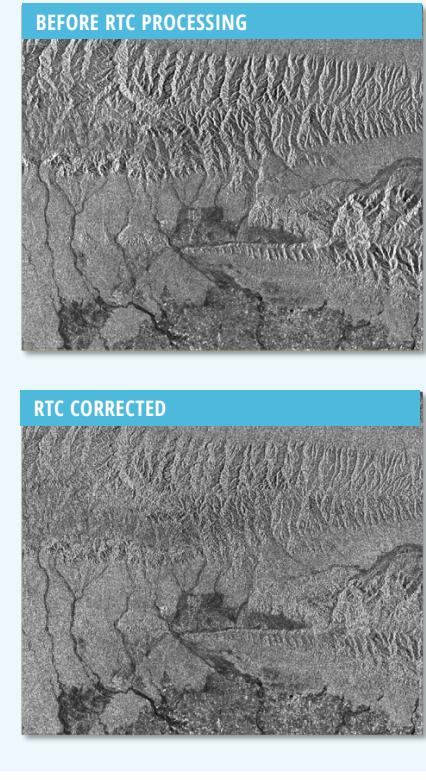
#### Potential necessary intermediary step – Multilooking:

Depending on the resolution of the DEM that can be found for your area of interest, the SAR data may have to be multilooked (reduced in resolution) before processing. If the DEM is of lower resolution than the SAR data, SNAP will enforce multilooking to the resolution of the DEM before RTC processing can be applied. To multilook your imagery, select the data set ending in “\_Orb\_Cal” in the Product Explorer window and then select Multilooking from SNAP’s Radar menu (found on the very bottom of the menu). In the emerging window, select the desired number of looks within the Processing Parameters tab and click Run. Once complete, use the output from this step (file ending in “\_Orb\_Cal\_ML”) as the input for Step 6 – RTC Processing.

## Small Data Analysis Exercise

It may be instructional to compare the SAR image data before and after RTC processing. Such a comparison will provide you with information both on the benefits and limitations of RTC correction for your area of interest.

To conduct a comparison, open both the image before (extension “\_Orb\_Cal\_ML”) and after (extension “\_Orb\_Cal\_ML\_TF”) RTC correction in the SNAP viewer by double-clicking the image bands in the respective data sets. Click on the symbol to synchronize views across multiple image windows and zoom into an area of interest (preferably an area with significant topography). Then toggle between images and compare content. You should see that most of the topographic shading was removed by the RTC processing step. Residual topography is mostly due to limitations in the resolution of the DEM and the small incidence angle dependence of  $\sigma^0(\theta_i)$ . An example of the performance of RTC correction is shown below. A significant (albeit not perfect) reduction of topographic shading was achieved.



## 2.2.5 Geocode the RTC-Corrected Data

### Step 7 – Geocoding:

Unfortunately, the nomenclature that is used in SNAP for the geocoding step is a bit opaque. You will have to pick Geometric > Terrain Correction > Range-Doppler Terrain Correction from the Radar menu to apply the geocoding procedure (**Figure 1.7**). Select the output of the RTC processing step (file ending in “\_Orb\_Cal\_ML\_TF”) as input for the geocoding procedure.

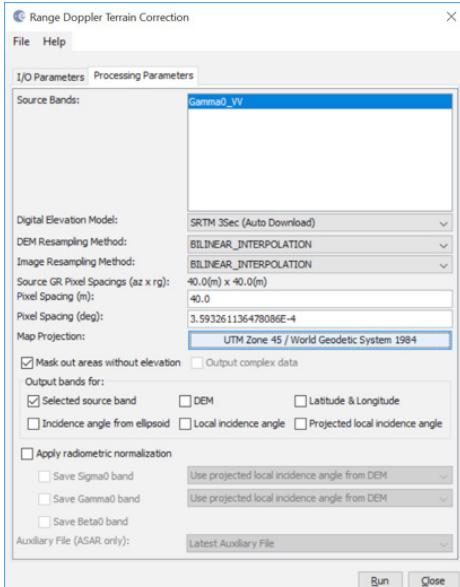
The processing box pops up, and the defaults for the I/O Parameters tab place the output files in the same directory as the source file. The Processing Parameters tab enables you to specify the map projection you need, pixel spacing if you wish to change it, and options for additional output files.

For the sake of this exercise, it is recommended to use the default options for output files but select “UTM / WGS 84” as your output map projection. Click Run after your settings are applied.

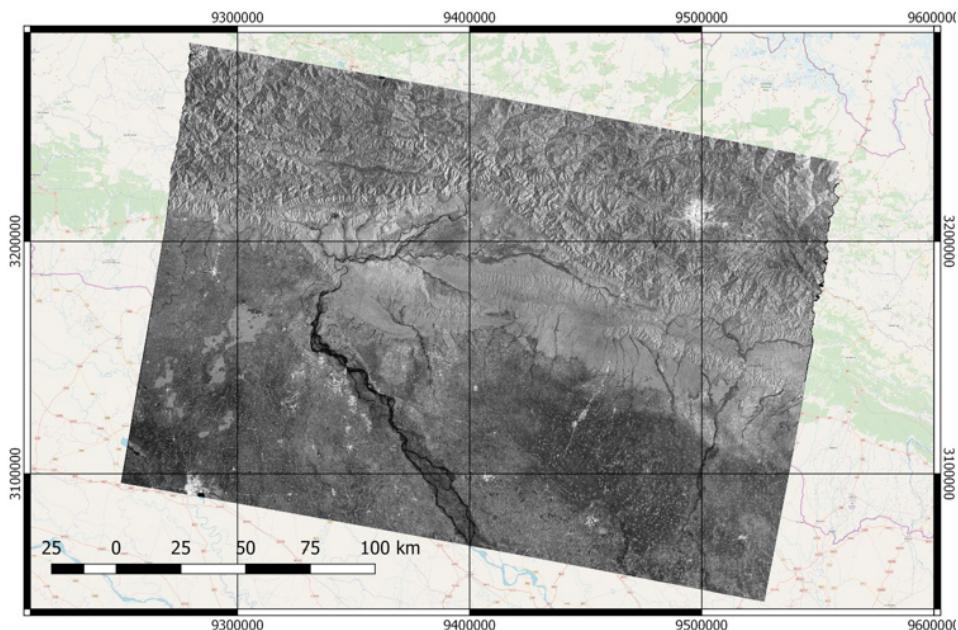
## 2.2.6 Visualizing Processing Results

The products of this processing flow can be visualized easily both within SNAP and within a GIS system of your choosing (e.g., ArcGIS or QGIS).

To view within SNAP, double click the generated file (ending in “Orb\_Cal\_TF\_TC”) in the Product Explorer window and explore within the SNAP interface.



**Figure 1.7** Geocoding interface in SNAP with relevant processing settings applied.



**Figure 1.8** Geocoded and RTC corrected Sentinel-1 SAR image over Kathmandu, Nepal.

To view your RTC image within a GIS, follow these steps: (1) open ArcGIS or QGIS; Select Add Data (ArcGIS) or Add Raster Layer (QGIS); (3) Navigate to the directory that contains the output from Step 7; (4) Within this folder, click on the sub-folder ending in “\_TC.data”; (5) Load the .img file(s) contained within. **Figure 1.8** shows the geocoded and RTC corrected image in QGIS.

## 2.2.7 Visualizing Processing Results

A summary of the geocoding and RTC processing steps is provided in **Figure 1.9**. The following links may be useful in case you want to dive deeper into the topic of geocoding and RTC processing:

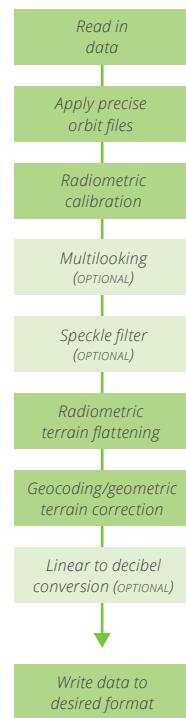
a) To learn a bit more about the theory behind geocoding and RTC processing, please visit Lecture 9 of UAF’s [Online Class on Microwave Remote Sensing](#). You can find Lecture 9 in Class Module 2 “[Imaging Radar Systems](#)”. To go directly to the slide deck, [click here](#).

b) To learn how to Radiometrically Terrain Correct (RTC) Sentinel-1 Data Using SNAP Scripting Languages, please visit ASF’s [SAR data recipe](#) on this topic.

c) For instructions on how to do geocoding and RTC processing using the GAMMA RS software, [please go here](#).

d) For instructions on how to geocode (no RTC) Sentinel-1 data using GDAL, [go here](#).

e) For information on how to effectively view RTC data in a GIS Environment, [go here](#).



**Figure 1.9** General workflow of geocoding and RTC processing.

## Small Data Analysis Exercise

As an additional exercise, geocode and RTC process the second data set (pre-earthquake image). After both data sets are available in geocoded and RTC corrected form, visually compare the images and see if you can identify changes in Kathmandu that might indicate earthquake damage.

### Pre-event image acquired on April 17, 2015:

- S1A\_IW\_GRDH\_1SSV\_20150417T001852\_20150417T001921\_005516\_0070C1\_17AA

### Post-event image acquired on April 29, 2015:

- S1A\_IW\_GRDH\_1SDV\_20150429T001909\_20150429T001934\_005691\_0074DC\_B016

## 3 INSAR PROCESSING USING SNAP

### 3.1 Introduction

In this lab, we will analyze a pair of Sentinel-1 images that bracket the devastating 2015 Gorkha earthquake near Kathmandu, Nepal, whose 7.8 magnitude main shock on April 25 together with several aftershocks (6.9M on April 26; 7.3M on May 12) triggered an avalanche on Mount Everest. 21 people were killed, making April 25, 2015 the deadliest day on the mountain in history. Another huge avalanche was caused in the Langtang valley, where 250 people were reported missing. Hundreds of thousands of people were made homeless with entire villages flattened, across many districts of the country. Centuries-old buildings were destroyed at UNESCO World Heritage Sites in the Kathmandu Valley, including some at the Kathmandu Durbar Square, the Patan Durbar Square, the Bhaktapur Durbar Square, the Changu Narayan Temple, the Boudhanath stupa, and the Swayambhunath Stupa.

**Figure 1.10** shows the USGS ShakeMap associated with the 7.8 magnitude main shock, showing both the violence of the event and the location of the largest devastation.

We will use ESA's Sentinel Application Platform (SNAP) to perform InSAR processing on these Sentinel-1 images. The advantages of the SNAP tool include (1) its graphical user interface, which renders the SNAP tool straightforward to use (compared to other InSAR processing tools); (2) the easy-to-access, free-of-charge, and public domain nature of the SNAP tool; and (3) the fact that SNAP is an integrative multi-sensor toolbox and enables processing data from all Sentinel sensors within one joint processing platform.

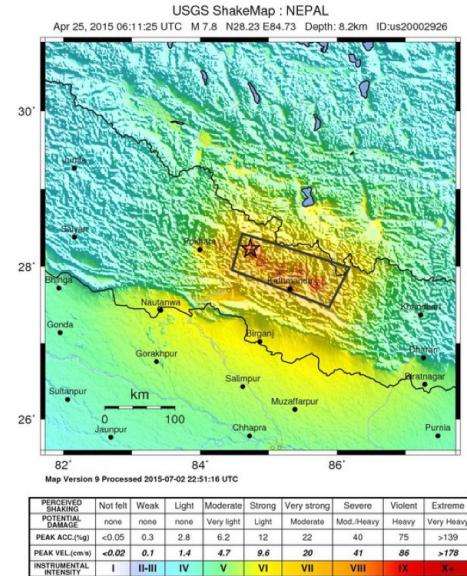
Should you be interested in using SNAP on your own work station, please visit <http://step.esa.int/main/download/> for download instructions.

### 3.2 Sentinel-1 and the 2015 Gorkha Earthquake

We will use a pair of repeated Sentinel-1A images for this lab that were acquired on April 17 and April 29, 2015, bracketing the main- and first aftershock of the Gorkha earthquake event. Hence, the phase difference between these image acquisitions capture the cumulative co-seismic deformation caused by both of these seismic events. The footprint of the Sentinel-1 images (**Figure 1.11**) shows good correspondence with the areas affected by the earthquake (**Figure 1.10**). Hence, Sentinel-1 data are a good basis for studying earthquake-related surface deformation.

SAR data for this exercise can be retrieved via the [ASF Vertex SAR data search client](#).

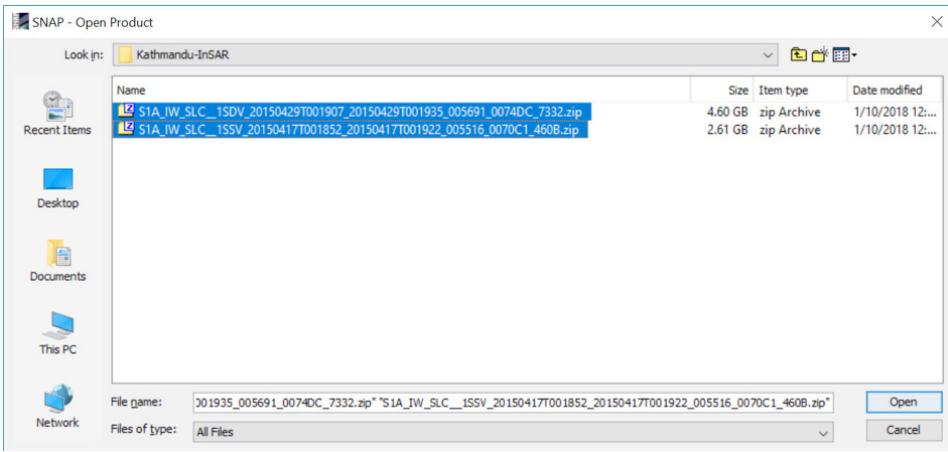
- Pre-event image:  
S1A\_IW\_SLC\_\_1SSV\_20150417T001852\_20150417T001922\_005516\_0070C1\_460B
- Post-event image:  
S1A\_IW\_SLC\_\_1SDV\_20150429T001907\_20150429T001935\_005691\_0074DC\_7332



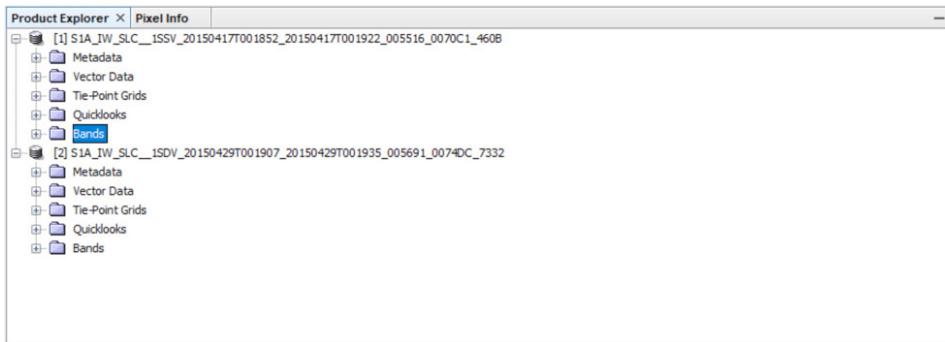
**Figure 1.10** USGS ShakeMap associated with the 7.8 main shock of the 2015 Gorkha Earthquake northwest of Kathmandu, Nepal.



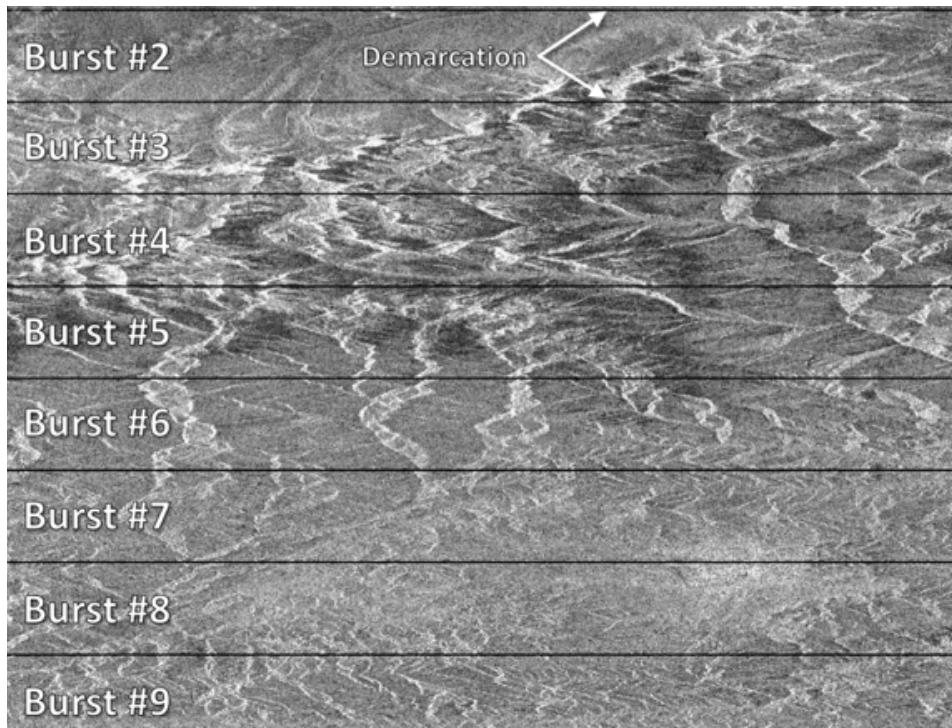
**Figure 1.11** Footprint of the Sentinel-1A SAR data used in this study.



**Figure 1.12** Open Product dialog in SNAP.



**Figure 1.13** Product Explorer tab within the SNAP user interface.



**Figure 1.14** Intensity image of IW1 swath with bursts and demarcation areas identified.

### 3.3 InSAR Processing using the SNAP Tool

Start the SNAP on your computer by either double-clicking on the related icon or by typing **snap** in your command window.

#### 3.3.1 Opening a Pair of SLC Products

In order to perform interferometric processing, the input products should be two or more SLC products over the same area acquired at different times.

##### Step 1 - open the products:

Use the Open Product button in the top toolbar of the SNAP interface and browse for the location of the Sentinel-1 Interferometric Wide (IW) swath products (**Figure 1.12**).

Select the \*.zip files containing the respective Sentinel-1 products that will be used in this lab and press Open Product. Press and hold the Ctrl button on the keyboard to select multiple products at a time.

##### Step 2 - view the products:

In the Product Explorer (**Figure 1.13**) you will see the opened products. Within the product bands, you will find two bands containing the real (i) and imaginary (q) parts of the complex data. The i and q bands are the bands that are actually in the product. The virtual Intensity band is there to assist you in working with and visualization of the complex data.

Note that in Sentinel-1 IW SLC products, you will find three subswaths labeled IW1, IW2, and IW3. Each subswath is for an adjacent acquisition by the TOPS mode.

##### Step 3 - view a band:

To view the data, double-click on the Intensity\_IW1\_VV band of one of the two images. Zoom in using the mouse wheel and pan by clicking and dragging the left mouse button. Within a subswath, TOPS data is acquired in bursts. Each burst is separated by demarcation zones (**Figure 1.14**). Any 'data' within the demarcation zones can be considered invalid and should be zero-filled but may contain garbage values.

#### 3.3.2 Coregistering the Data

For interferometric processing, two or more images must be co-registered into a stack. One image

is selected as the master and the other images are the slaves. The pixels in slave images will be moved to align with the master image to sub-pixel accuracy.

Coregistration ensures that each ground target contributes to the same (range, azimuth) pixel in both the master and the slave image. For TOPSAR InSAR, Sentinel-1 TOPS Coregistration should be used.

#### Step 4 - Coregister the images into a stack:

Select S-1 TOPS Coregistration in the Radar menu. TOPS Coregistration consists of a series of steps including the reading of the two data products, the selection of a single subswath with TOPSAR-Split, the application of a precise orbit correction with Apply-Orbit-File and the conduction of a DEM-assisted Back-Geocoding co-registration. All of these steps occur automatically once the process is kicked off via mouse click (inset at right).

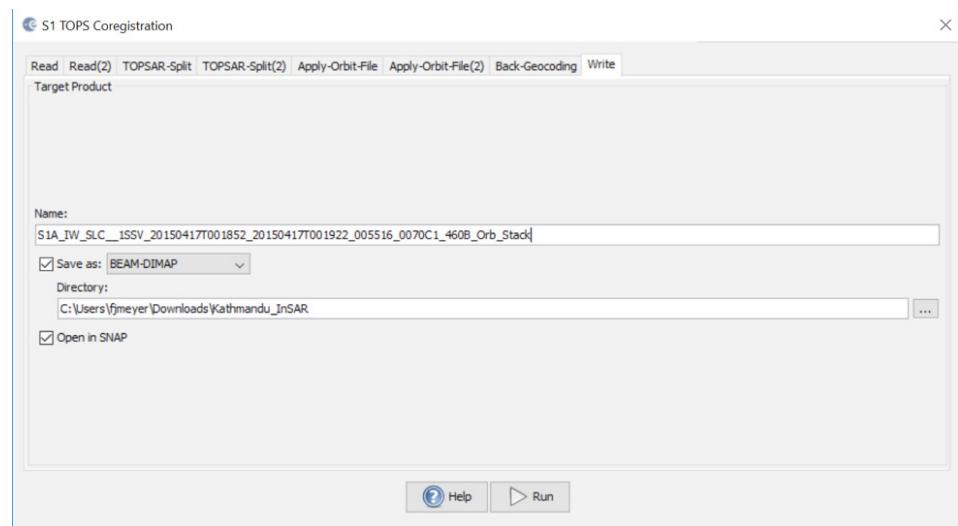
A window will appear allowing you to set a few parameters for the co-registration process (**Figure 1.15**). In the first Read operator, select the first product [1]. This will be your master image. In Read (2) select the other product. This will be your slave image.

In the TOPSAR-Split tab, select the IW1 subswath for each of the products. In the Apply-Orbit-File tab, select Sentinel Precise Orbits. Orbit auxiliary data contain information about the position of the satellite during the acquisition of SAR data. Orbit data are automatically downloaded by SNAP and no manual search is required by the user.

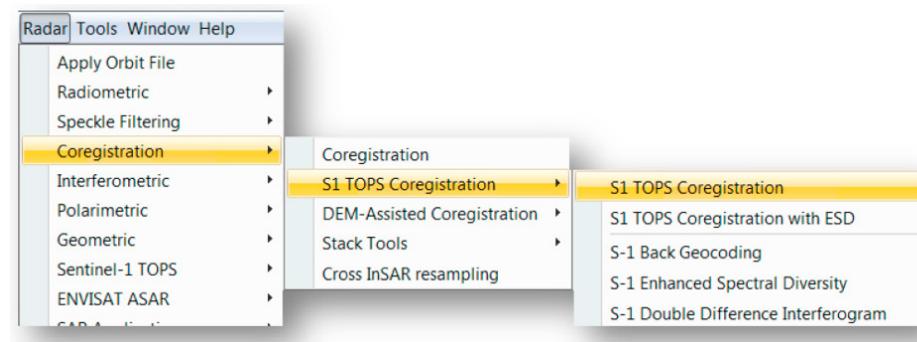
The Precise Orbit Determination (POD) service for SENTINEL-1 provides Restituted orbit files and Precise Orbit Ephemerides (POE) orbit files. POE files cover approximately 28 hours and contain orbit state vectors at fixed time steps of 10 seconds intervals. Files are generated one file per day and are delivered within 20 days after data acquisition.

If Precise orbits are not yet available for your product, you may select the Restituted orbits, which may not be as accurate as the Precise orbits but will be better than the predicted orbits available within the product.

In the Back-Geocoding tab, select the Digital Elevation Model (DEM) to use and the interpolation methods. Areas that are not covered by the DEM or are located in the ocean may be optionally masked



**Figure 1.15** SNAP co-registration interface.



out. Select to output the Deramp and Demod phase if you require Enhanced Spectral Diversity to improve the coregistration.

Finally, in Write, change the Directory path to a preferred location.

Press Process to begin co-registering the data. The resulting coregistered stack product will appear in the Product Explorer tab.

#### *3.3.3 Interferogram Formation and Coherence Estimation*

The interferogram is formed by cross-multiplying the master image with the complex conjugate of the slave. The amplitude of both images is multiplied while their respective phases are differenced to form the interferogram.

The phase difference map, i.e., interferometric phase at each SAR image pixel depends only on the difference in the travel paths from each of the two SARs to the considered resolution cell.

## Step 5 - Form the Interferogram:

Select the stack ([3] in Product Explorer) and select Interferogram Formation from the Radar/Interferometric/Products menu (see inset A at right). The information contained in the interferometric phase measurement is discussed in Lectures 12 - 14 referenced at the end of chapter 2. Please refer to the Supplemental Material on InSAR and associated lecture notes for further information.

Through the interferometric processing flow we will try to eliminate other sources of error to be left with only the contributor of interest, which is typically the surface deformation related to an event.

The flat-Earth phase removal is done automatically during Interferogram Formation step (**Figure 1.16**). The flat-Earth phase is the phase present in the interferometric signal due to the curvature of the reference surface. The flat-Earth phase is estimated using the orbital and metadata information and subtracted from the complex interferogram.

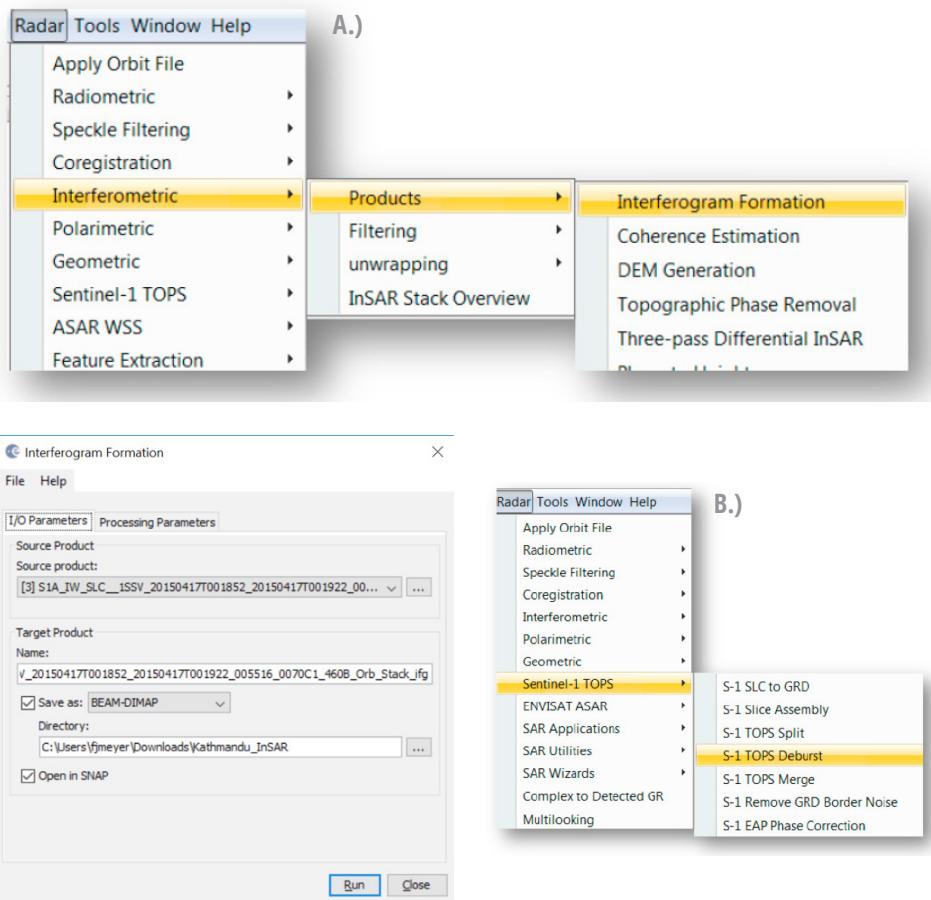
Once the interferogram product is created ([4] in Product Explorer), visualize the interferometric phase. You will still see the demarcation zones between bursts in this initial interferogram. This will be removed once TOPS Deburst is applied.

Interferometric fringes represent a full  $2\pi$  cycle of phase change. Fringes appear on an interferogram as cycles of colors, with each cycle representing relative range difference of half a sensor's wavelength. Relative ground movement between two points can be calculated by counting the fringes and multiplying by half of the wavelength. The closer the fringes are together, the greater the strain on the ground.

Flat terrain should produce a constant or only slowly varying fringes. Any deviation from a parallel fringe pattern can be interpreted as topographic variation.

### *3.3.4 TOPS Deburst*

To seamlessly join all bursts in a swath into a single image, we apply the TOPS Deburst operator from the Sentinel-1 TOPS menu.



**Figure 1.16** Interferogram Formation Interface.

## Step 6 – TOPS Deburst:

Navigate to the the Radar/Sentinel-1 TOPS menu item and select the S-1 TOPS Deburst step (inset B).

### *3.3.5 Topographic Phase Removal*

To emphasize phase signatures related to deformation, topographic phase contributions are typically removed using a known DEM. In SNAP, the Topographic Phase Removal operator will simulate an interferogram based on a reference DEM and subtract it from the processed interferogram.

### Step 7 - Remove Topographic Phase:

Select the Radar/Interferogram/Product menu item and select the Topographic Phase Removal step (inset C, right).

SNAP will automatically find and download the DEM segment required for correcting your interferogram of interest. After topographic phase removal,

the resulting product will appear largely devoid of topographic influence. A separated band showing the topographic phase component simulated based on the DEM is also included.

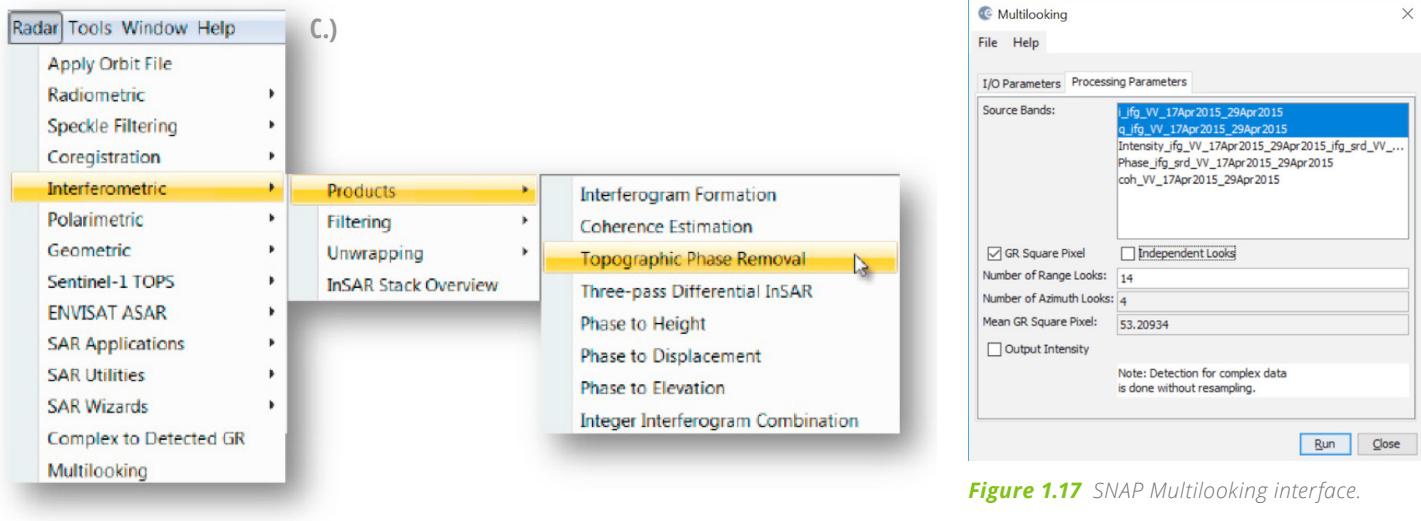
### *3.3.6 Multi-looking and Phase Filtering*

You will see that up to this stage, your interferogram looks very noisy and fringe patterns are difficult to discern. Hence, we will apply two subsequent processing steps to reduce noise and enhance the appearance of the deformation fringes.

As discussed in the previously referenced Lecture 12, interferometric phase can be corrupted by noise related to:

- Temporal decorrelation
- Geometric decorrelation
- Volume scattering
- Processing error

To be able to properly analyze the phase sig-



**Figure 1.17** SNAP Multilooking interface.

tures in the interferogram, the signal-to-noise ratio will be increased by applying multilooking and phase filtering techniques:

#### Step 8 – Multi-looking:

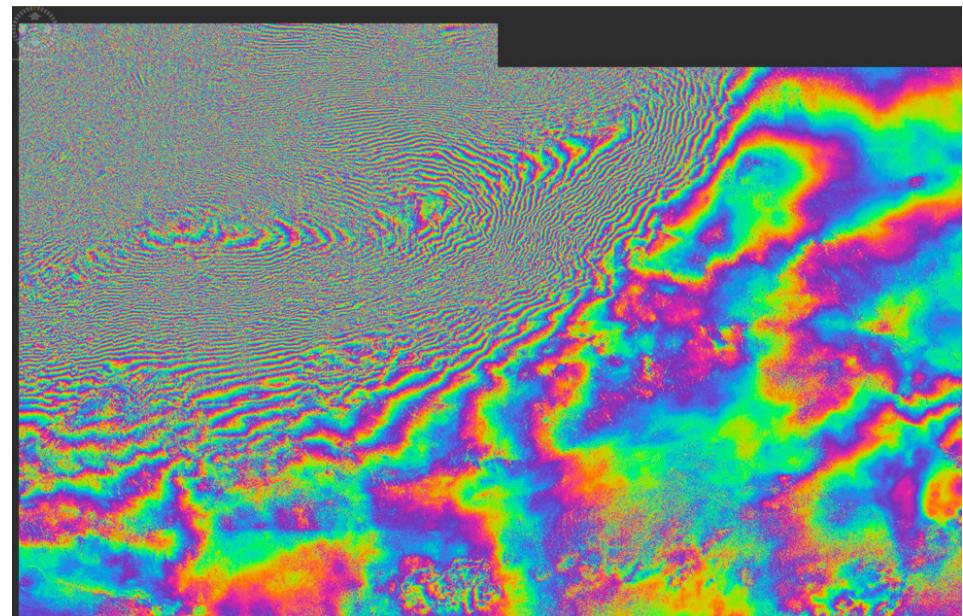
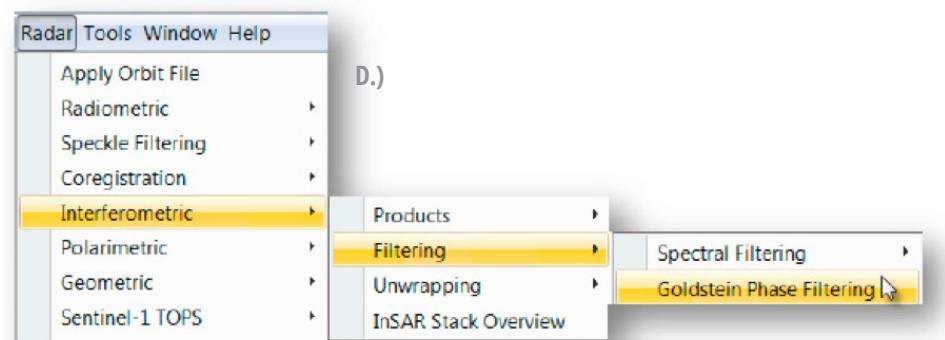
The first step to improve phase fidelity is called multi-looking. To run this step, navigate to the Radar dropdown menu and select the Multilooking option (bottom of the menu). A new window opens. In the Processing Parameters portion of this window, pick the I and q bands as your Source Bands to be multi looked. In the Number of Range Looks field, pick 6 range looks, resulting in a pixel size of about 25m (**Figure 1.17**).

In essence, multilooking performs a spatial average of a number of neighboring pixels (in our case 6x2 pixels) to suppress noise. This process comes at the expense of spatial resolution.

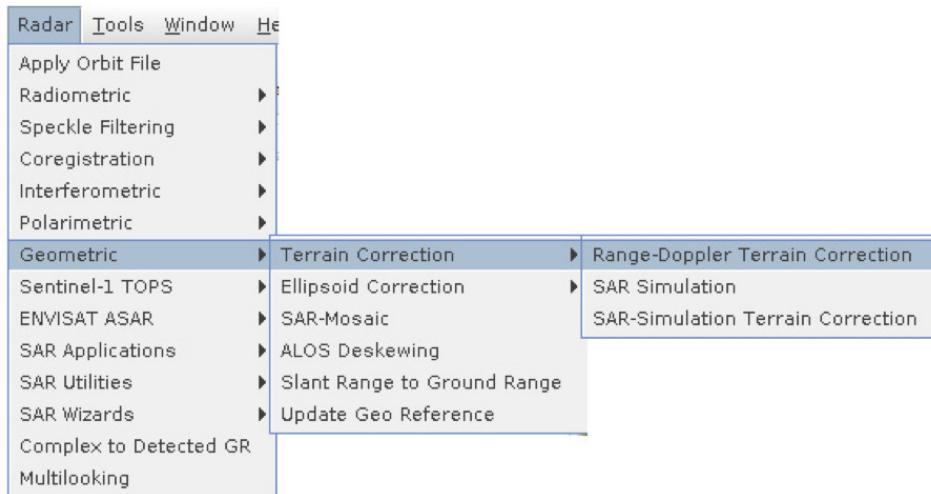
#### Step 9 - Phase Filtering:

In addition to multilooking we perform a phase filtering step using a state-of-the-art filtering approach. For this purpose, navigate to Radar/Interferometric/Filtering and select Goldstein Phase Filtering (inset D).

After phase filtering, the interferometric phase is significantly improved, and the dense earthquake deformation-related fringe pattern is now clearly visible (**Figure 1.18**).



**Figure 1.18** Deformation fringes related to the 2016 Kumamoto Earthquake show clearly after multilooking and phase filtering was applied.



### 3.3.7 Geocoding and Export in a User-Defined Format

To make the data useful to geoscientists, the interferometric phase image needs to be projected into a geographic coordinate system using a DEM-assisted geocoding step.

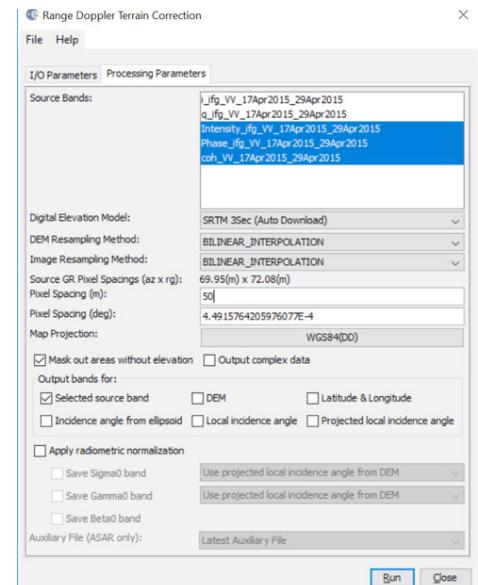
#### Step 10 - Geocoding:

To geocode the interferometric data, navigate to Radar/Geometric/Terrain Correction and select Range-Doppler Terrain Correction (inset above). In the Range-Doppler Terrain Correction window (**Figure 1.19**), select product [8] as source product

and pick the Intensity, Phase, and Coherence images as Source Bands to be geocoded. Adjust the pixel spacing if you want (e.g., 50m). See **Figure 1.20** for the resulting geocoded interferogram of IW1.

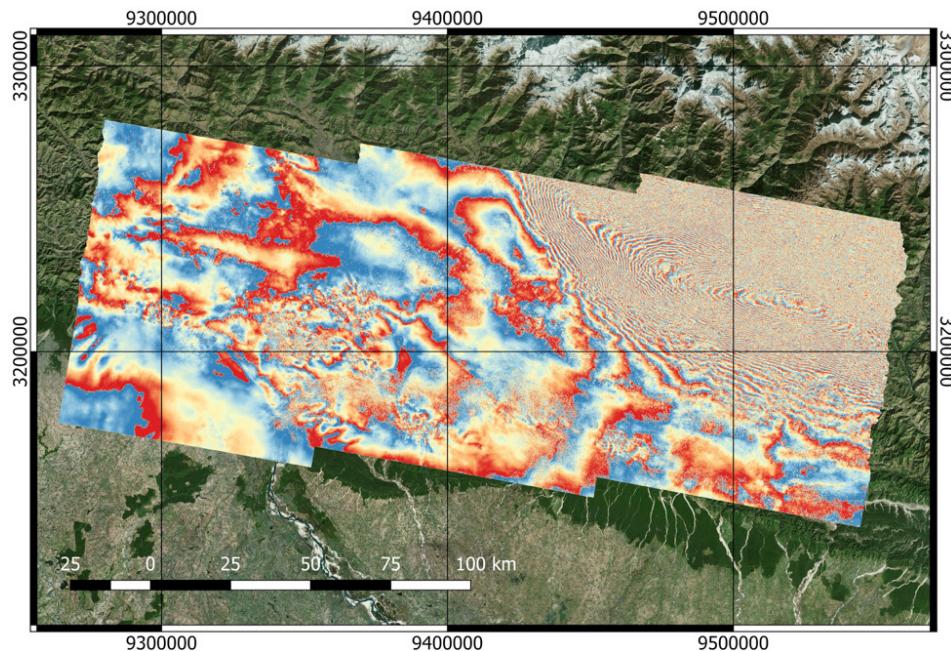
#### Step 11 – Export Data:

The final geocoded data can be exported from SNAP in a variety of formats. To find the export options navigate to File/Export. In addition to GeoTIFF and HDF5 formats, also KMZs and various specialty formats are supported. In addition to the Data Export functionalities, SNAP files can also be directly



**Figure 1.19** SNAP Range-Doppler Terrain Correction interface.

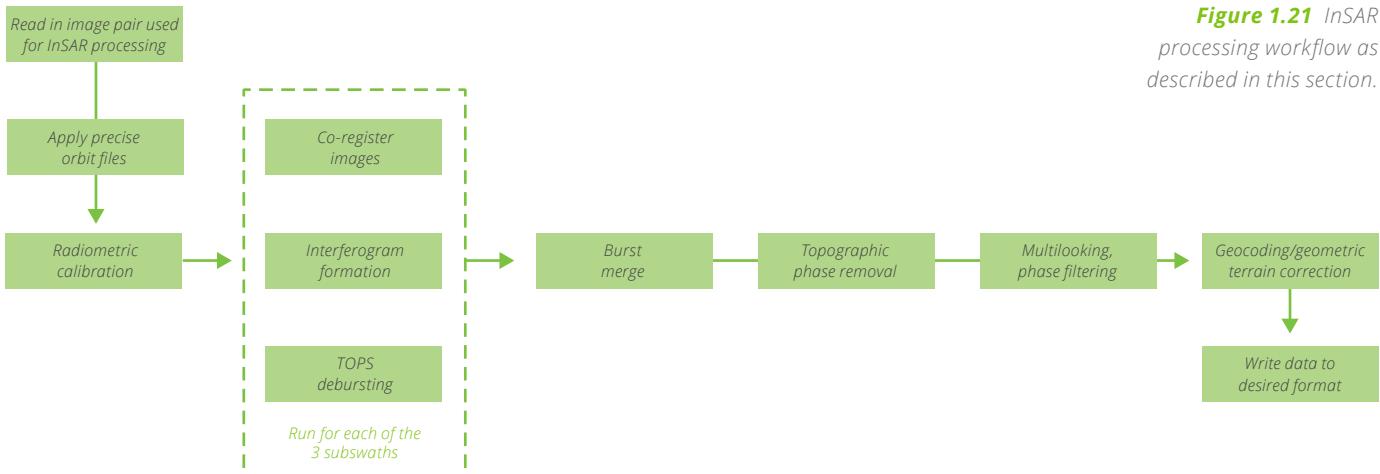
imported into most GIS packages such as QGIS. This is because SNAP uses the established ENVI format for its files, which breaks out each image in a binary data file accompanied by an ENVI-formatted metadata file. **Figure 1.20** shows the processed Gorkha interferogram mapped on top of reference data using QGIS.



**Figure 1.20** Geocoded Gorkha earthquake interferogram mapped in QGIS.

### 3.4 Summary of Processing Steps

A summary of the InSAR processing steps as described in this section can be found in **Figure 1.21**.



**Figure 1.21** InSAR processing workflow as described in this section.

### Small Data Analysis Exercise: Multi-Swath Processing

We processed sub-swath IW1 in the preceding exercise. You can extend this work by processing and merging multiple swaths. Specifically, you can create a geocoded differential interferogram by merging sub-swaths #1 and #2.

#### Step 1: Create a Geocoded Differential Interferogram of the Gorkha Earthquake by Merging Subswaths #1 through #3

To create this merged product, repeat the processing chain for the remaining two subswaths starting from Step 4:

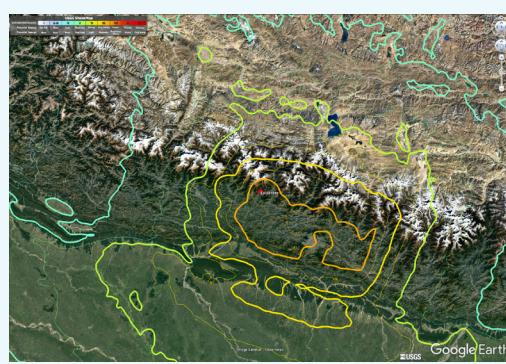
- Run “Step #4 – Coregistration” again but this time select IW2 (or IW3) in the TOPS Split Operator tab → coregistered InSAR pair for sub-swath IW2 (IW3) [Note: make sure to create a new filename under the “Write” tab to no overwrite the IW1 stack result]
- Run “Step #5 – Interferogram Formation” using the new IW2 (IW3) stack as input → IW2 (IW3) interferogram
- Run “Step #6 – Debursting” for the IW2 (IW3) interferogram → debursted IW2 (IW3) interferogram
- NEW STEP: Run Burst Merge: This step is combining the previously generated “debursted IW1 interferogram” with the newly generated “debursted IW2 interferogram” and “debursted IW3 interferogram”. To run burst merge, go to Radar/Sentinel-1 TOPS menu item and select the S-1 TOPS Merge step. Select the debursted IW1, debursted IW2, and debursted IW3 interferograms as inputs.
- Run Steps #7 - #11 for this merged product.
- Produce an image of the merged differential interferogram overlaid on Google Earth or on a QGIS basemap (see **Figure 1.20**).

#### Step 2: Compare InSAR Data to ShakeMap Information

The Earthquake Hazards Program of the U.S. Geological Survey is providing a wealth of information about all significant earthquakes around the globe. The ShakeMap® is one of many sets of information included in this USGS feed. It was developed by the USGS to facilitate communication of earthquake information beyond just magnitude and location. By rapidly mapping out earthquake ground motions, ShakeMap portrays the distribution and severity of shaking.

To access the ShakeMap for the 2015 Nepal Gorkha earthquake event, please visit <https://earthquake.usgs.gov/earthquakes/eventpage/us20002926#executive> and download the Event KML (you will find the KML download link on the bottom of the menu on the left side of the website).

Overlay the ShakeMap onto your interferogram and analyze how well the map conforms with the interferogram. Where do the two data sets match up? Are there places where they don’t match up?



## 4 A SIMPLE AMPLITUDE CHANGE DETECTION TECHNIQUE USING SNAP/ MAPREADY AND GIS

### 4.1 Introduction

Due to their 24/7 observation capabilities, SAR data are relevant for a broad range of applications in environmental monitoring and emergency response. This lab will touch on three examples of how SAR can be used to analyze various kinds of changes on the ground. Examples will include (1) repeated images over Huntsville, AL (or Fairbanks, AK), for the detection of environmental change; (2) imagery over Altamira, Brazil, a stronghold for illegal logging in the Amazon rainforest; and (3) data over Livingston Parish, LA, documenting the 2016 Louisiana flooding event.

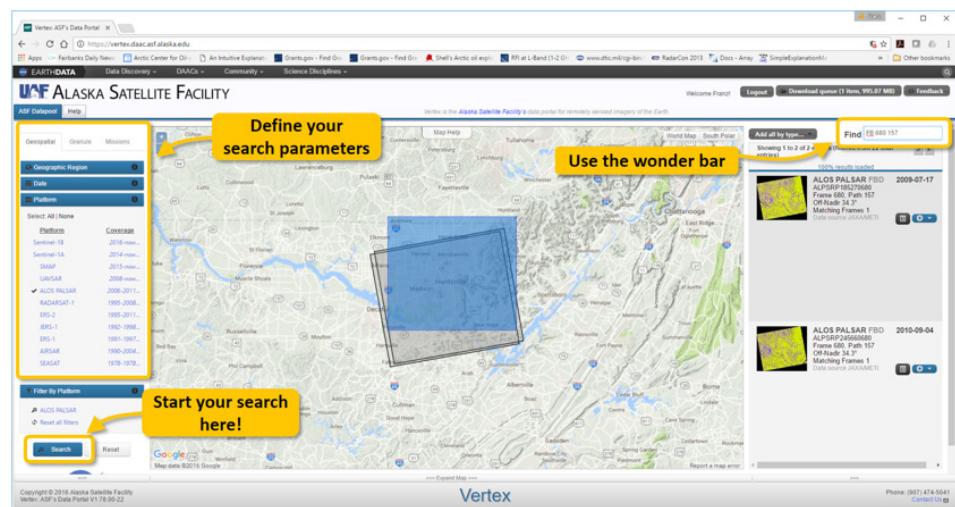
Processing of these data will be done in QGIS, and emphasis will be put on simple, yet effective processing techniques. While this lab can be done within the cloud-based Virtual SAR Lab, which is available to you for these exercises, it might be more effective to process the data locally.

### 4.2 Detecting Changes in and Around Huntsville, AL / (alternatively Fairbanks, AK) from a Pair of ALOS PALSAR RTC Images

The goal of this exercise to detect environmental changes around Huntsville, AL, Fairbanks, AK, or another area of interest through the observation period of the ALOS PALSAR SAR sensor system (2006 – 2011). A secondary goal is to demonstrate how quickly the ALOS PALSAR RTC products can be brought into a GIS system to aid in a geospatial analysis.

#### 4.2.1 Identify Suitable ALOS PALSAR RTC Images for Change Detection using Vertex

To identify ALOS PALSAR RTC images suitable for change detection, go to the ASF Vertex search engine (<http://vertex.daac.asf.alaska.edu>) and draw a box centered on Huntsville (Figure 1.22) / Fairbanks, AK. Specify the following search settings to find suitable data:



**Figure 1.22** Setting up a search mask around Huntsville, AL in Vertex.

#### Step 1 – Set Geographic Search Region:

Draw a bounding box on the Vertex map to include your study area.

#### Step 2 – Set Seasonal Search Range:

Setting a seasonal search range will limit your search to images from the same season. This is important for change detection operations as it avoids seasonal changes and focuses on true environmental changes in a change detection analysis.

- In the Search Tab on the left-hand side of the Vertex interface, activate the Seasonal Search radio button and set the search range to July to September.

Also, search for the time span of the ALOS PALSAR mission by setting the year range to 2006 – 2011 (see Figure 1.23a).

#### Step 3 – Select Platform:

This is to select the sensor of interest from the list of available sensors. Deactivate Sentinel-1A and -1B and activate ALOS PALSAR in the Platform selection interface (see Figure 1.23b).

#### Step 4 – Start Search and Down select Search Results:

Start the Vertex search by clicking on the Search button. Once the search is completed, down-select your results using the wonder bar (see Figure 1.22).

A.)



B.)

Select All   None	
Platform	Coverage
Sentinel-1B	2016-now...
Sentinel-1A	2014-now...
SMAP	2015-now...
UAVSAR	2008-now...
<input checked="" type="checkbox"/> ALOS PALSAR	2006-2011...
RADARSAT-1	1995-2008...
ERS-2	1995-2011...
JERS-1	1992-1998...
ERS-1	1991-1997...
AIRSAR	1990-2004...
SEASAT	1978-1978...

**Figure 1.23** Selecting date range (a) and platform type (b) on Vertex.

Once you have identified the images you are interested in, click on their preview image to enlarge the product info tab and download the Hi-Res Terrain Corrected data set via a mouse click per image of interest (**Figure 1.24**).

Every downloaded file is encapsulated in a zip file. Expand the files into separate folders. Once that's done, we are ready for an analysis of these scenes in a GIS system (QGIS or ArcGIS).

#### 4.2.2 Load Images into QGIS and Visualize

Load the downloaded images into QGIS (or ArcGIS) using the raster import feature (we will walk through the process in this exercise should you be unfamiliar with QGIS).

Ideally, load a background basemap image to be able to compare image features to known landmarks. I suggest loading the OpenLayers Plugin into your QGIS system via the Plugins/Manage and Install Plugins ... item in the menu bar at the top of the QGIS interface.



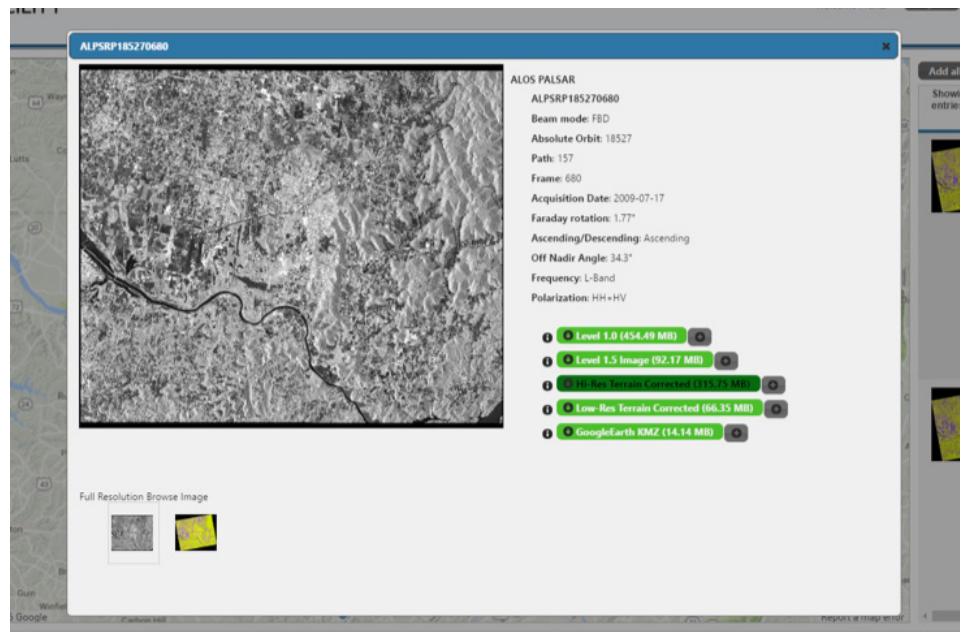
Visualize the images and explore the data set. To improve visualization at all spatial scales, apply a few changes to the image properties:

- Open the property editor by right clicking on an image in the image list and selecting Properties. The Layer Properties interface will appear.
- In the Style tab, change the resampling method for Zoomed in to Cubic and for Zoomed out to Average.

Analyze the data for geolocation quality, spatial resolution, and image content.

#### 4.2.3 Perform a Simple Change Detection Procedure in QGIS

The two images in your list were acquired approximately one year apart. As the data are seasonally coordinated, differences between the images



**Figure 1.24** Image info view in Vertex. Download the High-res Terrain Corrected product per image of interest.

should largely be due to environmental changes between the image acquisition times, such as urban development, changes in river flow, or differences in agricultural activity.

##### Step 1 – Perform visual change analysis:

Flicker between the images to try to identify changes.

- Which changes can you identify?
- How difficult is it for you to identify differences between images?
- What makes the identification of change difficult?

##### Step 2 – Perform Log-ratio scaling, a simple change detection routine:

Identifying changes in images with complex content (e.g., the complex landforms and urban structures in these PALSAR scenes) is hard, as the image content is masking the signatures of change. Our eyes and minds are overwhelmed and distracted by the wealth of information in the data.

Hence, the main goal of image-based change detection approaches is to effectively suppress the image background information, while preserving

the main change signatures of interest. A simple and effective change detection approach is the so-called log-ratio scaling method. It is based on a differential analysis of repeated images and has been shown to be effective in background suppression and change features enhancement. The callout on the next page provides a bit more details on this change detection method.

## Side-note: A Few Words on Log-Ratio Scaling

Log-ratio scaling is an effective means to suppress image background and enhance the change signatures in an image. To identify potential surface changes from SAR data using this approach, a ratio image is formed between a newly acquired image  $X_i$  and a reference data set  $X_p$ . Using ratio images in change detection was first suggested by Dekker (1998) and has since been the basis of many change detection methods (Ajadi et al., 2016; Bazi et al., 2005; Celik, 2010; Coppin et al., 2004; Meyer et al., 2015). To minimize the effects of seasonal variations as well as spurious changes of surface reflectivity on the change detection product, the reference image  $X_R$  should be selected in the same season as the newly-acquired image  $X_i$ . Before ratio image formation, all data should be geometrically and radiometrically calibrated. These steps were done by the data provider (ASF) in our case. The ratio image can be modeled as:

$$r = \frac{X_i}{X_R} = xR$$

where  $r$  is the observed intensity,  $x$  is a multiplicative speckle noise contribution, and  $R$  is the underlying true intensity ratio. The ratio image  $r$  has the disadvantage that the statistical distribution of its gray values is highly non-normal and that its multiplicative noise is difficult to remove. Therefore, a logarithmic scaling is applied to  $r$ , resulting in:

$$X_{LR} = \log_{10} \left( \frac{X_i}{X_R} \right) = y + Q$$

where  $y = 10 \log(x)$ ,  $Q = 10 \log(R)$ , and  $X_{LR}$  is the log-scaled ratio data.

### Related Literature:

Ajadi, O. A., Meyer, F. J., and Webley, P. W., 2016, *Change Detection in Synthetic Aperture Radar Images Using a Multiscale-Driven Approach*: *Remote Sensing*, v. 8, no. 6, p. 482.

Bazi, Y., Bruzzone, L., and Melgani, F., 2005, *An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images*: *IEEE Transactions on Geoscience and Remote Sensing*, v. 43, no. 4, p. 874-887.

Celik, T., 2010, *A Bayesian approach to unsupervised multiscale change detection in synthetic aperture radar images*: *Signal Processing*, v. 90, no. 5, p. 1471-1485.

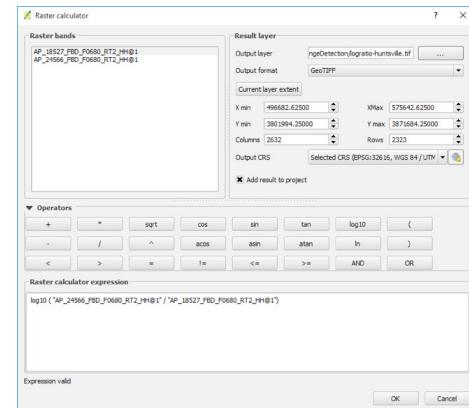
Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., and Lambin, E., 2004, *Review ArticleDigital change detection methods in ecosystem monitoring: a review*: *International Journal of Remote Sensing*, v. 25, no. 9, p. 1565-1596.

Dekker, R. J., 1998, *Speckle filtering in satellite SAR change detection imagery*: *International Journal of Remote Sensing*, v. 19, no. 6, p. 1133-1146.

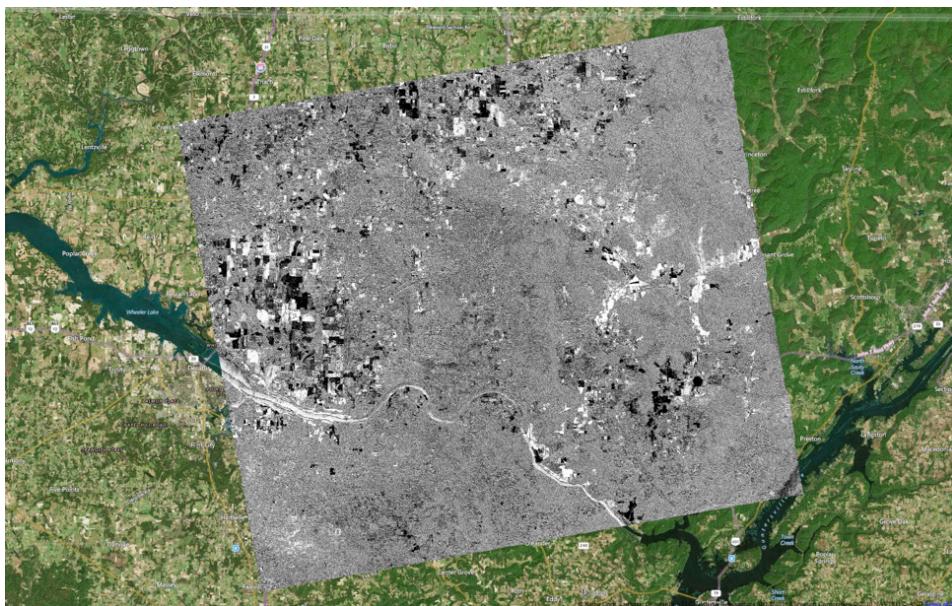
Meyer, F. J., McAlpin, D. B., Gong, W., Ajadi, O., Arko, S., Webley, P. W., and Dehn, J., 2015, *Integrating SAR and derived products into operational volcano monitoring and decision support systems*: *Isprs Journal of Photogrammetry and Remote Sensing*, v. 100, no. 0, p. 106-117.

To conduct log-ratio scaling in QGIS, apply the following procedure:

- In the Raster menu, select Raster Calculator
- In the Raster Calculator window (**Figure 1.25**), construct the following equation:  $\log_{10} ((\text{newer image}) / (\text{older image}))$
- Define an output layer name (e.g., HuntsvillePALSAR-Logratio.tif)
- Click Ok to calculate the log-ratio image.



**Figure 1.25** Raster Calculator Interface and entries for calculating a log-ratio image.



**Figure 7.26** Example log-ratio image for Huntsville, AL.

A screenshot of an example log-ratio image over Huntsville is shown in **Figure 1.26**. This image was created from a pair of images acquired on 7/17/2009 and 9/04/2010. It can be seen that most of the original image content (city of Huntsville, hills and vegetation structures near town, ...) was effectively suppressed from the image. In the log-ratio image, unchanged features have intermediate gray tones (gray value around zero) while change features are either bright white or dark black. Black features indicate areas where radar brightness decreased while in white areas, the brightness has increased.

### Step 3 – Analyze the log-ratio image:

Analyze the change image that you have created. What kind of changes do you see? Compare change features to the base map (e.g., Google Maps) to explain the meaning of observed changes.

## 4.3 Monitoring (Illegal) Logging Activities in the Amazon Rainforest

The region near Altamira, Brazil is one of the most active logging regions of the Amazon rainforest. While some of the logging activities in this area are legitimate, illegal logging operations have flourished over the last decade. Existing logging roads can be clearly identified in optical satellite images such as those used by Google Maps® (**Figure 1.27**). However, frequent rain and cloud cover make change detection based on optical remote sensing data impractical.

### 4.3.1 Retrieve Repeated ALOS PALSAR RTC Images over Logging Areas near Altamira, Brazil

Use the ASF Vertex interface to retrieve repeated images over the logging areas near the Brazilian city of Altamira. When searching for images, don't forget to target similar seasons. Due to the evergreen vegetation in this tropical area, there is no preference for which season you choose.

Once you have identified images of interest, download the High-Res Terrain Corrected images for your change detection analysis. The goal is to identify year-to-year changes in logging extent.

### 4.3.2 Map Logging Activities Using Log-Ratio-based Change Detection Procedures

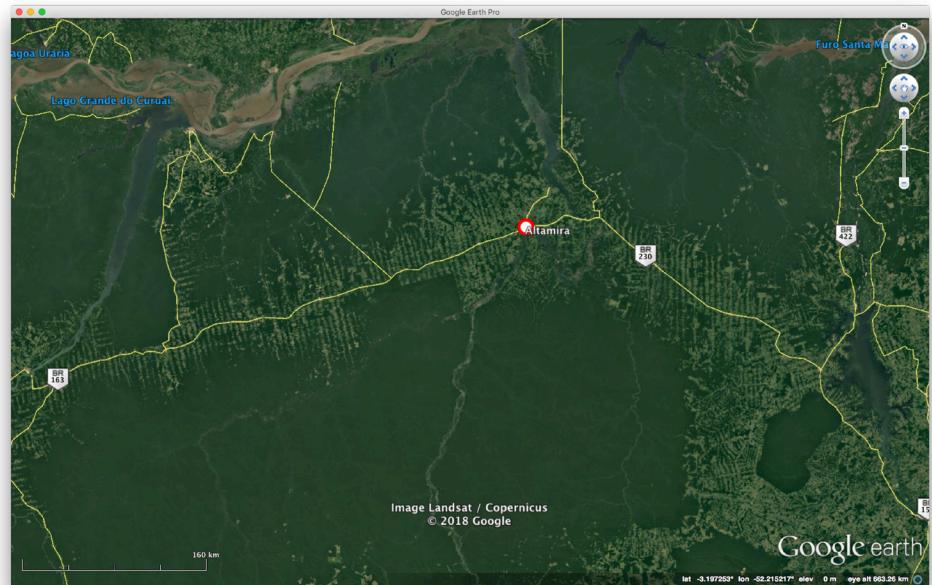
Experiment with Log-Ratio Scaling on your repeated ALOS PALSAR RTC data.

## 4.4 Flood and Inundation Mapping for the 2016 Louisiana Flooding

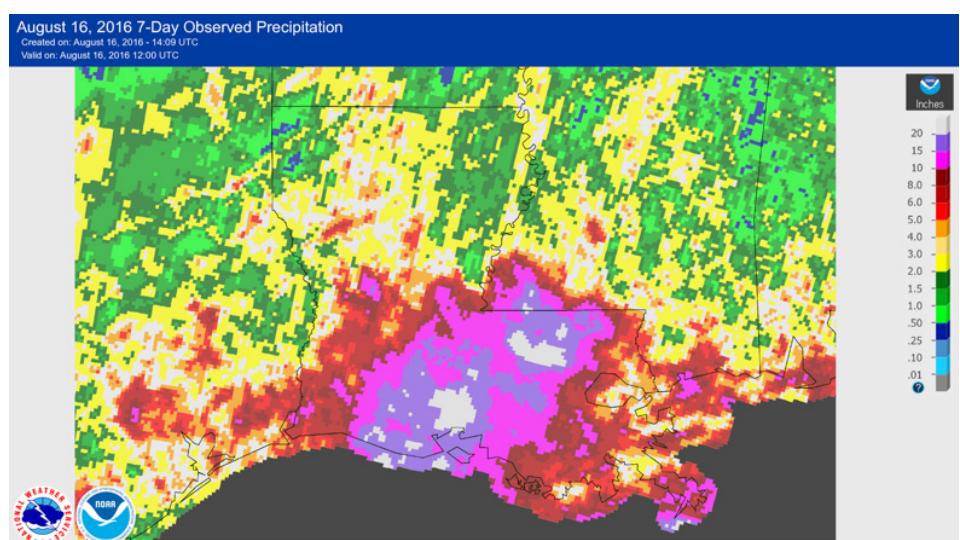
The Louisiana flood of 2016 was touted by many as the biggest U.S. natural disaster since Hurricane Sandy in 2012. In August 2016, prolonged rainfall in southern parts of the U.S. state of Louisiana resulted in catastrophic flooding that submerged thousands of houses and businesses. Many rivers and waterways, particularly the Amite and Comite rivers, reached record levels, and rainfall exceeded 20 inches (510 mm) in multiple parishes (**Figure 1.28**).

### 4.4.1 Meteorological History

Early on August 11, a mesoscale convective system flared up in southern Louisiana around a weak area of low pressure that was situated next to an outflow



**Figure 1.27** Logging roads can clearly be identified from optical satellite data around Altamira, Brazil.



**Figure 1.28** A map of radar-estimated rainfall accumulations across Louisiana between August 9 and 16, 2016; areas shaded in white indicate accumulations in excess of 20 in (510 mm).

boundary. It remained nearly stationary, and as a result, torrential downpours occurred in the areas surrounding Baton Rouge and Lafayette. Rainfall rates of up to 2–3 inches (5.1–7.6 cm) an hour were reported in the most deluged areas. Totals exceeded nearly 2 feet (61 cm) in some areas as a result of the system remaining stationary (**Figure 1.28**).

### 4.4.2 Flood History

Flooding began in earnest on August 12. On Au-

gust 13, a flash flood emergency was issued for areas along the Amite and Comite rivers. By August 15, more than ten rivers (Amite, Vermilion, Calcasieu, Comite, Mermentau, Pearl, Tangipahoa, Tchefuncte, Tickfaw, and Bogue Chitto) had reached a moderate, major, or record flood stage. Eight rivers reached record levels including the Amite and Comite rivers. The Amite River crested at nearly 5 ft (1.5 m) above the previous record in Denham Springs. Nearly one-third of all homes—approximately 15,000 structures—in

Ascension Parish were flooded after a levee along the Amite River was overtopped. Water levels began to slowly recede by August 15, though large swaths of land remained submerged. Livingston Parish was one the hardest hit areas; an official estimated that 75 percent of the homes in the parish were a "total loss". It was thought over 146,000 homes were damaged in Louisiana. This mass flooding also damaged thousands of businesses.

#### 4.4.3 Flood Mapping using Sentinel-1 SAR Data

In this exercise, we will look at the benefits (and limitations) of Sentinel-1 SAR data for mapping the extent of the 2016 Louisiana flood. As mentioned previously, Sentinel-1 is an operational SAR system acquiring images at a routine repeat frequency of either 12 or 24 days for all areas of the globe that were defined as hazard zones. Large parts of the western and central U.S. are part of this hazard map and, hence, are well-covered. The eastern U.S., however, is less well served by Sentinel-1 acquisitions (see [Figure 1.29](#)).

In addition to the coverage of hazard zones, Sentinel-1 is attempting to respond to natural disasters such as the Louisiana flood, through the scheduling of an acquisition on its next pass over this area.

For the Louisiana Flooding event, the following Sentinel-1 data are available:

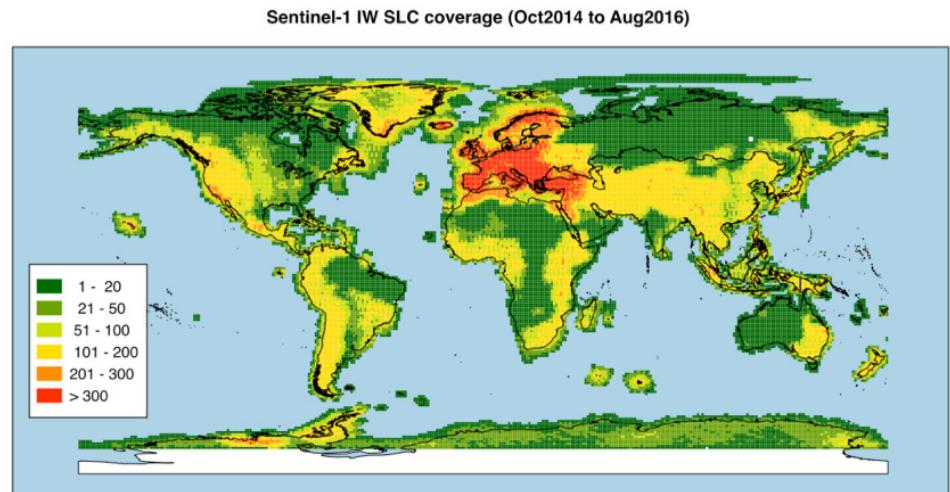
Pre-event data from Aug 7, 2016:

- S1A\_IW\_GRDH\_1SDV\_20160807T000141\_20160807T000210\_012487\_013866\_11BE
- S1A\_IW\_GRDH\_1SDV\_20160807T000210\_20160807T000235\_012487\_013866\_7D4F

Post-event data from August 19, 2016:

- S1A\_IW\_GRDH\_1SDV\_20160819T000211\_20160819T000236\_012662\_013E26\_5A79
- S1A\_IW\_GRDH\_1SDV\_20160819T000142\_20160819T000211\_012662\_013E26\_8995

The footprint of these data is shown in [Figure 1.30](#). The footprint matches the affected areas well. However, the post-event image on Aug 19 might come a bit late to detect the maximum extent of the event.



**Figure 1.29** Global IW-mode coverage of Sentinel-1 between October 2014 and August 2016.

#### 4.4.4 Data Processing

Sentinel-1 data are (currently) available as SLC and as so-called Ground-Range Detected (GRD) products. While the GRD images are georeferenced, neither of these products come fully geocoded and some pre-processing is needed before it is straightforward to work with these data in a GIS system. The images hosted on the website for this lab are pre-processed, using the steps below:

##### A – Extract Image Files:

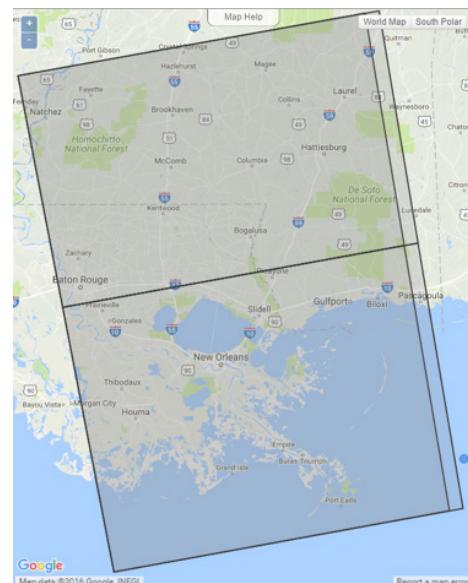
Extract the .tiff file that you are interested in. In each zip file there will be a base directory that is the name of the granule followed by ".SAFE" and then a number of lower level directories. One of these is named measurement, and within this directory will be the georeferenced .tiff files. Using the Unix (or Windows) unzip utility, you can extract only the file you want with a command like this:

```
unzip <filename.zip> */  
measurement/*vv*.tiff
```

This will extract just the VV polarized image from this zip package. If you know which file you want, this is usually much faster than extracting the whole zip file.

##### B – Project the Image Files:

Using the gdal ([www.gdal.org](http://www.gdal.org)) command line utilities you can project and export these files as more useful



**Figure 1.30** Footprint of Sentinel-1 data covering the 2016 Louisiana flooding event.

products. To begin, we want to project the files. The Sentinel-1's GCPs are provided in GCS (lat/lon) coordinates, but can easily be reprojected into another projection. In the case of the Louisiana data, this will be UTM zone 15 (EPSG 32615). To do this, you use the gdalwarp command:

```
gdalwarp -tps -r bilinear  
-tr 10 10 -t_srs  
EPSG:32615 <inputfile.tiff>  
<output-utm-file.tif>
```

Note that:

- t\_srs specifies the output projection
- r specifies the resampling method (default is nearest neighbor)
- tr specifies the output pixel size
- tps specifies to use the thin plate spline technique when interpolating control points

Executing this command will result in a map projected geotiff file that could be read by nearly any GIS application. For Sentinel-1, the datatype will still be uint16.

### C – Scale the data:

Scale to byte using gdal\_translate. You can easily scale the data to byte if you want, which may be more useful.

```
gdal_translate -ot Byte  
-scale 0 700 0 255 <infile.  
tif> <output.tif>
```

In this case, we convert from long int to byte and scale the range 0 to 700 into the 0 to 255 byte range. Values above 700 will be set to 255 and values below 0 (though there shouldn't be any) will be set to 0.

Once these pre-processing steps are completed, the data is ready for analysis in a GIS system.

For the purpose of this training module, these steps have been fully geo-coded and pre-processed. Download the required image data from the website ([radar.community.uaf.edu/lab-8-change-detection-from-sar-images/](http://radar.community.uaf.edu/lab-8-change-detection-from-sar-images/)).

Load the pre-processed data 20160807.tif and 20160819.tif into your GIS system and overlay them on a map. Inspect the images and flicker between them to get a first idea of potential flood extents in the area.

We will apply two different simple yet effective water/flood masking approaches to these data: (1) Image Thresholding and (2) Log-ratio scaling.

#### (1) Image Thresholding:

Image thresholding can be an effective method for the detection of open water bodies in a SAR image.

Here, we use the fact that water is often much darker than the true image content, causing the image histogram to be bimodal and enabling the separation of water from the rest of the image using simple thresholding operations. To conduct image thresholding on both images, please go through the following steps for both data, starting with image 20160819.

#### Step 1 - Perform a log-transformation:

This step creates image histograms for the data that are more Gaussian and simplifies the thresholding operations. In QGIS, go to the Raster menu and select the Raster Calculator. Perform a log transformation by applying an equation to the image (e.g.,  $\log_{10}(\text{"20160819@1"})$ ). Export the image as geotiff under the name 20160819-log.tif.

#### Step 2 – Analyze the image histogram:

Right click on the 20160819-log and select properties. Navigate to the histogram tab and zoom into the histogram to see its shape.

You should see a clearly bi-modal histogram (inset A) with water pixels appearing significantly darker than the main image data.

#### Step 3 – Pick a threshold:

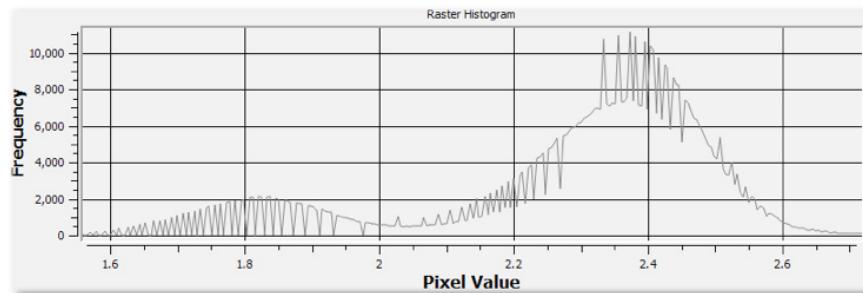
To separate water from the rest of the image, pick a threshold at (or near) the minimum between the two modes of the distribution (e.g., at 2.05)

#### Step 4 – Create a black/white water mask:

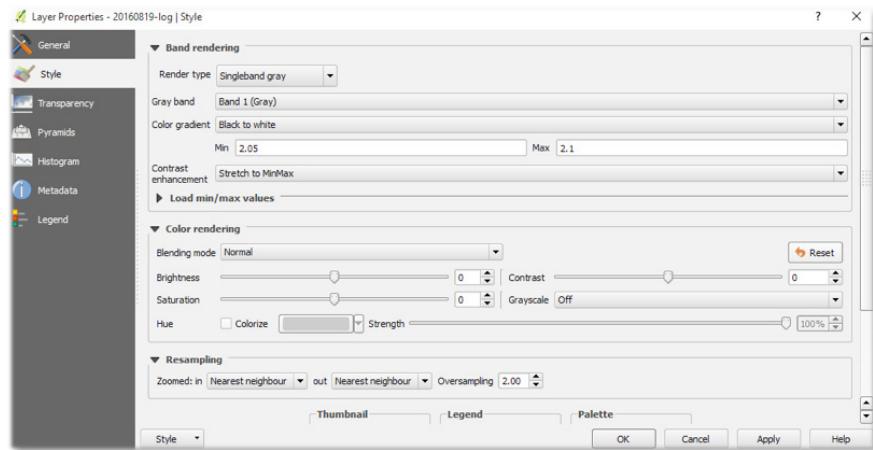
A simple approach is to navigate to the Style tab of the Layer Properties window and rest the minimum and maximum value of the image to values just below and just above the threshold. Click ok or Apply to view the result (inset B). Your image should look similar to the one in **Figure 1.31**.

Repeat this process for image 20160807 and compare the resulting flood masks:

- Analyze the quality of the water masks
- Compare flood mapping results to a base map (e.g., Google maps)
- Think about the benefit of such a map in emergency management situations



A.)



B.)

## (2) Log-Ratio Scaling:

Apply the log-ratio scaling approach from Section 4.2 to the Louisiana flood data. Use the following equation in your analysis:

$$\log_{10}((20160819@1)/("20160807@1"))$$

The result of your analysis should resemble the map in **Figure 1.32**. Analyze the log-ratio map:

- Which signatures can you see?
- What may the bright areas in the log-ratio image represent?
- How does the log-ratio result compare to the water masks created via image thresholding?
- Do you think this may be a useful layer for emergency management situations?

### 4.4.5 Take Away Points

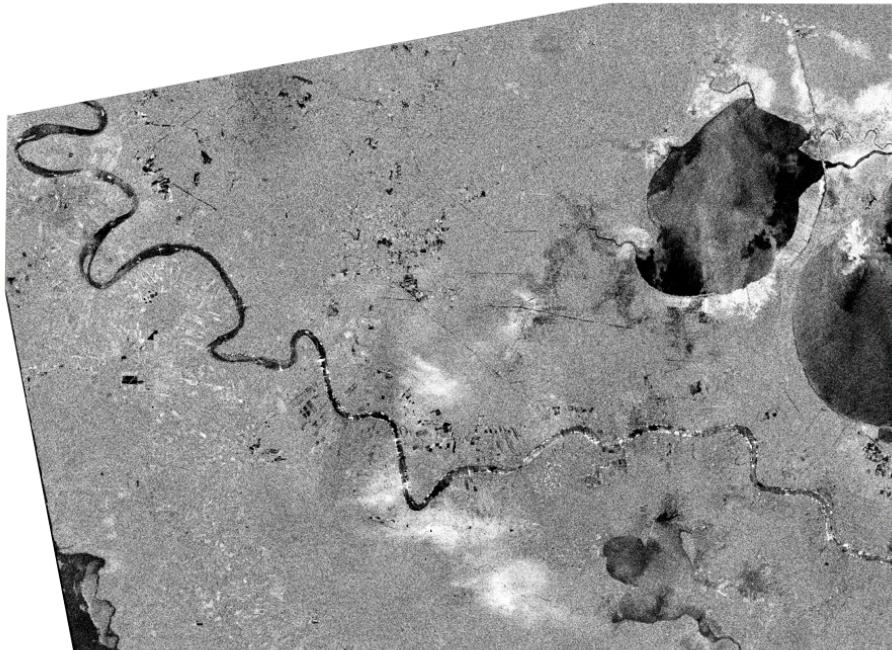
The goal of this last exercise was to demonstrate the potential of SAR data for flood mapping. We learned about two different simple water masking methods and compared them. You should have seen that even very simple techniques can achieve quite impressive results that can have bearing on emergency management situations.

At the same time, this lab should also showcase some of the remaining limitations that still plague SAR data in general and Sentinel-1 SAR data specifically:

- Currently, Sentinel-1 data does not come fully geocoded and some pre-processing is required to be able to manipulate it effectively in a GIS environment.
- While Sentinel-1 is an operational system, it doesn't cover all parts of the globe equally well (even though this is improving as Sentinel-1B is ramping up production).
- The temporal sampling of SAR systems is still limited. In the Louisiana flood example, we are missing the main flood pulse due to inconvenient image acquisition times. By the time of the post-event acquisition, most of the flood water has already receded.
- Freely-available SAR data are also a bit limited in their spatial sampling. In the Louisiana flood case, this may have reduced the quality of flood maps in urbanized environments.



**Figure 1.31** Threshold-based water mask for 20160818.



**Figure 1.32** Log-ratio image for 2016 Louisiana flood





**DR. JOSEF KELLNDORFER'S** research focuses on monitoring and assessing terrestrial and aquatic ecosystems, and disseminating Earth observation data products to policy makers to improve decision making and support capacity building. He is a distinguished visiting scientist at the Woods Hole Research Center and currently serves on various expert working groups within NASA, the Japanese Space Agency JAXA, the Group on Earth Observation, and GOFC-GOLD to advance the use of remote-sensing technology for natural resource mapping and monitoring. He is a member of the NASA Science Team for the US/Indian NISAR satellite. He founded Earth Big Data to provide scalable solutions to modern data-mining challenges.

*Writing this chapter and development of the training material was supported by the NASA SERVIR program. Sentinel-1 and ALOS examples were based on data from the Alaska Satellite Facility and work over many years with the JAXA Kyoto and Carbon Science Team. The following NASA grants supported some analysis of this work: NASA Carbon Monitoring System program, grant number 80NSSC18K0190; NASA NISAR Science Team, grant number 80NSSC18K0087.*

# CHAPTER 3

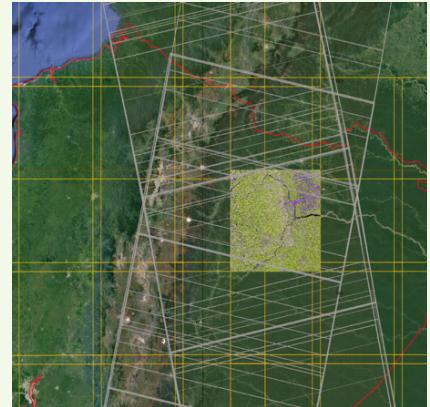
## Using SAR Data for Mapping Deforestation and Forest Degradation

Josef Kellndorfer, President and Senior Scientist, Earth Big Data, LLC

### ABSTRACT

This chapter focuses on Synthetic Aperture Radar (SAR) observations of forest cover change from deforestation and forest degradation. Discussed are SAR backscatter changes determined by sensor and target parameters. Sensor parameters include the wavelength/frequency of the SAR, as well as incidence angle, look directions, and transmit and receive polarization. Since sensor parameters are typically stable from a satellite SAR, backscatter variations over time can be attributed to two main target parameters: structure and moisture. For forests and other targets, this means observations of backscatter change can be linked directly to change in forest structure and moisture conditions of the vegetation and underlying soil. This makes observations with SAR complementary to optical data as (1) almost no atmospheric or Sun illumination variations play a role in SAR response, and (2) longer wavelengths and active penetration into forest canopies interact directly with structure and moisture conditions.

This chapter discusses the influence of sensor and target parameters on backscatter variations from forests and a time series analysis approach for forest change detection. Also discussed are proper methods for SAR data calibration for forest applications, including preprocessing and proper data scaling. Most image examples in this chapter stem from a time series stack of Sentinel-1 data acquired over Ecuador in the Universal Transverse Mercator (UTM) projection tile of the Military Grid Reference System (MGRS), tile number 18MTE (see Fig. 3.1). (The MGRS provides a global tiling scheme with UTM zone number, row designator, and two-letter tile identifier, i.e., 18MTE = Zone 18, Row M, Tile TE. More information may be found [here](#).) The tile is transected by the Napo and Coca rivers on the eastern slopes of the Andes.



**Figure 3.1** Location of the example Military Grid Reference System (MGRS) tile 18MTE in Ecuador used in this chapter.

### 3.1 SAR for Mapping Deforestation and Forest Degradation

As a vital natural resource, forests provide a host of ecosystem services, including carbon sequestration, diverse natural habitats for flora and fauna, and they are a key source of food and fiber for human consumption. Today, many nations have entered international or regional agreements (e.g., the United Nations' Framework Convention on Climate Change - Reducing Emissions from Deforestation or Forest Degradation (UNFCCC-REDD+)) to protect their forest resources. Tracking deforestation rates annually and developing early warning systems of forest loss (often from illegal activities) are essential. Remote sensing of forest change has an important role in this

monitoring effort. While optical data have long been the workhorse for forest monitoring, the advent of operational SAR data availability offers an invaluable complement with a crucial sensitivity: microwave remote sensors are largely cloud-penetrating and thus guarantee continuous monitoring, even under cloudy skies. For tropical nations, this is particularly important as continuous cloud cover severely limits the availability of optical data at medium resolution (Kellndorfer et al. 2014, Mitchell et al. 2017).

### 3.2 Brief Review of Color Theory for Interpreting SAR Images

SAR backscatter images are representations of the microwave portion of the electromagnetic spectrum,

and as such always represent grayscale or false color combinations mapped to the human visual color space. This is analogous to the false color representation of multispectral optical remote sensing imagery from bands outside the visual spectrum. (Please note that in this chapter, "SAR image" shall refer to a grayscale or multi-band image of SAR backscatter, calibrated to  $\gamma^0$  with a Radiometric Terrain Correction (RTC) approach (see [Chapter 2](#))).

#### 3.2.1 GRayscale Display of SAR Imagery

A single-band SAR image (i.e., from one frequency and one polarization) is displayed such that low backscatter values correspond to dark colors and high backscatter values correspond to bright colors. Enhancements can be applied, like linear or histogram stretches. Examples of SAR backscatter images

from Sentinel-1 are shown in **Figure 3.2** for a landscape scale subset in Ecuador and in **Figure 3.3** for a large oil palm plantation just to the north of Puerto Francisco.

### 3.2.2 COLOR DISPLAY OF SAR IMAGERY

For the interpretation of SAR imagery, it is useful to briefly review the basics of how multichannel SAR imagery is displayed. **Tables 3.1 and 3.2** may be used as resources for understanding colors when displaying false color SAR Images (see Henderson & Lewis 1998).

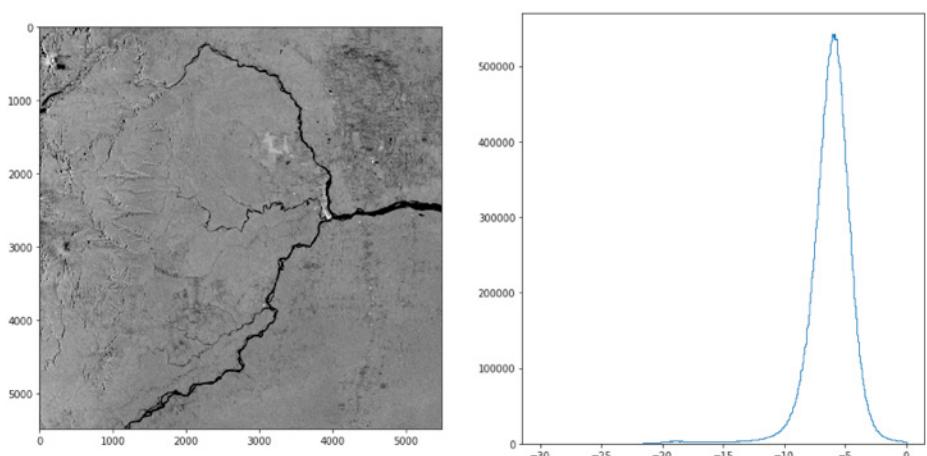
Img Layer 1	Img Layer 2	Img Layer 3	Resultant Color
Tonal Change on Image			
White	Black	Black	Blue
Black	White	Black	Green
Black	Black	White	Red
White	White	Black	Cyan
White	Black	White	Magenta
Black	White	White	Yellow
No Tonal Change on Image			
White	White	White	White
Black	Black	Black	Black
Grey	Grey	Grey	Grey

**Table 3.1** Color assignments and resultant colors for multi-dimensional SAR image composites (*Manual of Remote Sensing*, Vol. 2, 1998).

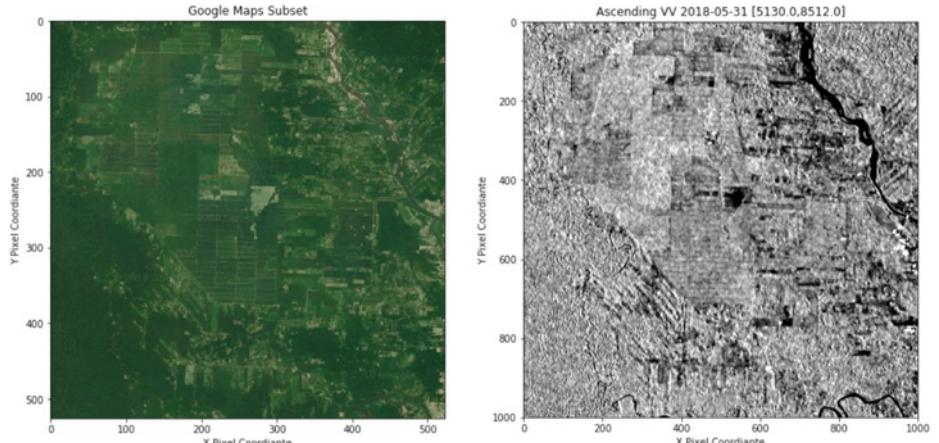
Type of Composite	Assigned Color		
	BLUE	GREEN	RED
Multifrequency/band	Shortest $\lambda$	Middle $\lambda$	Longest $\lambda$
Multitemporal (date)	First (earliest)	Second	Third (Latest)
Multipolarized	Most to Least Common (HH)	(HV/VH)	(VV)

**Table 3.2** Often-used color scheme for multi-dimensional false color SAR composites (*Manual of Remote Sensing*, Vol. 2, 1998).

**Table 3.1** describes how the combination of grayscale imagery assigned to the Red/Green/Blue (RGB) bands would lead to the resulting colors when the extreme dark (black) and bright (white) colors are combined. This is useful when interpreting an RGB



**Figure 3.2** Grayscale Sentinel-1 amplitude image in Ecuador. The area is mostly forested, with the Coca and Napo Rivers, Puerto Francisco, and an oil palm plantation being dark and bright prominent features. The Andes touch the western part of this image. The backscatter histogram in the right panel contains values ranging from about -23 to 0 dB, peaking at about -6 dB.



**Figure 3.3** Google Earth and Sentinel-1 images of a subset of the large oil palm plantation. While the river and most agricultural fields exhibit dark colors, the various states of regrowth in the oil palm plantation correspond to different gray values.

multiprotemporal color image. For example, assume that three dates are combined as per **Table 3.2**, with the earliest acquisition in red, the second acquisition in green, and the newest acquisition in blue. If a red color is seen for a pixel, according to **Table 3.1**, the red layer is close to white (bright backscatter), while the subsequent acquisitions are close to black (dark backscatter). Thus, the backscatter drops after the first acquisition, which is often a sign of deforestation or a degradation event. Note that for forest applications in particular, it is always useful to assign cross-polarized data, which are more related to volume scattering of the canopies to the green band. Co-polarized data

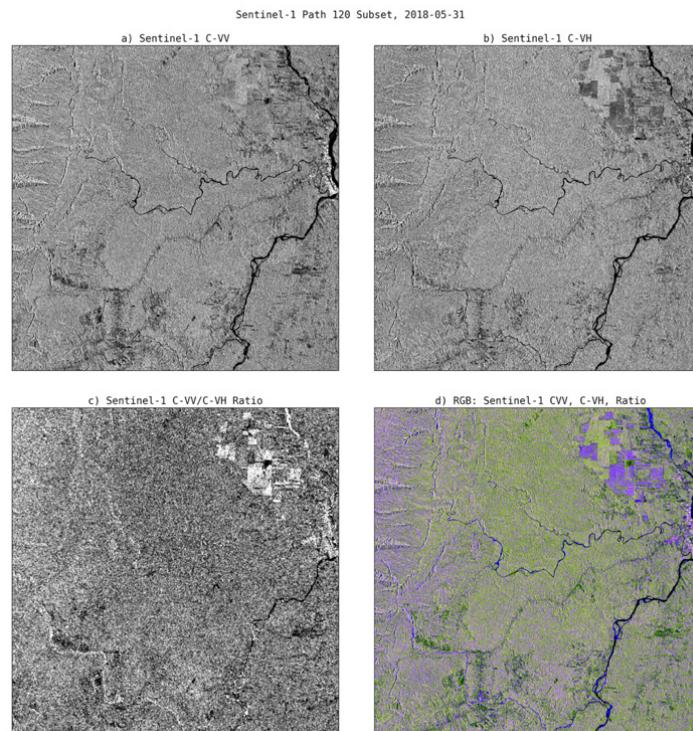
(VV or HH) are suited for the red band, where surface scattering components are more pronounced. When only dual-polarimetric data are available (e.g., L-HH/HV from ALOS, or C-VV/VH from Sentinel-1), a color SAR image is often constructed by assigning the ratio of co-polarized to cross-polarized data to the blue channel. Note that for multi-polarized images with only two polarizations, the co-polarized band is often assigned to red, the cross-polarized to green, and the ratio of co-/cross-polarized data to the blue channel.

Examples for Sentinel-1 C-band and ALOS-1 L-band data are shown in **Figures 3.4** and **3.5**, respectively. The images show the Napo river in the

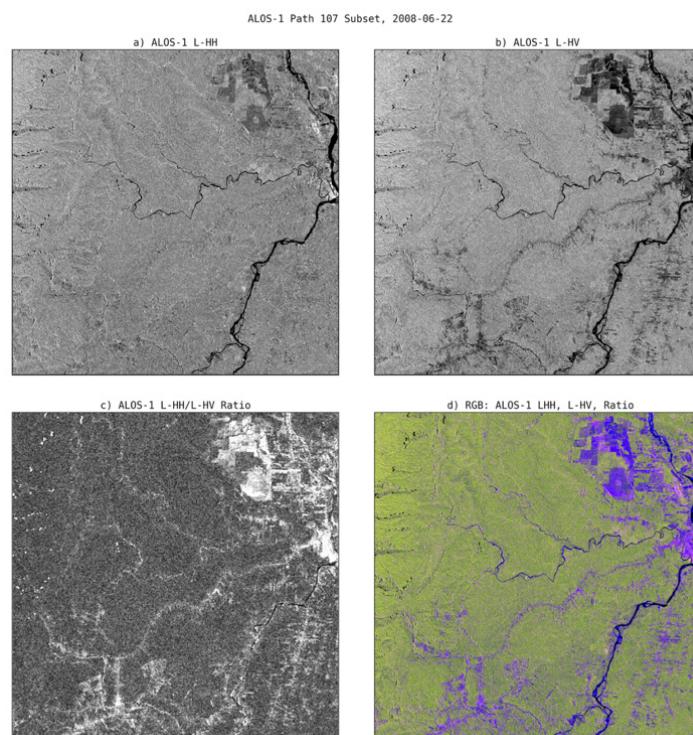
southeast, an oil palm plantation in the northeast, primary rainforest in the northwest, and active fishbone logging patterns in the southwest. The color composites are constructed from dual-polarimetric data with co-polarized data assigned to the red channel, cross-polarized data to the green channel, and the co-/cross-polarized ratio to the blue channel. A nice effect for forest applications with this color assignment strategy is that forests tend to be shown in shades of green, and typically the brightness of green corresponds to the amount of biomass in the forest. Also, water tends to be represented in blue colors, which also represent other surface scattering components. Naturally, different histogram stretches may be applied to enhance various surface components. In these examples, it is remarkable that both C-VV/VH and L-HH/HV false color SAR composites over this predominantly forested landscape exhibit similar color impressions. Differences are notable, however, foremost by the appearance of some dark green color in agricultural areas in the C-band composite. This likely stems from higher sensitivity to volume scattering from agricultural crops, which have less of a volume scattering component at L-band.

### 3.3 Review of SAR Characteristics for Forest Mapping

SAR backscatter values are determined by two main groups of characteristics: sensor and target characteristics. The first group includes the frequency/wavelength of the SAR, polarization of the transmitted and received SAR signal, incidence angle of the radar beam interacting with the ground, and look direction of the sensor. The combination of these characteristics needs to be considered when interpreting and analyzing SAR imagery. It is often ill-advised to combine SAR imagery from a set of varying sensor parameters if the backscatter data are not carefully cross-calibrated. For time series analysis in particular, it is advisable to analyze data from the same sensor characteristics, otherwise signal variations can be misinterpreted as true change, though no change has actually occurred. The following sections review with examples the main sensor characteristics to point to these differences.



**Figure 3.4** Sentinel-1 C-band dual polarimetric VV and VH data: (a) VV, (b) VH, (c) VV/VH ratio, and (d) SAR false color composite with  $RGB = VV/VH/ratio$  channel assignment. Image acquired on May 31, 2018.

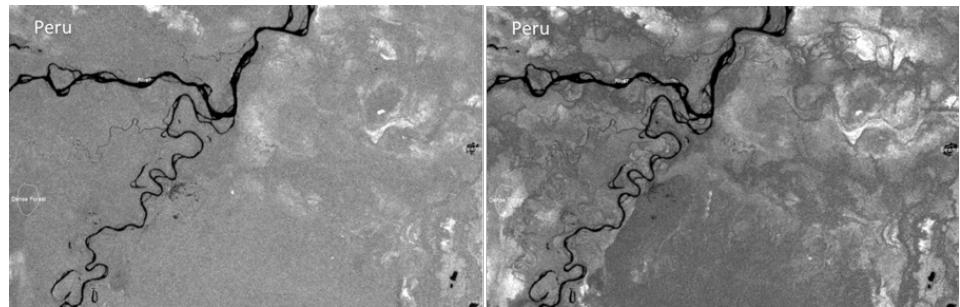


**Figure 3.5** ALOS-1 L-band dual-polarimetric HH and HV data: (a) HH, (b) HV, (c) HH/HV ratio, and (d) SAR false color composite with  $RGB = HH/HV/ratio$  channel assignment. Same area as in **Figure 3.4**, acquired ~10 years earlier on June 22, 2008.

The other group of characteristics determining SAR backscatter of forests and other natural and manmade targets are related to target characteristics. In general, assuming constant imaging sensor characteristics, SAR backscatter is a function of a target's moisture content and structural characteristics. For forests, this means that forest volume (biomass) and structural complexity (forest trunks, branches, and leaves) can indicate species present (e.g., pines vs. deciduous). Unlike optical imagery, if sensor parameters are stable—as is the case with most repeat-pass orbiting SAR sensors—signal variations at any given pixel location are only a function of these target characteristics. Sun angle variations seen in optical data do not affect the active SAR sensing system. Also, atmospheric variations (including clouds) have (almost) no impact on the SAR signal; however, there are notable and important exceptions at shorter wavelength SARs when heavy active rain events are encountered, as seen in C-band observations over tropical environments. Thus, when analyzing radar signals, it is important to recognize that moisture changes in both soil and vegetation strongly determine SAR backscatter. For some key concepts in understanding SAR backscatter from forests and natural vegetation, see Ulaby et al. 1986, 1989, 1990, 2014; Henderson & Lewis 1998; Woodhouse 2006; and Kellndorfer & McDonald 2008.

### 3.3.1 ROLE OF FREQUENCY IN FORESTS

SAR frequency determines the wavelength of the electromagnetic wave interacting with targets such as forests. In a nutshell, the longer the wavelength (i.e., the smaller the frequency), the more a wave penetrates into forest canopies and interacts with larger parts of the forest volume. In a simplistic view, one can attribute X-band (at about 3 cm) to mostly crown and small branch and leaf/needle scattering. C-band (5 cm) penetrates somewhat deeper into crowns and scatters on medium-sized branches. L-band (23 cm) and P-band (40 cm) have strongest penetration capacity and interact with larger parts of trees like big branches and trunks (see **Chapter 2, Fig. 2.6**). As such, L-band and longer wavelengths are often connected with a strong “double-bounce” scattering component, where the incident energy is scattered



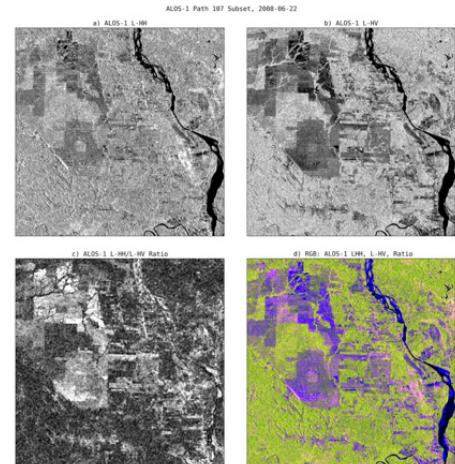
**Figure 3.6** Double-bounce effect from below-canopy flooding at L-HH polarization from ALOS-1: (a) Low-water season and (b) high-water season. Note the brightening of the forests during inundation.

forward towards the ground where it bounces back to the sensor (similar to a racquetball or squash). This double-bounce effect is invaluable for detecting below-canopy flooding effects where inundation with standing water below a tree acts as a strong reflecting surface in the forward direction back to the SAR instrument. In tropical forest environments, riparian forests are thus extremely bright in SAR imagery when flooded (**Fig. 3.6**).

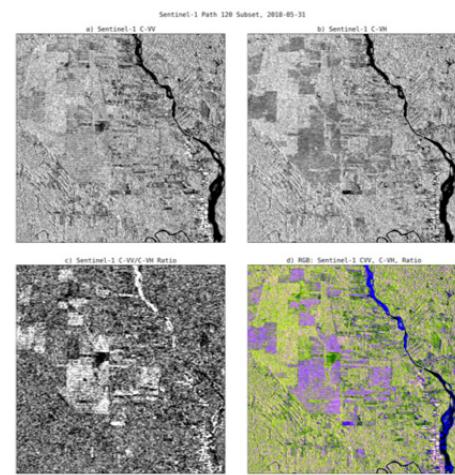
**Figures 3.7 and 3.8** show L- and C-band backscatter images of the oil palm plantation in Ecuador. Although the C-band data are from a timeframe of 10 years after the L-band acquisitions, most notably, the relative absence of very dark surfaces in the C-band data points to strong backscatter from rough surfaces at the shorter wavelengths, whereas at the L-band, surfaces appear smoother (hence, darker) when little or no vegetation is present.

### 3.3.2 ROLE OF POLARIZATION IN FORESTS

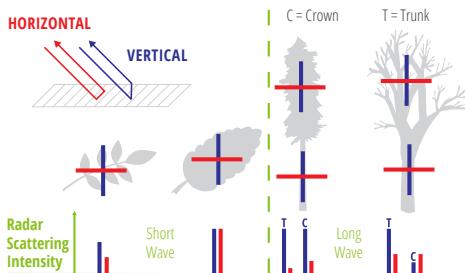
It is important to consider the polarization of radar waves interacting with forests, as it determines how the signal interacts with trunks and crown components. **Figure 3.9** shows a simplified diagram of how long and short wavelengths at horizontal and vertical polarizations interact with forests. Most important is that backscatter from co-polarization (VV, HH) (i.e., same transmit and receive components) is typically stronger for surface scattering components, whereas energy measured from cross-polarized (VH or HV) detection (i.e., measuring energy returning at a 90° offset to the transmitting wave) is associated with measuring volume scattering. **Chapter 2, Section 2.2.3** provides a good background about polarization and sur-



**Figure 3.7** ALOS-1 L-band imagery for the oil palm plantation: (a) L-HH, (b) L-HV, (c) ratio, and (d) RGB composite LHH/LHV/ratio.



**Figure 3.8** Sentinel-1 C-band imagery for the oil palm plantation: (a) C-VV, (b) C-VH, (c) ratio, and (d) RGB composite CVV/CVH/ratio.



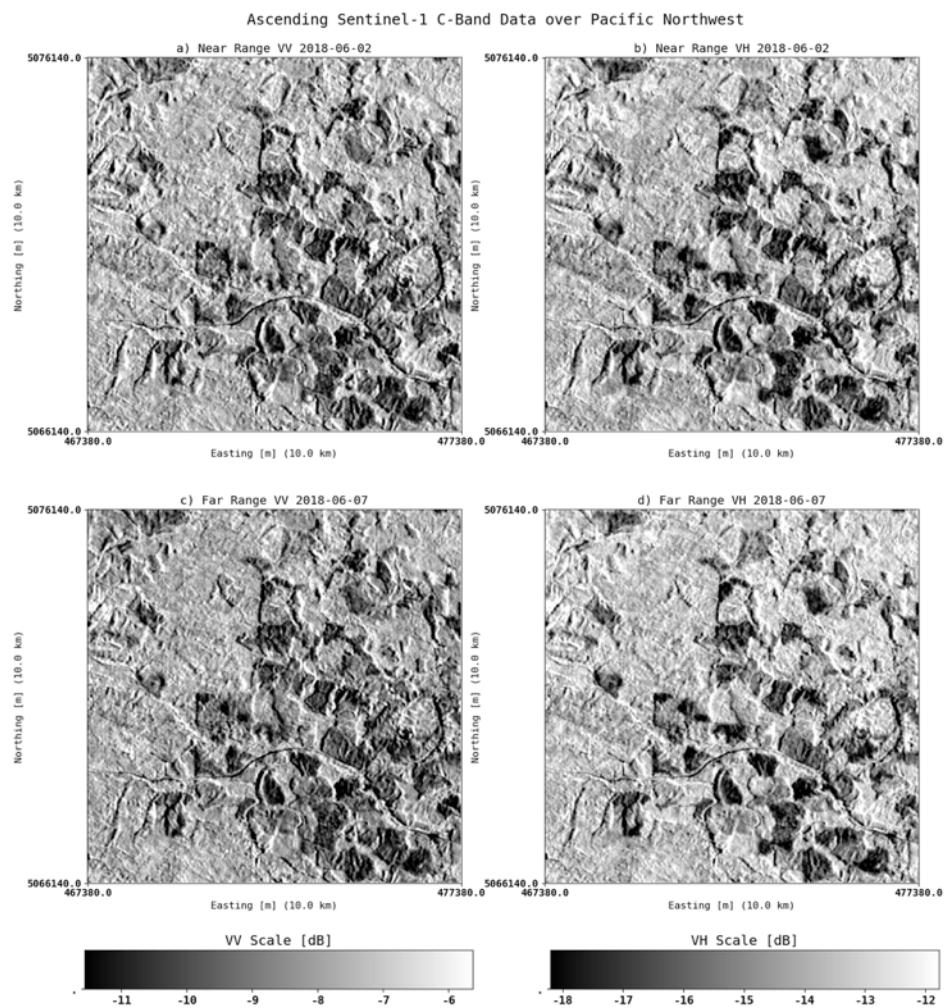
**Figure 3.9** Schematic effects of polarization on backscatter of long and short wavelengths scattering from trunks and crowns.

face scattering types. Thus, for biomass applications, forest degradation tracking, and identifying changes from volumes to surfaces, cross-polarized observations with SAR imagery are essential. The differences between like and cross-polarized imagery from the C- and L-bands of the oil palm plantation are visible in **Figures 3.7 and 3.8**. It can clearly be seen at both L-HH or C-VV that large gray value ambiguities exist between forest canopies and non-forest regions. In the cross-polarized images, these distinctions are more readily made and less ambiguous. Note for example in the L-band image's lower part in **Figure 3.5** that the fishbone logging pattern visible in the HV polarization is not visible in the HH polarization.

### 3.3.3 ROLE OF INCIDENCE ANGLE

The incidence angle describes the angle between the sensor and ground and the surface normal of the illuminated surface (see **Chapter 2**). SAR backscatter is strongly influenced by this angle, as it determines scattering in the crown layer, trunks, and interactions with the ground. If slopes are tilted toward the sensor, stronger backscatter can be expected. If slopes are tilted away from the sensor, weaker backscatter is to be expected. RTC will account for these effects to some degree; however, scattering behavior is strongly dependent on the type of surface cover. This effect is weaker over dense forested environments and stronger over sparse vegetation or bare soils.

**Figure 3.10** is an example from the Pacific Northwest of the United States where timber management involves clearcutting, selective logging, and replanting. The Sentinel-1 images show acquisitions in the subset from overlapping paths, one imaging the area closer to near range (steeper incidence angle) of the SAR sensor and one closer to far range (shallower incidence angle)



**Figure 3.10** Near- and far-range acquisitions of Sentinel-1 CVV and CVH data over a forested site in the Pacific Northwest.

of the sensor. While not immediately obvious, close inspection of the figure shows differences in the near- and far-range acquisitions only five days apart where no significant rain events have changed moisture conditions. The rows show near- and far-range data for VV and VH data in the columns. A comparison of the top and bottom figures in each column illustrates the differences stemming from variations in incidence angles from the overlapping paths.

### 3.3.4 ROLE OF LOOK DIRECTION (ASCENDING/DESCENDING) DATA TAKES

The look direction of a SAR refers to the direction the radar antenna is pointed when emitting and receiving the radar beam. A SAR look direction is determined with respect to the flight direction of the

sensor (see **Chapter 2, Sec. 2.1**). It is analogous to sitting on the right or left side of an airplane and looking out the window. Typically, SAR sensors are configured to look either right or left. If the satellite is rotated, that direction can change. How an area is illuminated by a radar beam changes foremost with image acquisitions during ascending and descending overpasses of an area. **Figure 3.11** exemplifies the effect of look direction from ascending or descending data. The image subset is from the Sentinel-1 cross-over pass in northeast Ecuador at the location shown in the right-hand part of the figure. The left side of the figure shows from top to bottom the combined layover and shadow masks from ascending and descending paths over a Google Earth subset. The center figure shows the descending path, and the bottom

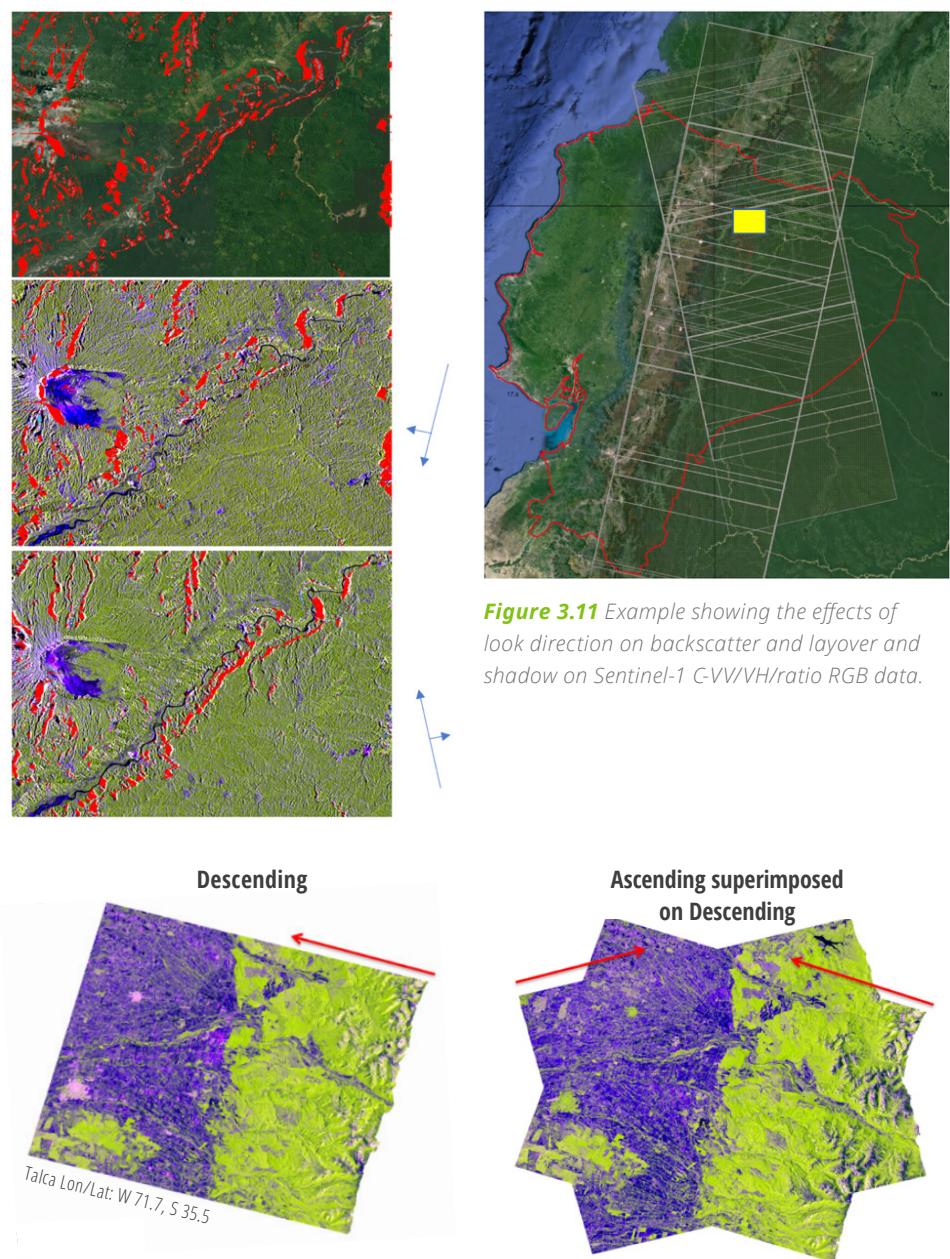
figure shows the ascending path. Differences in the backscatter can be seen as well as the varying locations of the layover and shadow masks (red color). Forest monitoring applications benefit from combining different look directions, as different regions will be mapped and complementary backscatter information can be retrieved.

**Figure 3.12** shows an example of look direction effects for forest observations in Chile from L-band. The city of Talca lies in the western part of the images and can be seen as a rose-colored blob, similar another smaller city farther north. Note that in the ascending data, these two cities turn green in the multi-polarization L-HH/L-HV/ratio image to assume the same backscatter levels as the forests south of Talca and on the Andean slopes in the eastern part of the images. Incidence angle might also contribute with near- and far-range observations, although the gamma naught values mostly flatten the backscatter in the narrow ALOS-1 swath of about a 70-km swath width. Thus, here look direction is mostly causing a change in how the city and forests are seen structurally. Again, if time series analysis for change detection is targeted for forest monitoring, it is advisable to analyze time series by repeat-pass orbits and not mix ascending and descending datasets.

### 3.3.5 ROLE OF MOISTURE

SAR is very sensitive to moisture in soils and vegetation, and also to standing open water and below-canopy standing water. Increased moisture content in soils and vegetation tend to increase the backscatter signals. Standing open water has very dark image characteristics due to most of its energy being scattered in the forward direction away from the sensor; however, when wind, currents, or boat engines rough up water surfaces, strong backscatter can originate from open water surfaces. In particular, shorter wavelengths like C- and X-bands have strong open water surface backscatter from rough water surfaces. At longer wavelengths, the aforementioned double-bounce effect under canopies can have a strong backscatter signal (**Fig. 3.6**).

**Figure 3.13** shows an example of moisture influence on the Sentinel-1 C-band data over Ecuador. The

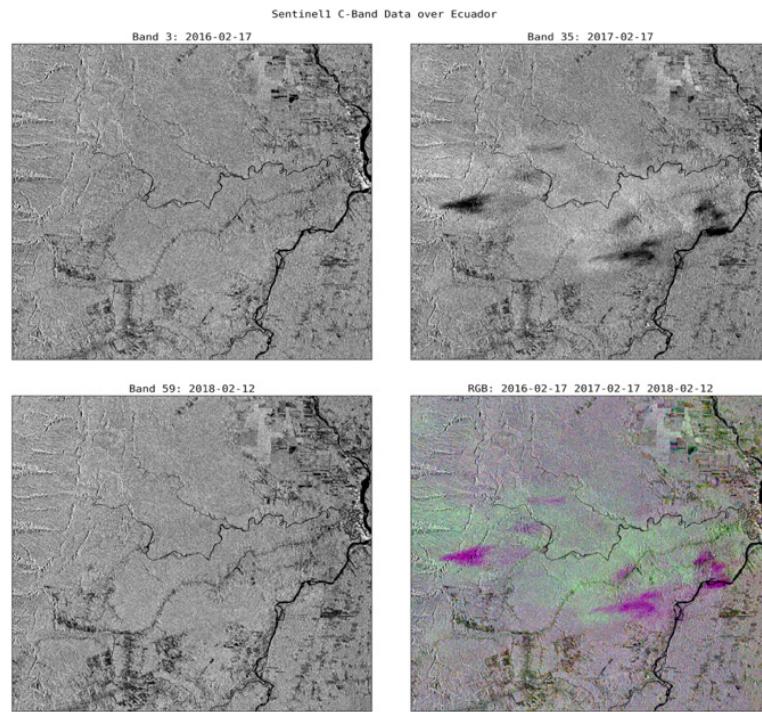


**Figure 3.11** Example showing the effects of look direction on backscatter and layover and shadow on Sentinel-1 C-VV/VH/ratio RGB data.

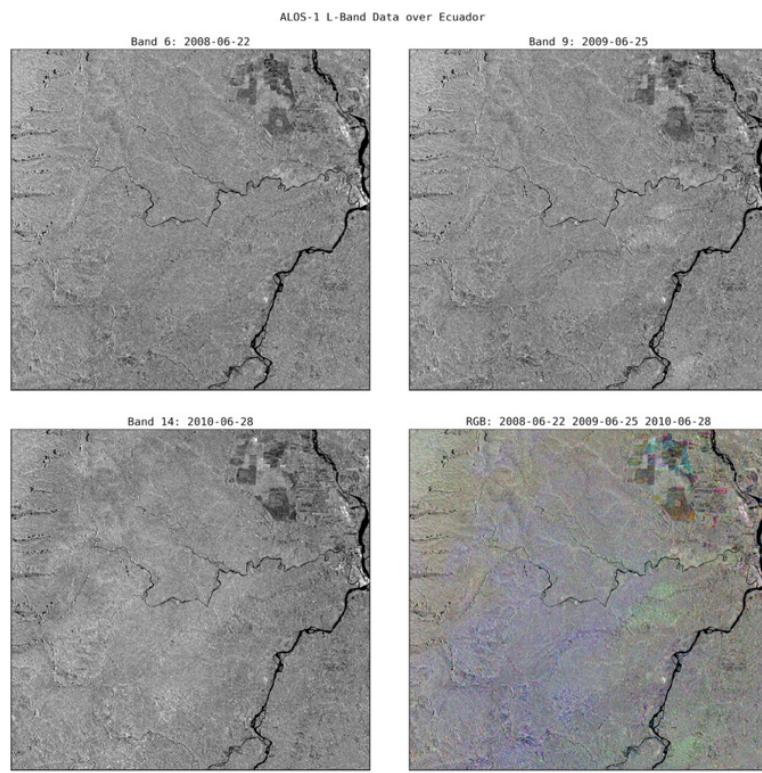
darkening effects are associated with actively raining strong tropical convection systems that cause signal attenuation. The brightening effects stem from wet vegetation and soils from the rain events associated with the tropical frontal system. Riverbeds are still seen in the midst of brightened backscatter areas in the affected image from February 27, 2017, confirm-

ing that the SAR signals indeed stem from an increase in vegetation and soil moisture.

**Figure 3.14** shows the effects of vegetation and soil moisture on signal brightening in L-band HH polarization from ALOS-1 at the Ecuador site. Three acquisitions from the end of June 2008, 2009, and 2010 are compared. While 2008 seems to have few effects



**Figure 3.13** Sentinel-1 CVV example of moisture influence on enhancing and darkening backscatter



**Figure 3.14** ALOS-1 L-HH example of moisture influence on enhancing backscatter.

from moisture-related backscatter enhancements, the year 2009 shows some effects in the eastern part of the image. In 2010, a strong moisture-related brightening is visible. As a result, the multitemporal color composite shows large-scale color variations that are moisture-related. Care must be taken when performing multitemporal image change detection for forest degradation so as to not to interpret darkening in a time series as a degradation signal when moisture variations can be the cause for decreases or increases in backscatter. Time series analysis can help to separate these effects, as moisture variations are shorter in time and space and exhibit a more random pattern compared to real disturbance or deforestation signals.

### 3.3.6 ROLE OF STRUCTURE

In addition to moisture conditions, vegetation structural characteristics determine SAR backscatter from forests. This includes both horizontal structure (i.e., canopy density, row plantations, texture) and vertical structure (i.e., crown depth, crown and trunk biomass, leaf and branching structure, life forms of trees, excurrent or decurrent growth). **Figure 3.15** provides a schematic overview of these structural classes (Dobson et al. 1996).

**Figure 3.16** provides an example of backscatter response for C-VV and C-VH data for the oil palm plantation and its various growth, disturbance, and regrowth stages (including backscatter from undisturbed primary forest). The timing of the Google Earth subset corresponds to the C-band acquisition dates in September 2017.

For L-band sensors, **Figure 3.17** provides an example from a timber management area in Louisiana, U.S. The area is heavily managed, and various stages of clearcutting, selective logging (row thinning), and regrowth can be seen. The cross-polarized data clearly show increased brightness where there are more mature, higher biomass forests.

### 3.3.6 SUMMARY: DEFORESTATION AND FOREST DEGRADATION FROM A SAR POINT OF VIEW

In simple terms, broad characteristics of backscatter behavior can be summarized as follows:

- **Deforestation**—Predominantly a change from volume to surface scattering. This means

Growth Form	Herbaceous			Woody		
	Blade-like	Broadleaf	Shrubs	Trees		
<b>Structural Characteristics:</b>	(i.e. grass, corn)	(i.e. soybeans)	(i.e. alder)	Gymnosperms (i.e. pine)	Angiosperms Dicots (i.e. oak)	Angiosperms Monocots (i.e. palm)
Trunks	None	None	Many small trunks with characteristic orientations	Conical trunk with layered dielectric	Cylindrical, forked trunk with layered dielectric	Cylindrical trunk of homogeneous dielectric
Branches	Non-woody stalks or stems	Non-woody stems	Many small branches & stems	Branch size/orientation varies with height; branches often long/thin	Forked branches, few horizontal elements; branches often short/thick	None
Foliage	Blade-like erectophile	Broad leaves	Blade-like or broad leaves	Needles	Broad leaves	Blade-like clump at top of trunk

**Figure 3.15** Description of simple structural classes of vegetation (Dobson et al. 1996).

cross-polarized (VH, HV) backscatter decreases significantly. However, if deforestation results in rough soil conditions (e.g., slash) or if site preparations rough up soils, backscatter can be significantly enhanced, to the point where actual felling events increase (e.g., until logs are removed). In time series observations, however, trends are towards reduced backscatter. Moisture conditions of soils that are more visible now can enhance signals at C-band significantly and can introduce ambiguities. Time series signals will reveal those transitions.

- **Degradation**—Degradation of forests typically reduces volume scattering and (depending on the amount of degradation) how much soil contributes to the backscatter signal at the observing wavelength. At C-band, degradation is tough to detect unless larger patches of forest are removed. L-band tends to have a detectable signal drop from forest thinning. However, the type of degradation also determines the scattering mechanisms. For example, storm damage may be such that vegetation volumes and scattering mechanisms have enhanced backscatter from slanted trunks, which is difficult to separate from before-disturbance signal strength. Fire events have a strong increase at L-band, where stronger soil contributions enhance double-bounce and hence brighten the backscatter signal. Over time,

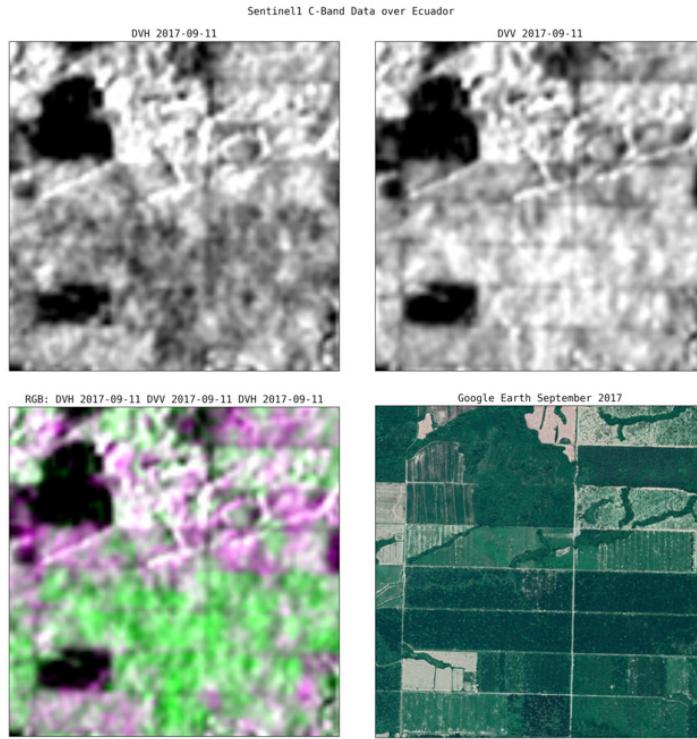
as volume starts to significantly degrade, the SAR signal follows a pattern of backscatter decrease in degraded forests.

**Table 3.3** gives an overview of the expected backscatter characteristics for different vegetation transition scenarios.

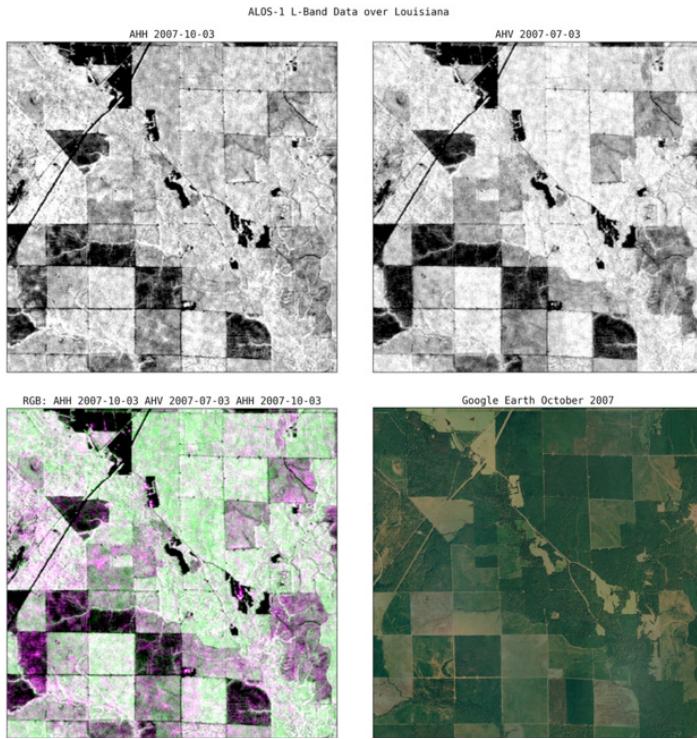
## 3.4 APPROPRIATE SAR PREPROCESSING METHODS FOR FOREST APPLICATIONS

### 3.4.1 WELL-CALIBRATED, RADIOMETRICALLY TERRAIN CORRECTED SAR DATA

Proper RTC of SAR data is a crucial starting point for any analysis of change detection, either bitemporal, in time series, or in combination with optical datasets (see **Chapter 2** for RTC processing discussion). A word of caution: as of this writing, the open source software SNAP delivered by the European Space Agency (ESA) has two known shortcomings: (1) geolocation inaccuracies up to 40 m in the range direction and (2) radiometric correction that is suboptimal given the novel approach by Small et al. (2012). For change detection purposes, careful co-registration after processing with SNAP (i.e., with image matching postprocessing) might overcome some of these issues. However, it is important to assess whether backscatter change stems from geometric



**Figure 3.16** Sentinel-1 C-band example of VV/VH backscatter in the oil palm plantation in Ecuador for different growth stages. Descending orbit (D).



**Figure 3.17** ALOS-1 L-band data over a timber management region in southern Louisiana, U.S., showing various stages of clear cuts, selective logging, and regrowth. Ascending orbit (A).

WAVELENGTH	POLARIZATION	RESPONSE BY FOREST TYPE					
		Sparse Forest (dry)	Sparse Forest (flooded)	Degraded Forest (dry)	Degraded Forest (flooded)	Dense Forest (dry)	Dense Forest (flooded)
C-band backscatter ( $g_0$ )	VV	Medium to high; Depending on the roughness of the forest floor and moisture, there is lots of variation in this category	Low to medium; Depending on forest density, lots of forward scattering	Medium to high; most scattering from crown	Medium to high; most scattering from crown	Medium to high; most scattering from crown (Can be low in scenarios where absorption dominates and diminishes backscatter)	Medium to high; most scattering from crown (Can be low in scenarios where absorption dominates and diminishes backscatter)
	VH	Medium to high; Depending on the roughness of the forest floor and moisture, there is lots of variation in this category	Low to medium; Depending on forest density, lots of forward scattering	Medium to high; most scattering from crown	Medium to high; most scattering from crown	Medium to high; most scattering from crown (Can be low in scenarios where absorption dominates and diminishes backscatter)	Medium to high; most scattering from crown (Can be low in scenarios where absorption dominates and diminishes backscatter)
	VV/VH Ratio	Medium to high	Medium to high	Medium	Medium	Medium	Medium
L-band backscatter ( $g_0$ )	HH	Low to medium; lower than dense forest and flooded sparse forest. At steep incidence angles, backscatter can be medium to high	Medium to high, depending on how much double bounce is contributing to the signal	Medium to high	High to very high, double bounce contributes to high backscatter	High to very high; higher than degraded forest, however at very high biomass levels we see saturation and no distinction with degraded forests	High to very high, double bounce contributes to high backscatter
	HV	Low to very low, depending on how dry the soils are	Low to very low. Most scattering is in the forward direction due to specular reflection	Medium to high	Medium to high, no seasonal variation with flooded forest floor	High to very high; volume scattering is dominant – best sensitivity to biomass	Medium to high, no seasonal variation with flooded forest floor
	HH/HV Ratio	Medium	High	Medium	High	Medium	High

**Table 3.3** Expected backscatter characteristics for different vegetation transition scenarios. Note: Cross-polarized backscatter is generally lower than like polarized backscatter; backscatter values range from very low, low, medium, high, to very high.

offsets rather than real change, particularly in hilly terrain. The quality of the DEM as an input to any orthorectification process is also critical. Note that SRTM-derived DEMs are often adequate for ~20- to 30-m resolution SAR processing; however, improvements in backscatter mapping could be achieved with better resolution DEMs. This is in some ways a question of cost/benefit ratios, as higher resolution DEMs are available, yet often not open source. All datasets shown in this chapter were produced with the Gamma Remote Sensing software, which is also employed by the Alaska SAR facility for RTC production and used by Earth Big Data, LLC, for all SAR geocoding. In preparation for the NISAR mission, the Jet Propulsion Laboratory (JPL) developed the InSAR Scientific Computing Environment (ICSE) software which will eventually be available to the community. A well-suited open source software for post-RTC processing is available in the [Geospatial Data Abstraction Library \(GDAL\)](#) packages from command line or as Python API bindings.

### 3.4.2 MULTITEMPORAL SPECKLE NOISE REDUCTION

If properly stacked SAR data are available (such as in a tiling scheme for manageable data volume handling), it is advisable to preprocess time series data stacks with a multitemporal speckle filter (e.g., by Quegan et al. 2001). Multitemporal speckle filters have been shown to preserve spatial detail while significantly reducing speckle noise at each time step. Multitemporal speckle filters estimate speckle characteristics along the time domain rather than the spatial domain. The resulting speckle statistics can be used to estimate a noise-reduced mean backscatter of a pixel, preserving the backscatter estimate at any time step, but at reduced noise. As such, spatial detail is preserved.

**Figure 3.18** contains an example of L-band data from ALOS. Sixteen multitemporal scenes were available to reduce speckle noise using multitemporal speckle diversity. After filter application, various forest growth and logging states are much

more discernible than before filter application. Given the color theory in [Section 3.2.2](#) and an understanding of volume backscatter changes in L-band HV for forests, the multitemporal image can be readily interpreted as to what areas underwent clearcutting or selective logging (red and yellow colors) and what areas are in regrowth (blue colors) or unchanged stage (white and black colors). Note that perfect alignment of pixels over the temporal domain is a prerequisite of successful multitemporal speckle filtering. Thus, it is advisable to apply the filter on data of the same repeat path.

### 3.4.3. A WORD ON POWER, AMPLITUDE, AND DB SCALES

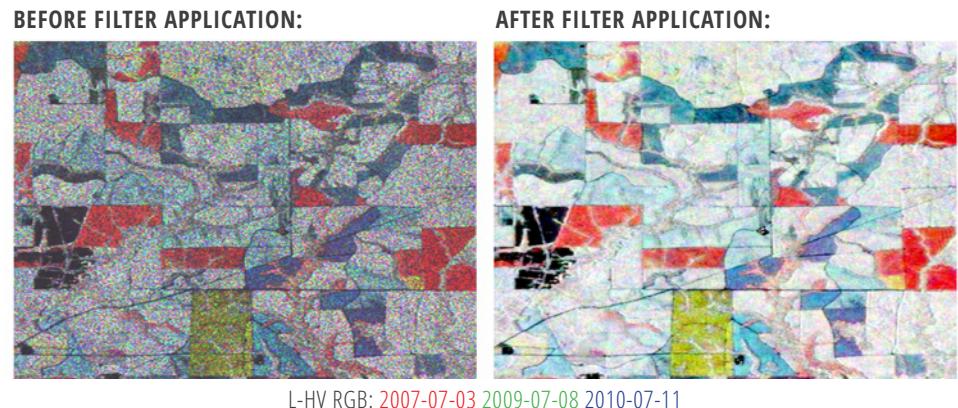
With SAR data handling, it is important perform all spatial and temporal averaging operations in power scale. SAR data expressed in dB (logarithmic transformation) or amplitude scale (square root transformation) introduce mathematical errors when using these averaging or spatial convolution

operations. This is also true for warping operations when convolutions on the SAR data are performed. Therefore, it is recommended that data be converted to the power domain during processing, such as the Earth Big Data's (EBD's) processing software for multitemporal filtering. The QGIS plugin of EBD's open source SAR [time series visualization tool](#) also uses power transformations behind the scenes when displaying time series in dB scale.

#### 3.4.4 TILING AND CONSTRUCTION OF TIME SERIES FROM GEOTIFFS WITH VIRTUAL RASTER TABLES

With the advent of SAR sensors with global acquisitions at high temporal frequency, the era of time series analysis for SAR data has begun. Sentinel-1, with its two-sensor formation flights, now monitors most of the planet at 12-day repeat cycles, denser at higher latitudes. With swath width in high-resolution Interferometric Wide Swath mode at ~250 km, SAR data volumes become massive quite quickly. Thus, it is imperative that appropriate tiling schemes and data handling strategies are employed. For many reasons, the GeoTIFF image format has evolved as a standard for handling remote sensing imagery. In concert with the [Virtual Raster Table \(VRT\)](#) format from the GDAL library, GeoTIFFs can be very efficiently tied together into time series that can readily be subset or rearranged without the need for large raster data operations. VRTs are just XML-based headers that form the metadata for building image band stacks. But even more so, many raster operations can be prescribed as VRT processing in multiple steps, only to be executed on the data when the raster output is generated.

A tiling approach was developed for Sentinel-2 optical data at 20-m resolution based on the Military Grid Reference System (MGRS). This globally consistent Universal Transverse Mercator (UTM) projection-based approach keeps data consistent in spatial extent and projection across the globe. The pixel area of an MGRS UTM tile at the equator is the same as in a tile at higher latitudes. Arguably, this approach keeps data globally minimally distorted, and algorithms for spatial convolutions



**Figure 3.18** Multitemporal speckle filter application on a perfectly co-registered time series data stack of ALOS L-band data over Louisiana, U.S

like speckle filters would work consistently on UTM data. This is not true for data in latitude/longitude spacing, where longitudinal pixel resolution changes with latitude. Using the [Sentinel-2 MGRS tiling scheme](#) also for Sentinel-1 data enables readily optical/SAR fusion without the need for further reprocessing. Hence, the EBD production suite readily provides Sentinel-1 SAR time series data stacks in MGRS tiling format.

A data guide explaining the naming conventions and tiling of VRT/GeoTIFF time series data stacks used by EBD products can be found [here](#). GDAL can be used directly to build VRT stacks solely based in open source components.

## 3.5 Change Detection Approaches for SAR Data

### 3.5.1 BITEMPORAL METHODS

Classic image change detection methods for bitemporal image comparison can be applied to well-calibrated RTC SAR imagery. The log-ratio method was explained in [Chapter 2](#). The Iteratively reweighted Multivariate Alteration Detection (iMAD) algorithm (Nielsen 2007) holds promise for change detection between two images; however, as shown in previous sections, it is important to understand possible impacts on backscatter change that are not linked to real changes such as deforestation. While forest changes are easier to detect in bitemporal

analyses at L-band, C-band data often present a challenge, as surface roughness and moisture components can lead to significant SAR signal ambiguities.

### 3.5.2 TIME SERIES ANALYSIS METHODS

In the past, the availability of SAR data was sparse in space and time; however, the Sentinel-1 mission has been a game changer in moving SAR into operational use. The upcoming NISAR mission—with its open data policy and L-band data at 12-day repeat intervals at medium resolution—will be the next big push for SAR data availability. With near-continuous availability of SAR observations of the ground, real time forest monitoring can thus be achieved. Time series analysis techniques developed for optical imagery are somewhat applicable, although SAR characteristics of backscatter sensitivity to structure and moisture warrant a closer look at new methods. Change point detection with cumulative sums (Manogaran & Lopez 2018) is an established time series analysis technique stemming from the financial sector. With the general SAR backscatter trending to decrease with biomass loss due to deforestation or forest degradation, the application of cumulative sum analysis to SAR time series data seems potentially simple, yet powerful.

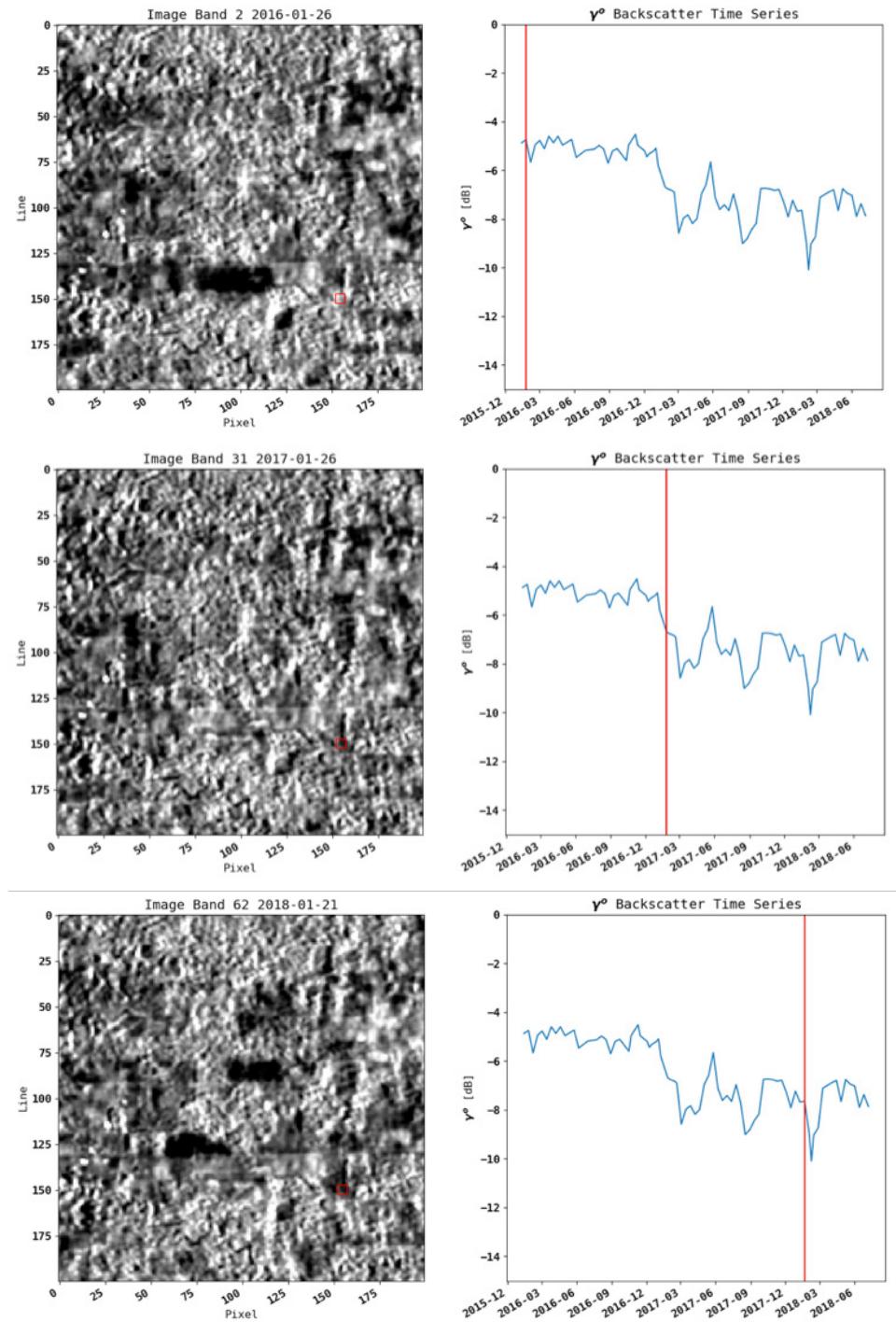
The following figures show time series signals over a deforestation event in Ecuador observed



**Figure 3.19** Ecuador logging test site

with Sentinel-1 data from 2016 to 2018 that exemplify the strength of SAR time series for forest change detection. **Figure 3.19** shows a 4-x-4-km<sup>2</sup> subset of an active logging region in the northeastern part of Ecuador, and **Figure 3.20** shows the time series profile and associated imagery for a logging event in January 2017. While some noise exists in the time series, a clear backscatter decrease in early 2017 is visible in the center image and time series plot. As is typical for deforested areas at C-band, lower backscatter at higher variability is observed in the C-band profile after the deforestation event. This disturbance observation can be identified from the longer trends visible compared to more short-term random noise due to moisture variations. After applying a kernel filter to smooth the time series somewhat, a cumulative sum curve can be constructed from the residuals of the time series data, minus the mean observation of the entire time series.

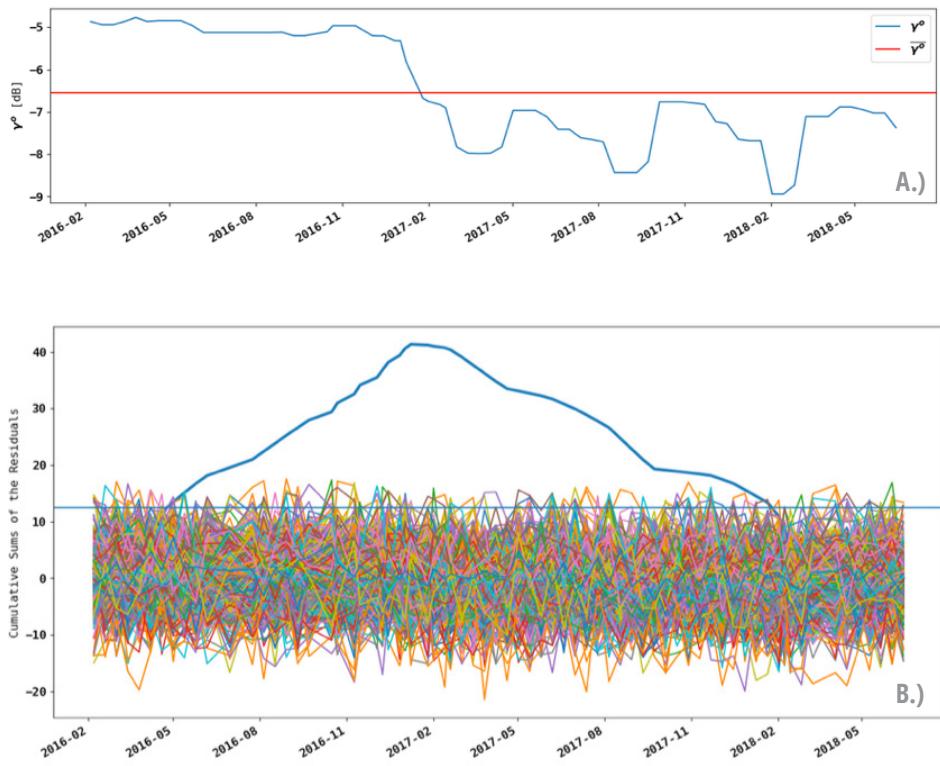
**Figure 3.21(a)** shows the smoothed time series profile and the mean of the time series used to calculate the residuals. The cumulative sum of the residuals is shown as the peaking blue curve in the bottom panel. A way to establish the validity and significance of a candidate change point is to perform a bootstrap analysis in which the time steps are randomly reordered and cumulative sums of the randomized residuals are computed. If the randomization ( $n > 500$ ) shows few or no



**Figure 3.20** Time series profile of red square with associated Sentinel-1 descending VV data.

curves reaching the same maximum value of the peak of the cumulative sum curve (which is the change point in time) the point can be labeled valid. The bootstrapping thus provides a confidence level for a detected change point. Other metrics

can aid in the confirmation of change points in a SAR time series, as elaborated with formulas and Python code in the training Jupyter Notebooks that go along with this chapter. As can be seen in **Figure 3.21(b)**, the 500-fold randomization shows

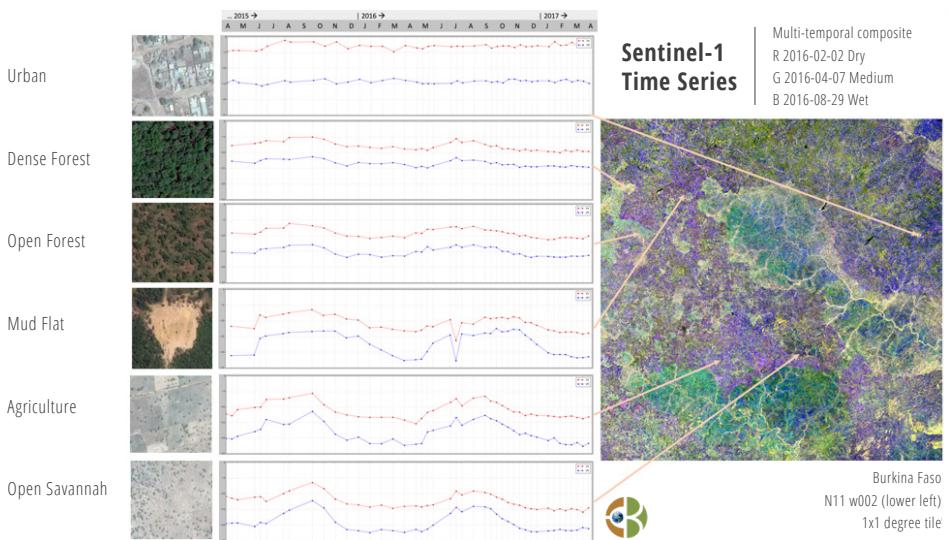


**Figure 3.21** (a) Smoothed time series and mean backscatter, and (b) 500 cumulative sums of the residuals of the time series, minus the mean and 500-fold bootstrapped cumulative sum curves.

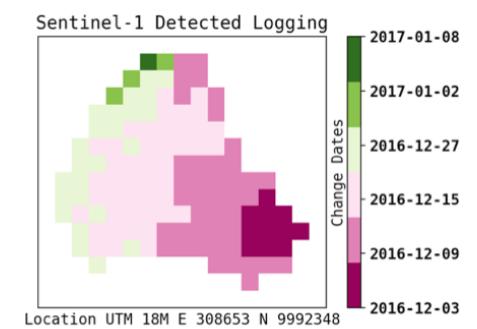
that all randomized S-curves are significantly lower in their peak values compared to the candidate change point in the observed time series.

Applying this approach to all pixels in the subset results in the identification of change pixels and the detected dates of change shown in **Figure 3.22** (right panel). The color codes correspond to the change dates, at a time resolution of about 12 days. The left panel in this figure shows a multitemporal color composite of Sentinel-1 descending VV acquisitions from 2016-11-15 (red), 2017-08-29 (green), and 2018-05-21 (blue). Note that many of the red and yellow color tones in this multitemporal composite correspond to the expected and detected deforestation and forest degradation events. However, some red tones also are more associated with changes in agricultural patterns, which were correctly not mapped as forest degradation events, as their time series profiles did not match the type of curves seen in the previous profiles.

Lastly, to confirm the capability of Sentinel-1 SAR time series to map logging progression, a close-up of the earliest detected event in this region is shown in **Figure 3.23**. Change dates show the progression of the logging of a 5-ha area over the course of four months starting in the southeast corner of the patch and progressing to the west.



**Figure 3.22** Sentinel-1 time series profiles of forest and non-forest land cover patches. Red profiles are C-VV, and blue profiles are C-VH backscatter curves. The backscatter range in each subset shows backscatter from 0 to -20 dB for the SAR  $g^0$  values. The timeframe covers dates from April 2015 to April 2017.



**Figure 3.23** Logging progression detected from Sentinel-1 satellites. A 20-m pixel spacing the subset covers 300 x 320 m<sup>2</sup>. The logged area is 5 ha.

### 3.5.3 SUMMARY ON TIME SERIES SIGNAL ANALYSIS FOR SAR BACKSCATTER DATA

In summary, SAR time series data, such as those now available from Sentinel-1, are an invaluable resource for detailed forest change mapping with quasi-continuous mapping capacity from the sensors. Note that several regions of the planet might be covered more often with ascending or descending data, single-polarization VV or dual-polarization VV+VH datasets. The upcoming NISAR mission will bring the same datasets and temporal frequency at L-band, which will increase forest change detection capability, as fewer signal ambiguities in the time series exist with clear drops in backscatter from deforestation and forest degradation activities.

An example of a semi-arid region and time series signal variation at C-band is provided for Burkina Faso. **Figure 3.22** exemplifies the moisture and structure dependency of various dense forests. Note in this figure how backscatter varies by season due to an increase in moisture and agricultural activity. Even a strong rain event seems to be detected in April 2016, leading to a spike in almost all curves but urban and the mud flat. The mud flat profile shows a strong drop at one date (which is most likely associated with a flash flood event from the heavy rain event), leading to open water surface detection in the time series. Also note that the amplitude in the time series signal increases with decreasing canopy cover, which can be attributed to an increase in soil moisture signal contribution during the rainy season. It can be seen that with decreasing density, the seasonal moisture changes contribute to the rise and fall of backscatter. Thus, it is again important to keep in mind that backscatter signals vary over time, which is vital for careful selection of seasons for time series analysis. A compilation by Ulaby et al. (2014) entitled *Microwave Radar and Radiometric Remote Sensing* contains in-depth resources for SAR data backscatter behavior from soil and vegetation targets.

### 3.5.4 OPTICAL/SAR FUSION FOR FOREST MAPPING

SAR and optical data provide complementary information for forest monitoring, as different im-

aging principles underlie the SAR backscatter and optical multispectral reflectance measurements. As previously noted, SAR measures changes in vegetation and soil moisture content as well as the structural composition of the vegetation (life-forms). Optical remote sensing measures changes in the chemical composition of leaves and their reflectance when illuminated by sunlight, also including measurements of shadow fractions within canopies. Indices like the Normalized Difference Vegetation Index (NDVI) (Tucker 1979) normalize optical reflectance values and provide a measure of the vegetation density or leafiness. Thus, studies of SAR backscatter and NDVI can be used to compare time series of optical and SAR data. Several studies have exploited these similarities, fusing SAR data from Sentinel-1 and ALOS and Landsat time series (Reiche et al. 2016). Various approaches for fusing time series data can be applied. Attempts have been made to fuse time series at the *signal level*, where optical and SAR signals are normalized to simulate similar trends in a fused time series (e.g., filling NDVI gaps with simulated SAR backscatter assuming similar behaviors). This is problematic, however, given that the signals have different underlying principles, although some successes have been demonstrated (Reiche et al. 2015).

Another approach is fusion at the *prediction level*, that is, optical and SAR time series are analyzed separately, and probabilities for deforestation and forest degradation events are computed and compared in the time domain. This has an advantage in that inherent sensor characteristics are optimally analyzed, and probabilities as dimensionless measures can readily be fused in a time series. As such, SAR can fill time gaps in optical observations, and joint probabilities can confirm detections from separate optical or SAR analyses. Holden et al. (forthcoming) developed and tested two approaches for fusing time series of Landsat reflectance and L-band backscatter time series for mapping deforestation for a site with both small- and large-scale agroforestry near Yurimaguas, Peru. This “Probability Fusion” approach—similar to the approaches used by Reiche et al. (2015, 2018)—performed slightly better for finding deforestation with radar

data in terms of map accuracy (78.9% vs. 75.6%) and change detection timing, even with a relative abundance of Landsat data and only 11 radar observations. The improvement when using radar data was much higher when simulating reductions to Landsat data availability. Their “Residual Fusion” algorithm relies on time series regression forecasts (similar to BFAST Monitor (Verbesselt et al. 2010) or CCDC (Zhu et al. 2012)) and was less accurate when fusing data sources than when using Landsat alone, likely because there were too few radar observations to reliably develop forecast regression models. The authors encourage further development of time series fusion algorithms that can incorporate data from current and upcoming radar missions, especially approaches that can go beyond just deforestation mapping to provide class transition labels for IPCC reporting.

## 3.6 Conclusions

With the launch of Sentinel-1 and its associated open data distribution, monitoring forest resources at medium resolution with SAR has now reached operational levels. The C-band mission of the Sentinel-1 sensors are already projected to 2030 in ESA’s budget. NASA and ISRO are poised to launch the L-band NISAR missions at the beginning of the next decade, which will provide 12-day repeat global L-band and regional S-band acquisitions, also with an open data policy. As shown in this chapter, SAR data have a strong sensitivity to forest change. Careful preprocessing is required to build good time series data stacks. Seasonal and moisture variations need to be separated from structural changes in change detection approaches. This requires potentially filtering of the time series to remove “outliers.” Cumulative sum-based change detection of SAR backscatter mean shifts are amongst efficient change detection techniques of the continuously available time series signals.

## 3.7 References

- Craig Dobson, M., Pierce, L. E., & Ulaby, F. T. (1996). Knowledge-Based Land-Cover Classification Using ERS-1/JERS-1 SAR Composites. *IEEE Transactions on Geoscience and Remote Sensing*, 34(1). <http://doi.org/10.1109/36.481896>
- Henderson F., Lewis A. (Eds.) (1998). *Principles and Applications of Imaging Radar. Manual of Remote Sensing, 3rd Edition, Volume 2*. John Wiley and Sons.
- Holden, C., Woodcock, C., Kellndorfer, J. (Forthcoming). Forest change detection by radar/optical time series fusion: a comparison of two approaches using Landsat and ALOS PALSAR.
- Kellndorfer, J. M., & McDonald, K. C. (2008). Active and Passive Microwave Systems. In *The SAGE Handbook of Remote Sensing*. <http://doi.org/10.4135/9780857021052.n13>
- Kellndorfer, J., Cartus, O., Bishop, J., Walker, W., & Holecz, F. (2014). Large Scale Mapping of Forests and Land Cover with Synthetic Aperture Radar Data. In *Land Applications of Radar Remote Sensing*. <http://doi.org/10.5772/58220>
- Manogaran, G., & Lopez, D. (2018). Spatial cumulative sum algorithm with big data analytics for climate change detection. *Computers and Electrical Engineering*, 65, 207–221. <http://doi.org/10.1016/j.compeleceng.2017.04.006>
- Mitchell, A. L., Rosenqvist, A., & Mora, B. (2017). Current remote sensing approaches to monitoring forest degradation in support of countries measurement, reporting and verification (MRV) systems for REDD+. *Carbon Balance and Management*, 12(1), 9. <http://doi.org/10.1186/s13021-017-0078-9>
- Nielsen, A. A. (2007). The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data. *IEEE Transactions on Image Processing*, 16(2), 463–478. <http://doi.org/10.1109/TIP.2006.888195>
- Quegan, S. (2001). Filtering of multichannel SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 39(11), 2373–2379. <http://doi.org/10.1109/36.964973>
- Reiche, J., Verbesselt, J., Hoekman, D., & Herold, M. (2015). Fusing Landsat and SAR time series to detect deforestation in the tropics. *Remote Sensing of Environment*, 156, 276–293. <http://doi.org/https://doi.org/10.1016/j.rse.2014.10.001>
- Reiche, J., de Bruin, S., Hoekman, D., Verbesselt, J., Kellndorfer, J., Rosenqvist, A., Lehmann, E., Woodcock, C., Seifert, F., Herold, M. (2016). Combining satellite data for better tropical forest monitoring. *Nature Climate Change*, 6, 120-122.
- Reiche, J., Hamunyela, E., Verbesselt, J., Hoekman, D., & Herold, M. (2018). Improving near-real time deforestation monitoring in tropical dry forests by combining dense Sentinel-1 time series with Landsat and ALOS-2 PALSAR-2. *Remote Sensing of Environment*, 204, 147–161. <http://doi.org/https://doi.org/10.1016/j.rse.2017.10.034>
- Small, D. (2011). Flattening Gamma: Radiometric Terrain Correction for SAR Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 49(8), 3081–3093. <http://doi.org/10.1109/TGRS.2011.2120616>
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2), 127-150. [http://doi.org/10.1016/0034-4257\(79\)90013-0](http://doi.org/10.1016/0034-4257(79)90013-0)
- Ulaby, F. T., Moore, R. K., & Fung, A. K. (1986). *Microwave remote sensing: active and passive. From theory to applications (Vol. 3)*. Reading, MA: Addison-Wesley.
- Ulaby, F. T., & Dobson, M. C. (1989). *Handbook of radar scattering statistics for terrain. The Artech House remote sensing library*. Norwood, MA: Artech House.
- Ulaby, F. T., & Elachi, C. (1990). *Radar polarimetry for geoscience applications*. Norwood, MA: Artech House.
- Ulaby, F. T., D. G. Long, W. Blackwell, C. Elachi, A. Fung, C. Ruf, K. Sarabandi, J. van Zyl, and H. Zebker (2014). *Microwave Radar and Radiometric Remote Sensing*, 1116 pp., Artech House, Norwood, Mass.
- Verbesselt, J., Hyndman, R., Newham, G., & Culvenor, D. (2010). Detecting trend and seasonal changes in satellite image time series. *Remote Sensing of Environment*, 114(1), 106–15. <http://doi.org/10.1016/j.rse.2009.08.014>
- Woodhouse, I. (2006). *Introduction to Microwave Remote Sensing*. Boca Raton: CRC Press.
- Zhu, Z., Woodcock, C. E., & Olofsson, P. (2012). Continuous monitoring of forest disturbance using all available Landsat imagery. *Remote Sensing of Environment*, 122, 75–91. <http://doi.org/10.1016/j.rse.2011.10.030>

## APPENDIX B

### Chapter 3 Training Module Overview

These training modules are provided as Python scripts in the form of Jupyter Notebooks (<http://jupyter.org>). The appeal of using Jupyter Notebooks is that Jupyter Notebook servers can be deployed from any platform/operating system and natively deploys a range of open source programming and scripting languages that are supported by a very large, growing open source developer community. Jupyter notebooks support the Python and R languages, amongst others. This means that all open source packages supported by Python (e.g., GDAL) are readily available in Jupyter. Also, Jupyter Notebooks are cloud-friendly, as servers can be launched on high-performance cloud instances and displayed via any web browser. Terminals are also supported. The developed Notebooks for the SERVIR training courses have instructions for setting up and using Notebooks. A suite of Notebooks has been developed to cover various aspects of SAR data processing and analysis with a focus on forest mapping. To exercise the Notebooks, several example SAR time series data stacks have been provided for:

- West Africa Region - Sub-Saharan Forest and Savanna ecosystems
- HKH Region - Foothills of the Himalaya

The training datasets are hosted by SERVIR and can be downloaded from [SERVIRglobal.net](#). The Notebooks cover the following topics with embedded exercises (and their solutions):

- Part 1 - Getting to Know SAR Images and Forest Signatures
- Part 2 - SAR Time Series Visualizations and Animations
- Part 3 - Change Detection with Time Series Metrics and Log Ratio Method
- Part 4 - SAR Time Series Change Point Detection
- Part 5 - SAR/Optical (NDVI) Time Series Analysis
- Part 6 - How to Make RGB Composites from Dual-Polarimetric SAR Data

Another Notebook is available that describes how to use the GDAL Virtual Raster Table (VRT) format for efficient stacking of SAR data into an analysis-ready time series data stack. All notebooks, dataset descriptions and installation instructions are also hosted on an open source GitHub repository that can be accessed from [http://github.com/jkellndorfer/servir\\_training](http://github.com/jkellndorfer/servir_training) and <http://github.com/earthbigdata/openSAR>

A time series visualization QGIS plugin tool is also available on the openSAR site.

# SAR Training Workshop for Forest Applications

## Part 1 - Getting to Know SAR Images and Forest Signatures

Josef Kellndorfer, Ph.D., President and Senior Scientist, Earth Big Data, LLC

Revision date: January 2019

This book chapter on SAR data analysis for forest applications with a focus on deforestation and forest degradation monitoring is implemented as an interactive notebook. The digital format (*a jupyter notebook*) of this chapter can readily be launched in any web browser for interactive data exploration with provided or new training data. The notebook is comprised of text written in a combination of executable python code and markdown formatting including latex style mathematical equations. With this approach, the trainees can readily expand, change, and share the entire work with new data sets in new regions or newly available time series steps.

While we are only scratching the surface of available open source tools, the course will provide a broad overview on what modern tools can be employed for SAR focused data analysis, or remote sensing data analysis in general.

### Software Installation and Data Sets

Please refer to the documents **INSTALLATION** and **DATA\_HOWTO**.

The time series data sets for this training course were pre-processed with the EARTH BIG DATA *Software for Earth Big Data Processing, Prediction Modeling and Organization* (SEPP0) using cloud-based processing on Amazon Web Services. SEPP0 allows for the fully automated processing of large SAR (and other remote sensing) data sets to construct time series data effectively. The data format guide **EBD\_README** explains data structures and filenames conventions for data sets produced by EARTH BIG DATA, LLC.

### Notes on Working with this Notebook

1. After launching the notebook server and opening a notebook navigate to the **Kernel** menu and choose ebd:> Kernel > Change Kernel > Python \conda env:ebd\
2. To execute code in a cell, position your blinking cursor inside a cell and either select the *Run* Button from the notebook menu bar, or use the following keystroke combination:
  - CTRL+Enter to run a cell
  - ALT+Enter to run a cell and insert a new cell below
3. To comment lines inside code cells use as first character #. You can mark several lines and use a keystroke combination to comment/uncomment the block with:
  - Windows: CTRL+ /
  - MacOS: CMD+ /

### Importing relevant python packages

First step in the time series analysis approach after obtaining the preprocessed data stacks is the import of necessary python packages.

See the comments below as to what packages are needed and their functions. Note that all these packages should have been installed when the python anaconda environment was created.

```
In [1]: import pandas as pd
import gdal
import numpy as np
import time,os, glob
```

```
In [2]: %matplotlib inline  
import matplotlib.pyplot as plt
```

## Set Project Directory and Filenames

Edit and uncomment the respective cell entries below to activate the wanted project data directory. Take a look at the EBD Data Guide:  
[https://github.com/EarthBigData/openSAR/blob/master/documentation/EBD\\_DataGuide.md](https://github.com/EarthBigData/openSAR/blob/master/documentation/EBD_DataGuide.md) for an explanation of the filenames used for image and date files.

### How to specify data directories:

**Linux:** /path/to/file

**Windows:** d:/path/to/file

D: is the drive letter # IMPORTANT: Always use '/' instead '\' in Windows

NOTE: Directories and filenames are specified in python as strings enclosed in single or double quotes: 'string' "string"

## West Africa - Biomass Site

```
In [5]: datadirectory='/Users/rmuench/Downloads/wa/BIOss1'  
datefile='S32631X398020Y1315440ss1_A_vv_0001_mtfil.dates'  
imagefile='S32631X398020Y1315440ss1_A_vv_0001_mtfil.vrt'  
imagefile_cross='S32631X398020Y1315440ss1_A_vh_0001_mtfil.vrt'
```

## West Africa - Niamey Deforestation Site

```
In [6]: # datadirectory='/dev/shm/projects/c401servir/wa/cra/'  
# datefile='S32631X402380Y1491460ss1_A_vv_0001_A_mtfil.dates'  
# imagefile='S32631X402380Y1491460ss1_A_vv_0001_A_mtfil.vrt'
```

## West Africa - Dam Site

```
In [7]: # datadirectory='/dev/shm/projects/c401servir/wa/DAMss1/'  
# datefile='S32631X232140Y1614300ss1_A_vh_0001_A_mtfil.dates'  
# imagefile='S32631X232140Y1614300ss1_A_vh_0001_A_mtfil.vrt'
```

## HKH Site

```
In [8]: # datadirectory='C:/data/hkh/time_series/S32644X696260Y3052060ss1-EBD'  
# datefile='S32644X696260Y3052060ss1_D_vv_0092_mtfil.dates'  
# imagefile='S32644X696260Y3052060ss1_D_vv_0092_mtfil.vrt'  
# imagefile_cross='S32644X696260Y3052060ss1_D_vh_0092_mtfil.vrt'
```

## Switch to the data directory

```
In [9]: os.chdir(datadirectory)
```

```
In [10]: os.getcwd() # Uncomment this line to display the present working directory
Out[10]: '/Users/rmuench/Downloads/wa/BIOss1'

In [11]: # glob.glob("*.vrt") # Uncomment this line to see a List of the files
```

## Acquisition Dates

**Read from the dates file the dates in the time series and make a pandas date index**

```
In [12]: dates=open(datefile).readlines()
tindex=pd.DatetimeIndex(dates)

In [13]: # From the index we make and print a lookup table for
# band numbers and dates
j=1
print('Bands and dates for',imagefile)
for i in tindex:
    print("{:4d} {}".format(j, i.date()),end=' ')
    j+=1
    if j%5==1: print()

Bands and dates for S32631X398020Y1315440sS1_A_vv_0001_mtfil.vrt
  1 2015-03-22      2 2015-04-03      3 2015-04-15      4 2015-05-09      5 2015-05-21
  6 2015-06-02      7 2015-06-14      8 2015-06-26      9 2015-07-08     10 2015-07-20
 11 2015-08-01     12 2015-08-13     13 2015-08-25     14 2015-09-06     15 2015-09-18
 16 2015-09-30     17 2015-10-12     18 2015-10-24     19 2015-11-17     20 2015-11-29
 21 2015-12-11     22 2015-12-23     23 2016-01-04     24 2016-01-28     25 2016-02-09
 26 2016-03-04     27 2016-03-16     28 2016-03-28     29 2016-04-09     30 2016-04-21
 31 2016-05-03     32 2016-05-15     33 2016-05-27     34 2016-06-08     35 2016-07-02
 36 2016-07-14     37 2016-07-26     38 2016-08-07     39 2016-08-19     40 2016-08-31
 41 2016-09-12     42 2016-09-24     43 2016-10-06     44 2016-10-18     45 2016-10-30
 46 2016-11-11     47 2016-11-23     48 2016-12-05     49 2016-12-17     50 2016-12-29
 51 2017-01-10     52 2017-01-22     53 2017-02-03     54 2017-02-15     55 2017-02-27
 56 2017-03-11     57 2017-03-23     58 2017-04-04     59 2017-04-16     60 2017-04-28
 61 2017-05-10     62 2017-05-22     63 2017-06-03     64 2017-06-15     65 2017-06-27
 66 2017-07-09     67 2017-07-21     68 2017-08-02     69 2017-08-14     70 2017-08-26
 71 2017-09-07     72 2017-09-19     73 2017-10-13     74 2017-10-25     75 2017-11-06
```

## Image data

To **open** an image file and make it readable use the gdal.Open() function. This generates an image handle that can be used for further interactions with the file:

```
In [14]: img=gdal.Open(imagefile)
```

To explore the image, e.g. number of bands, pixels, lines you can use several functions associated with the opened image object, e.g.:

```
In [15]: print(img.RasterCount) # Number of Bands  
print(img.RasterXSize) # Number of Pixels  
print(img.RasterYSize) # Number of Lines
```

```
77  
4243  
3776
```

## Reading data from an image band

To access any band in the image, use the img.GetRasterBand(x) function. E.g. to access the first band x=1, the last band: x=60.

```
In [16]: band=img.GetRasterBand(1)
```

Once a band is selected, several functions associated with the band are available for further processing, e.g.

- band.ReadAsArray(xoff=0,yoff=0,xsize=None,ysize=None)

So, to read the entire raster layer for the band:

```
In [17]: raster=band.ReadAsArray()
```

## Subsets

Because of the potentially large data volume when dealing with time series data stacks, it may be required to read only a subset of data.

With the gdal .ReadAsArray() function, subsets can be requested with offsets and size:

**img.ReadAsArray(xoff=0, yoff=0, xsize=None, ysize=None)**

xoff,yoff are the offsets from the upper left corner in pixel/line coordinates.

xsize,ysize specify the size of the subset in x-direction (left to right) and y-direction (top to bottom).

E.g., to read only a **subset** of 5x5 pixels with an offset of 5 pixels and 20 lines:

```
In [18]: raster_sub=band.ReadAsArray(5,20,5,5)
```

The result is a two dimensional numpy array with the datatype the data were stored in. We can inspect these data in python by simply typing the array name on the command line:

```
In [19]: raster_sub
```

```
Out[19]: array([[4308, 4616, 4944, 4850, 4130],  
                 [3639, 4142, 4789, 5224, 4745],  
                 [3361, 3980, 4785, 5364, 4999],  
                 [3383, 3946, 4674, 5118, 4936],  
                 [3359, 3687, 4155, 4711, 5004]], dtype=uint16)
```

## Displaying Bands in the Time Series of SAR Data

From the look-up table we know that bands 5 and 18 in the Niamey dataset are from late March and late October. Let's take a look at these images.

**HINT: Because python is an object oriented scripting language, we can often combine several steps (or function calls) into one command. See the trick below to access a raster band and read the data in one step.**

```
In [20]: # These will select the two bands
raster_1 = img.GetRasterBand(5).ReadAsArray()
raster_2 = img.GetRasterBand(18).ReadAsArray()
```

## Plotting in Python to Visualize the Image Bands

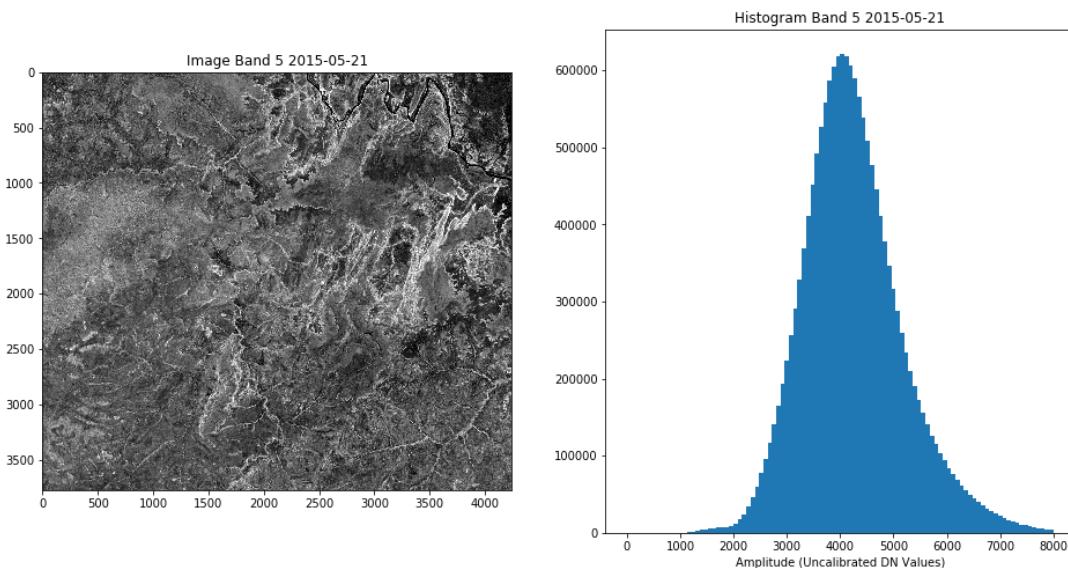
Matplotlib's plotting functions allow for powerful options to display imagery. We are following some standard approaches for setting up figures.

First we are looking at a **raster band** and its associated **histogram**.

```
In [21]: fig = plt.figure(figsize=(16,8)) # Initialize figure with a size
ax1 = fig.add_subplot(121) # 121 determines: 1 row, 2 plots, first plot
ax2 = fig.add_subplot(122) # 122 determines: 1 row, 2 plots, second plot

# First plot: Image
bandnbr=5
ax1.imshow(raster_1,cmap='gray',vmin=2000,vmax=8000)
ax1.set_title('Image Band {} {}'.format(bandnbr,
                                         tindex[bandnbr-1].date()))

# Second plot: Histogram
# IMPORTANT: To get a histogram, we first need to *flatten*
# the two-dimensional image into a one-dimensional vector.
h = ax2.hist(raster_1.flatten(),bins=100,range=(0,8000))
ax2.xaxis.set_label_text('Amplitude (Uncalibrated DN Values)')
ax2.set_title('Histogram Band {} {}'.format(bandnbr,
                                             tindex[bandnbr-1].date()))
```



## Writing a plotting function for the task

Below, the plotting commands used above are defined in a function named `showImage`. Several parameters can be passed to the function, some with default values listed at the end:

- raster = a numpy two dimensional array
- tindex = a panda index array for dates
- bandnbr = the band number the corresponds to the raster
- vmin = minimum value to display
- vmax = maximum value to display

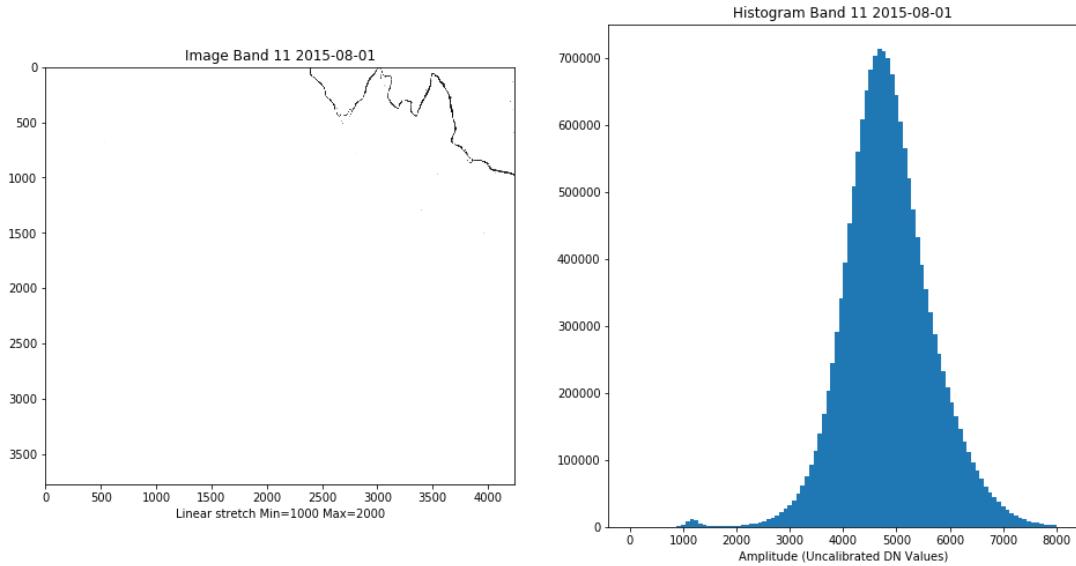
Note: By default, data will be linearly stretched between vmin and vmax.

```
In [22]: def showImage(raster,tindex,bandnbr,vmin=None,vmax=None):  
    fig = plt.figure(figsize=(16,8))  
    ax1 = fig.add_subplot(121)  
    ax2 = fig.add_subplot(122)  
  
    ax1.imshow(raster,cmap='gray',vmin=vmin,vmax=vmax)  
    ax1.set_title('Image Band {} {}'.format(bandnbr,  
                                             tindex[bandnbr-1].date()))  
    vmin=np.percentile(raster,2) if vmin==None else vmin #change vmin & vmax to  
    change what values are displayed  
    vmax=np.percentile(raster,98) if vmax==None else vmax  
    ax1.xaxis.set_label_text(  
        'Linear stretch Min={} Max={}'.format(vmin,vmax))  
  
    h = ax2.hist(raster.flatten(),bins=100,range=(0,8000))  
    ax2.xaxis.set_label_text('Amplitude (Uncalibrated DN Values)')  
    ax2.set_title('Histogram Band {} {}'.format(bandnbr,  
                                              tindex[bandnbr-1].date()))
```

### EXERCISE 1: Read different bands and display them using the function `showImage()`

Use as a variable name for bands `bandnbr`. Use the already open image handle `img` to obtain the raster data from a band.

```
In [23]: # ENTER YOUR CODE HERE  
showImage(raster_2,tindex,11,1000,2000)
```



## EXERCISE 2: Read two different bands and display them side by side

The output should display two bands with a heading of the band numbers. Use the concept for figures with subplots from the function showImage(). Try your code to compare images from different years and different seasons.

```
In [24]: # ENTER YOUR CODE HERE
```

## Time Series Data Stacks

Just as we can use the `ReadAsArray()` function on a band, we can actually use it on the entire image data stack. To read an entire stack, i.e. all bands use the function on the image data handle:

```
img.ReadAsArray()
```

CAUTION: Since this could potentially result in large memory need, it is wise to do some preliminary calculations as to how large of a data set would be read in. For that we can do the following calculation:

$$DataVolume[GB] = \frac{N_{bands} \times N_{pixels} \times N_{lines} \times Bytes\_pixel}{1024^3}$$

For SAR data we typically use datatypes of:

- Float 32 bit (4 bytes per pixel) for power and dB data,
- Unsigned Integer 16 bit (2 bytes per pixel) linearly scaled amplitudes, and
- Unsigned Byte (1 byte per pixel) for dB-scaled to 8 bit data

The following table gives an overview of typically used data types for SAR data analysis in python:

Data Type	Numpy Name	GDAL Name	GDAL Code	Bytes per pixel
Float 32 bit	<code>np.float32</code>	<code>gdal.GDT_Float32</code>	6	4
Unsigned Integer 16 bit	<code>np.uint16</code>	<code>gdal.GDT_UInt16</code>	2	2
Unsigned Integer 8 bit	<code>np.uint8</code>	<code>gdal.GDT_Byte</code>	1	1

Compare the result of the computation with the available RAM on the computer running the notebook.

### EXERCISE 3: Compute the Data Volume of the Raster Stack

Compute the estimated data volume from the data set opened with `gdal.Open()` using the `img` object information `img.RasterXSize`, `img.RasterYSize`, `img.RasterCount`, `img.GetRasterBand(1).DataType`

```
In [25]: # ENTER YOUR CODE HERE (if you need help see the bottom of the document)
```

## Reading the SAR Time Series Subset

Let's read a image subset (offset 2000, 2000 / size 1024, 1024) of the entire time series data stack. The data are representatios of linearly scaled amplitudes scaled to unsigned 16 bit interger

We use the gdal ReadAsArray(xoff,yoff,xsize,ysize) function where:

- xoff = offset in pixels from upper left
- yoff = offset in lines from upper left
- xsize = number of pixels
- ysize = number of lines

If ReadAsArray() is called without any parameters set, the entire image data stack is read. ReadAsArray() returns a numpy array of the form:

```
[bands,lines,pixels]
```

```
In [26]: # Alternatively you can make a subset and use
# it in the ReadAsArray function prefixed with a star
subset=(2000,2000,1024,1024)
rasterDN = img.ReadAsArray(*subset)
```

The numpy .shape object tells us the dimensions of this data stack as *bands (here:time steps), lines, and pixels*:

```
In [27]: rasterDN.shape
Out[27]: (77, 1024, 1024)
```

## Data conversion from linear scaled amplitudes to dB, power and amplitude data

The values of the raw image data show the linearly scaled amplitude values. These digital number (DN) values need to be converted to proper backscatter values of  $\gamma^o$ .

We consider conversion to dB scale (logarithmic scale) for the expression of the SAR backscatter, power, or amplitude scale.

SAR backscatter data of radiometrically terrain corrected data are often expressed as  $\sigma^o$  or the terrain flattened  $\gamma^o$  backscattering coefficients. For forest and land cover monitoring applications  $\gamma^o$  is the preferred metric.

Conversion from power to the logarithmic decibel (dB) scale follows:

$$\gamma_{dB}^o = 10 \times \log_{10}(\gamma_{power}^o)$$

As per widely used convention SAR backscatter data are often stored in 16bit unsigned integer values as linearly scaled amplitude data (referred to below as digital numbers **DN**), conversion to dB scale from the linear scaled amplitudes is performed with a standard **calibration factor of -83 dB**. This is how ALOS SAR data are distributed by JAXA, how Earth Big Data LLC produces all SAR data including Sentinel-1 data, and how NISAR data will likely be scaled:

Conversion from amplitude to dB:

$$\gamma_{dB}^o = 20 * \log_{10}(DN) - 83$$

```
In [28]: rasterdB=20*np.log10(rasterDN)-83
```

Conversion from dB to power:

$$\gamma_{pwr}^o = 10^{\frac{\gamma_{dB}^o}{10}}$$

```
In [29]: rasterPwr=np.power(10.,rasterdB/10.)
```

Conversion from power to amplitude:

$$\gamma_{amp}^o = \sqrt{\gamma_{pwr}^o}$$

```
In [30]: rasterAmp=np.sqrt(rasterPwr)
```

## Explore the image bands of the time steps

Let's explore how a band looks in the various image scales

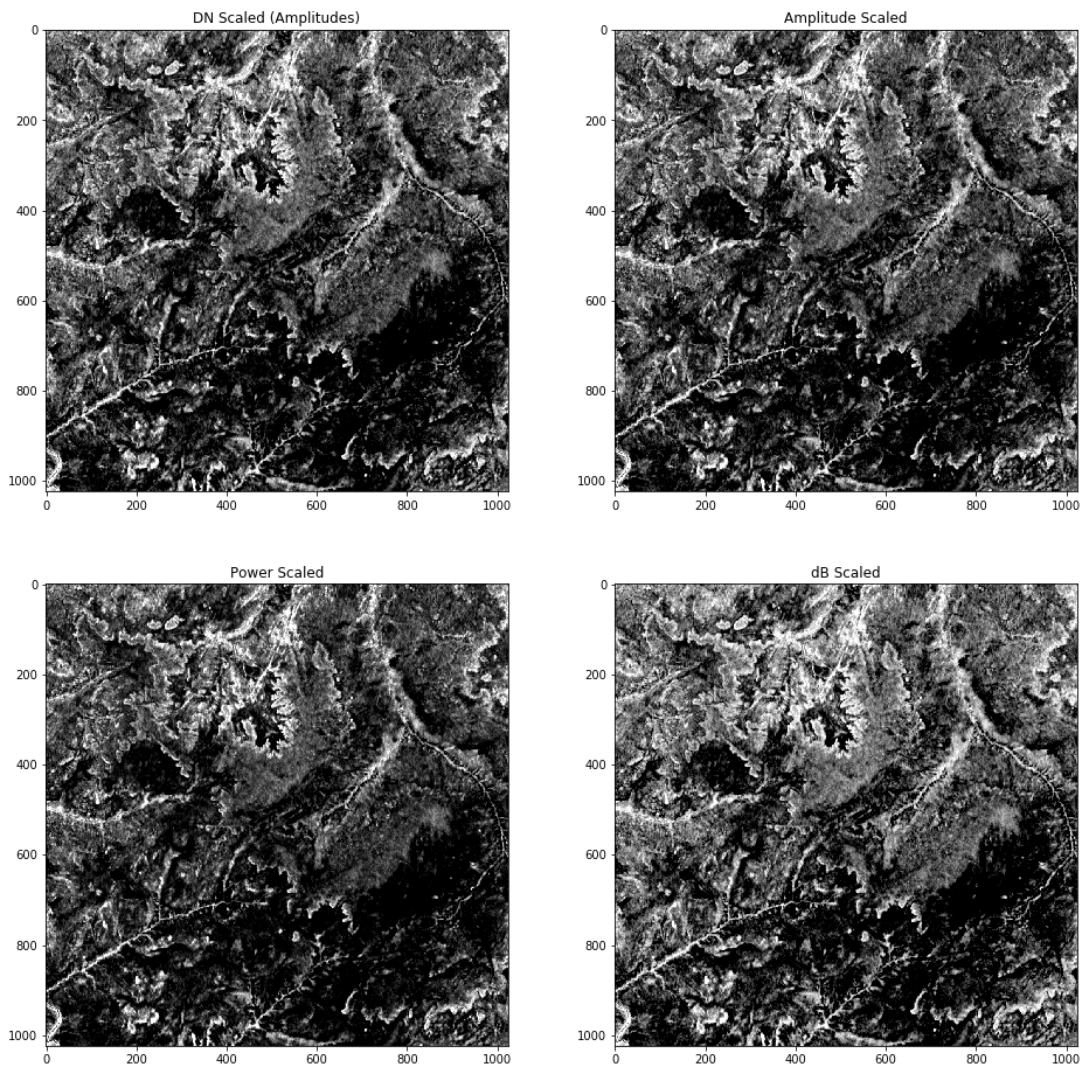
### Choose the band number and find which date it is

```
In [31]: bandnbr=20  
tindex[bandnbr-1]
```

```
Out[31]: Timestamp('2015-11-29 00:00:00')
```

Below is the python code to create a four-part figure comparing the effect of the representation of the backscatter values in the DN, amplitude, power and dB scale.

```
In [32]: fig=plt.figure(figsize=(16,16))  
  
ax1=fig.add_subplot(221)  
ax2=fig.add_subplot(222)  
ax3=fig.add_subplot(223)  
ax4=fig.add_subplot(224)  
  
ax1.imshow(rasterDN[bandnbr],cmap='gray',  
           vmin=np.percentile(rasterDN,10),  
           vmax=np.percentile(rasterDN,90))  
ax2.imshow(rasterAmp[bandnbr],cmap='gray',  
           vmin=np.percentile(rasterAmp,10),  
           vmax=np.percentile(rasterAmp,90))  
ax3.imshow(rasterPwr[bandnbr],cmap='gray',  
           vmin=np.percentile(rasterPwr,10),  
           vmax=np.percentile(rasterPwr,90))  
ax4.imshow(rasterdB[bandnbr],cmap='gray',  
           vmin=np.percentile(rasterdB,10),  
           vmax=np.percentile(rasterdB,90))  
  
ax1.set_title('DN Scaled (Amplitudes)')  
ax2.set_title('Amplitude Scaled')  
ax3.set_title('Power Scaled')  
_ = ax4.set_title('dB Scaled')
```



## Comparing histograms of the amplitude, power, and dB scaled data

```
In [33]: # Setup for three part figure
fig=plt.figure(figsize=(16,4))
fig.suptitle('Comparison of Histograms of SAR Backscatter in Different Scales',font-size=14)
ax1=fig.add_subplot(131)
ax2=fig.add_subplot(132)
ax3=fig.add_subplot(133)

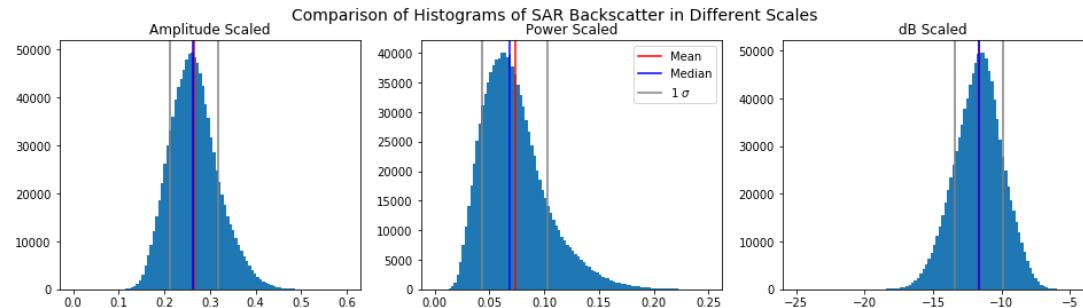
# Important to "flatten" the 2D raster image to produce a histogram
ax1.hist(rasterAmp[bandnbr].flatten(),bins=100,range=(0.,0.6))
ax2.hist(rasterPwr[bandnbr].flatten(),bins=100,range=(0.,0.25))
ax3.hist(rasterdB[bandnbr].flatten(),bins=100,range=(-25,-5))

# Means, medians and stddev
amp_mean=rasterAmp[bandnbr].mean()
amp_std=rasterAmp[bandnbr].std()
pwr_mean=rasterPwr[bandnbr].mean()
pwr_std=rasterPwr[bandnbr].std()
dB_mean=rasterdB[bandnbr].mean()
dB_std=rasterdB[bandnbr].std()

# Some lines for mean and median
ax1.axvline(amp_mean,color='red')
ax1.axvline(np.median(rasterAmp[bandnbr]),color='blue')
ax2.axvline(pwr_mean,color='red',label='Mean')
ax2.axvline(np.median(rasterPwr[bandnbr]),color='blue',label='Median')
ax3.axvline(dB_mean,color='red')
ax3.axvline(np.median(rasterdB[bandnbr]),color='blue')

# Lines for 1 std dev
ax1.axvline(amp_mean-amp_std,color='gray')
ax1.axvline(amp_mean+amp_std,color='gray')
ax2.axvline(pwr_mean-pwr_std,color='gray',label='1 $\sigma$')
ax2.axvline(pwr_mean+pwr_std,color='gray')
ax3.axvline(dB_mean-dB_std,color='gray')
ax3.axvline(dB_mean+dB_std,color='gray')

ax1.set_title('Amplitude Scaled')
ax2.set_title('Power Scaled')
ax3.set_title('dB Scaled')
_=ax2.legend()
```



## Why is the scale important?

It is critical to use the correct scaling of SAR data for image processing operations. As we can see from the comparison of the histograms, the amplitude, power, and dB scales have different statistical distributions.

In time series analysis we often compare measurements at any given time step against the mean of the time series and compute its residuals. When we compute the mean of observations, it makes a difference whether we do that in power or dB scale. Since dB scale is a logarithmic scale, we cannot simply average data in that scale. Consider the following backscatter values and their mean:

$$\gamma_1^o = -10 \text{ dB}$$

$$\gamma_2^o = -15 \text{ dB}$$

Let's compute the mean of these values in power and dB scale and compare the result in dB scale:

```
In [34]: g1_dB = -10
g2_dB = -15
g1_pwr = np.power(10., -10/10.)
g2_pwr = np.power(10., -15/10.)

mean_dB = (g1_dB+g2_dB)/2.
mean_pwr = (g1_pwr+g2_pwr)/2.
mean_pwr_inDB = 10.*np.log10(mean_pwr)

print('Mean averaging dB values      : {:.1f}'.format(mean_dB))
print('Mean averagin power values in dB : {:.1f}'.format(mean_pwr_inDB))

Mean averaging dB values      : -12.5
Mean averagin power values in dB : -11.8
```

As one can see, there is a 0.7 dB difference in the average of these two  $\gamma^o$  backscatter values. If we make mean estimates of backscatter values, the **correct scale** in which operations need to be performed **is the power scale**. This is critical, e.g. when speckle filters are applied, spatial operations like block averaging are performed, or time series are analyzed. Very often we implement models that relate backscatter to biophysical variables like biomass, forest height, or use thresholds to determine change. Ensure that the proper scaling is done when working with the SAR data applying these models.

Another example of the effects can be illustrated with our backscatter data from the images we extracted. Consider a 1 hectare window extracted from our data sets with an offset of 500, 500 for band 20. We compute the mean over time and space of all the pixels.

```
In [35]: offset=500
size=5
o1=offset
o2=offset+size

In [36]: mean_dB = rasterdB[:,o1:o2,o1:o2].mean()
mean_dB

Out[36]: -11.302698

In [37]: mean_pwr = rasterPwr[:,o1:o2,o1:o2].mean()
mean_pwr_in_dB = 10.* np.log10(mean_pwr)
mean_pwr_in_dB

Out[37]: -10.75519323348999
```

As one can see, a difference of more than 0.5 dB is found simply by operating in the different scales. Hence: CAUTION!

## Exploring Polarization Differences

We look at the backscatter characteristics in SAR data from like-polarized (same transmit and receive polarization, hh or vv) and cross-polarized (vh or hv polarization). For this, we read a timestep in both polarizations, plot the histograms, and display the images in dB scale. First, we open the images, pick the bands from the same acquisition date, read the raster bands and convert them to dB scale.

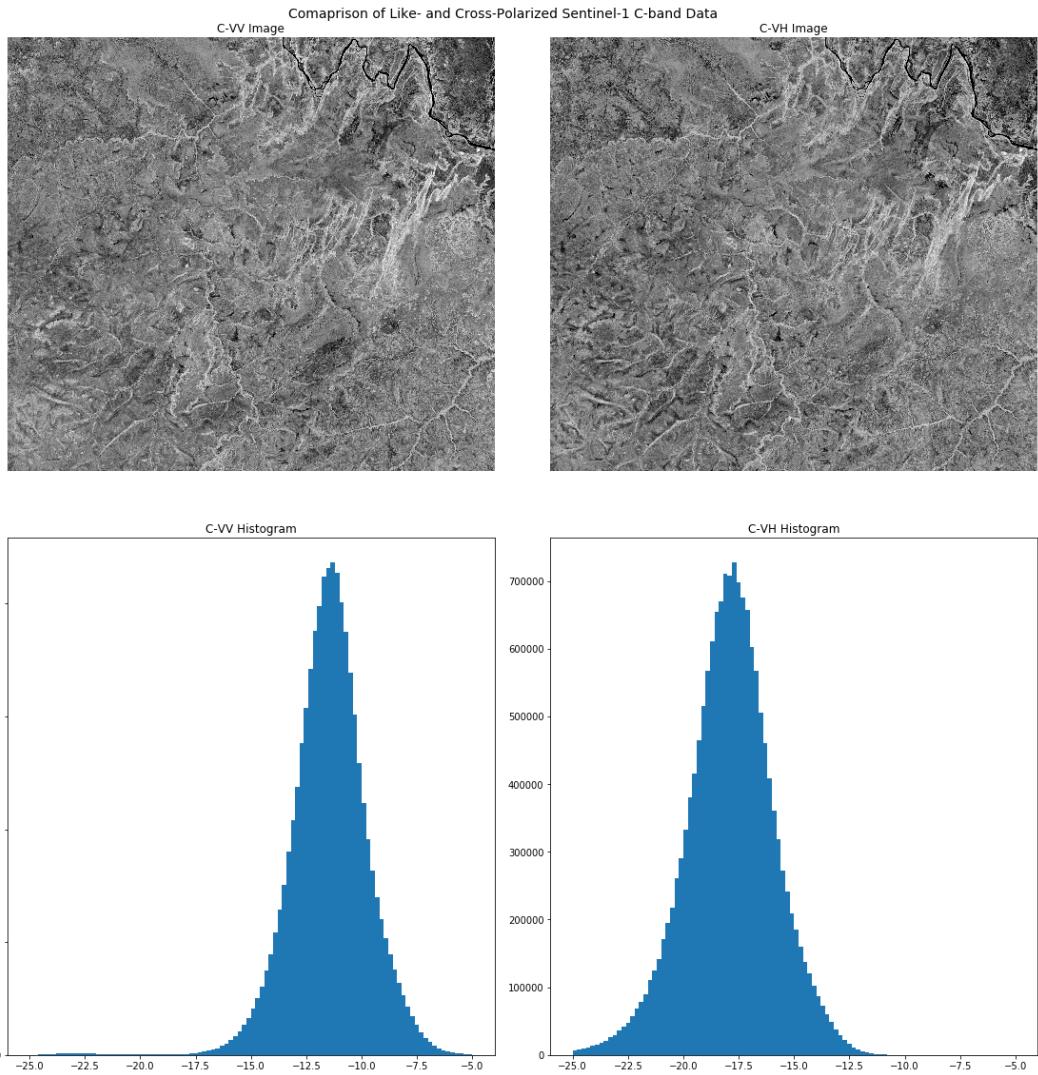
```
In [38]: # Open the Images
img_like=gdal.Open(imagefile)
img_cross=gdal.Open(imagefile_cross)
# Pick the bands, read rasters and convert to dB
bandnbr_like=20
bandnbr_cross=20
rl=img_like.GetRasterBand(bandnbr_like).ReadAsArray()
rc=img_cross.GetRasterBand(bandnbr_cross).ReadAsArray()
rl_dB=20.*np.log10(rl)-83
rc_dB=20.*np.log10(rc)-83
```

Now, we explore the differences in the polarizations by plotting the images with their histograms. We look at the dB ranges over which the histograms spread, and can adjust the linear scaling in the image display accordingly to enhance contrast. In the case below:

- C-w like-polarized data are mostly spread from -17.5 to -5 dB
- C-vh cross-polarized data are mostly spread from -25 to -10 dB

Thus, we note that the cross-polarized data exhibit a larger dynamic range of about 2.5 dB

```
In [39]: fig,ax=plt.subplots(nrows=2,ncols=2,figsize=(16,16))
fig.suptitle('Comaprison of Like- and Cross-Polarized Sentinel-1 C-band Data',
             fontsize=14)
ax[0][0].set_title('C-VV Image')
ax[0][1].set_title('C-VH Image')
ax[1][0].set_title('C-VV Histogram')
ax[1][1].set_title('C-VH Histogram')
ax[0][0].axis('off')
ax[0][1].axis('off')
ax[0][0].imshow(rl_dB,vmin=-17.5,vmax=-5,cmap='gray')
ax[0][1].imshow(rc_dB,vmin=-25,vmax=-10,cmap='gray')
ax[1][0].hist(rl_dB.flatten(),range=(-25,-5),bins=100)
ax[1][1].hist(rc_dB.flatten(),range=(-25,-5),bins=100)
fig.tight_layout() # Use the tight layout to make the figure more compact
```



#### EXERCISE 4: Explore different Seasons in different polarizations

Change the band numbers `bandnbr_like` and `bandnbr_cross` in the cell above to explore different bands.

# EXERCISE SOLUTIONS

## Solution 1

```
In [ ]: # Pick different band numbers for the exercise.  
# Adjust scaling factors and see the effect.  
bandnbr=40  
raster=img.GetRasterBand(bandnbr).ReadAsArray()  
showImage(raster,tindex,bandnbr,4000,8000)
```

## Solution 2

```
In [ ]: # Enter your code for the exercise here.  
bandnbr1=51  
raster1=img.GetRasterBand(bandnbr1).ReadAsArray()  
  
bandnbr2=66  
raster2=img.GetRasterBand(bandnbr2).ReadAsArray()  
  
fig=plt.figure(figsize=(16,8))  
ax1=fig.add_subplot(121)  
ax2=fig.add_subplot(122)  
ax1.imshow(raster1,vmin=2000,vmax=8000,cmap='gray')  
ax2.imshow(raster2,vmin=2000,vmax=8000,cmap='gray')  
ax1.set_title('Band {} Date {}'.format(bandnbr1,tindex[bandnbr1-1].date()))  
_ =ax2.set_title('Band {} Date {}'.format(bandnbr2,tindex[bandnbr2-1].date()))
```

## Solution 3:

```
In [ ]: # Get the Data type  
img.GetRasterBand(1).DataType
```

```
In [ ]: #Use the lookup table for the number of bytes per pixel for this type:  
bytespp=2  
size=img.RasterCount*img.RasterXSize*img.RasterYSize*bytespp/(1024*1024*1024)  
print('Data Volume for {}: {:.1f} Gigabytes'.format(img.GetDescription(),size))
```

# SAR TRAINING WORKSHOP: Forest Applications

## PART 2 - SAR TIME SERIES VISUALIZATION AND ANIMATIONS

Josef Kellndorfer, Ph.D., President and Senior Scientist, Earth Big Data, LLC

Revision date: January 2019

This section introduces more sophisticated animations for time series visualization which allow us to inspect time series in more depth. Note that html animations are not exported into the pdf file, but will display interactively.

```
In [1]: # Turn on inline presentations
%matplotlib inline

In [2]: # Imports
import os
import time
import gdal
import pandas as pd

import numpy as np
import matplotlib.pyplot as plt
import matplotlib.patches as patches # Needed to draw rectangles
from matplotlib import animation, rc
from IPython.display import HTML
```

### West Africa - Biomass Site

```
In [3]: #datadirectory='/dev/shm/projects/c401servir/wa/BIOsS1/'
datadirectory='C:\\\\Users\\\\loaner.SERVIRLOAN-5057.001\\\\Downloads\\\\BIOsS1\\\\'
#if using a PC you may need to add two forward slashes between folders

datefile ='S32631X398020Y1315440sS1_A_vv_0001_mtfil.dates'
imagefile='S32631X398020Y1315440sS1_A_vv_0001_mtfil.vrt'
subset=None
# subset=(2000,2000,1000,1000)
# # Browse image
# # datefile ='S32631X398020Y1315440sS1_A_vh_browse.dates'
# # imagefile='S32631X398020Y1315440sS1_A_vh_browse.tif'
# subset=None
# subset=(3700,1500,500,500)
# subset=(3000,700,500,500)
```

### West Africa - Niamey Deforestation Site

```
In [ ]: # datadirectory='/Users/rmuench/Downloads/wa/cra/'
# datefile ='S32631X402380Y1491460sS1_A_vv_0001_A_mtfil.dates'
# imagefile='S32631X402380Y1491460sS1_A_vv_0001_A_mtfil.vrt'
# subset=None
```

### West Africa - Dam Site

```
In [ ]: # datadirectory='/Users/rmuench/Downloads/wa/DAMsS1/'
# datefile='S32631X232140Y1614300sS1_A_vh_0001_A_mtfil.dates'
# imagefile='S32631X232140Y1614300sS1_A_vh_0001_A_mtfil.vrt'
# subset=None
# # subset=(2000,1500,500,500)
# # subset=(1500,500,500,500)
```

## HKH Site

```
In [ ]: # datadirectory=/Users/rmuench/Downloads/hkh/time_series/S32644X696260Y3052060sS1
-EBD'
# datefile='S32644X696260Y3052060sS1_D_vv_0092_mtfil.dates'
# imagefile='S32644X696260Y3052060sS1_D_vv_0092_mtfil.vrt'
# imagefile_cross='S32644X696260Y3052060sS1_D_vh_0092_mtfil.vrt'
```

## Prepare the Animations

```
In [4]: os.chdir(datadirectory)
```

```
In [5]: # Get the date indices via pandas
dates=open(datefile).readlines()
tindex=pd.DatetimeIndex(dates)
```

```
In [6]: tindex
```

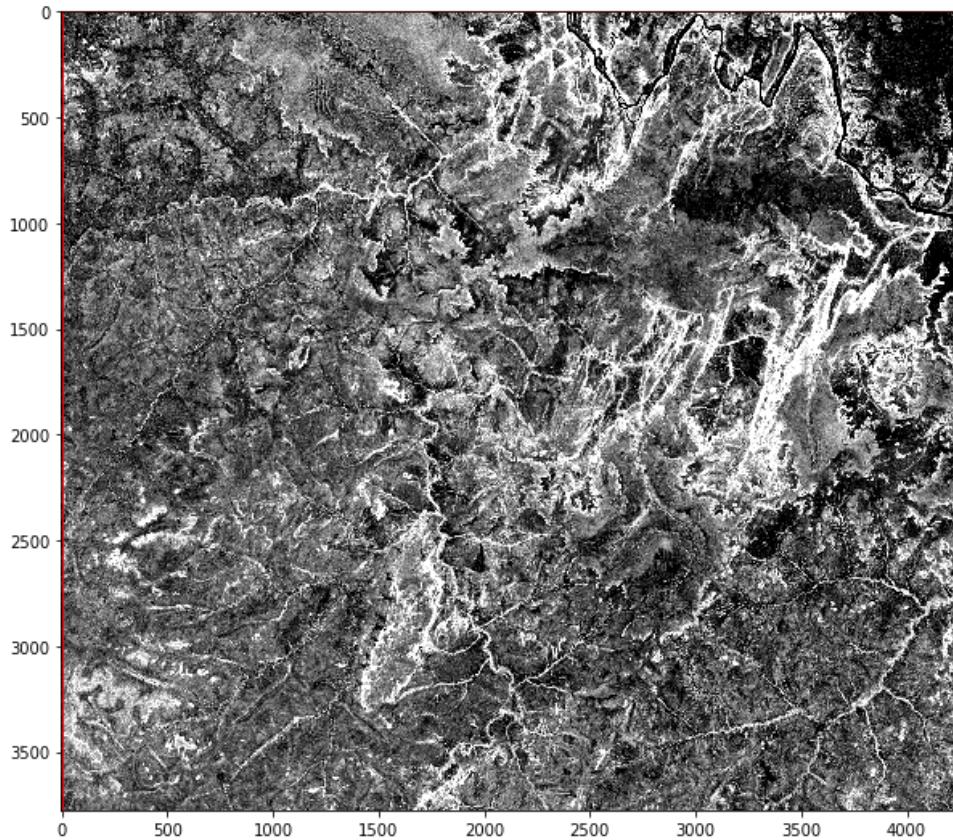
```
Out[6]: DatetimeIndex(['2015-03-22', '2015-04-03', '2015-04-15', '2015-05-09',
'2015-05-21', '2015-06-02', '2015-06-14', '2015-06-26',
'2015-07-08', '2015-07-20', '2015-08-01', '2015-08-13',
'2015-08-25', '2015-09-06', '2015-09-18', '2015-09-30',
'2015-10-12', '2015-10-24', '2015-11-17', '2015-11-29',
'2015-12-11', '2015-12-23', '2016-01-04', '2016-01-28',
'2016-02-09', '2016-03-04', '2016-03-16', '2016-03-28',
'2016-04-09', '2016-04-21', '2016-05-03', '2016-05-15',
'2016-05-27', '2016-06-08', '2016-07-02', '2016-07-14',
'2016-07-26', '2016-08-07', '2016-08-19', '2016-08-31',
'2016-09-12', '2016-09-24', '2016-10-06', '2016-10-18',
'2016-10-30', '2016-11-11', '2016-11-23', '2016-12-05',
'2016-12-17', '2016-12-29', '2017-01-10', '2017-01-22',
'2017-02-03', '2017-02-15', '2017-02-27', '2017-03-11',
'2017-03-23', '2017-04-04', '2017-04-16', '2017-04-28',
'2017-05-10', '2017-05-22', '2017-06-03', '2017-06-15',
'2017-06-27', '2017-07-09', '2017-07-21', '2017-08-02',
'2017-08-14', '2017-08-26', '2017-09-07', '2017-09-19',
'2017-10-13', '2017-10-25', '2017-11-06', '2017-11-18',
'2017-11-30'],
dtype='datetime64[ns]', freq=None)
```

```
In [7]: # Open the image and read the first raster band
img = gdal.Open(imagefile)
band = img.GetRasterBand(1)
# Set the subset
if subset==None:
    subset=(0,0,img.RasterXSize,img.RasterYSize)
```

```
In [8]: subset
```

```
Out[8]: (0, 0, 4243, 3776)
```

```
In [9]: # Plot one band and subset outline to see which subset we are interested in
raster=band.ReadAsArray()
vmin=np.percentile(raster.flatten(),5)
vmax=np.percentile(raster.flatten(),95)
fig=plt.figure(figsize=(10,10))
ax=fig.add_subplot(111)
ax.imshow(raster,cmap='gray',vmin=vmin,vmax=vmax)
# plot the subset as rectangle
_ax=ax.add_patch(patches.Rectangle((subset[0],subset[1]),subset[2],subset[3],
fill=False,edgecolor='red'))
```



```
In [10]: raster0 = band.ReadAsArray(*subset)
bandnbr=0 # Needed for updates
rasterstack=img.ReadAsArray(*subset)
```

```
In [11]: %%capture
import matplotlib.pyplot as plt
import matplotlib.animation
import numpy as np

fig=plt.figure(figsize=(10,10))
ax = fig.add_subplot(111)
ax.axis('off')
vmin=np.percentile(rasterstack.flatten(),5)
vmax=np.percentile(rasterstack.flatten(),95)

im = ax.imshow(raster0,cmap='gray',vmin=vmin,vmax=vmax)
ax.set_title("{}".format(tindex[0].date()))

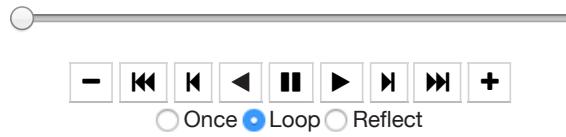
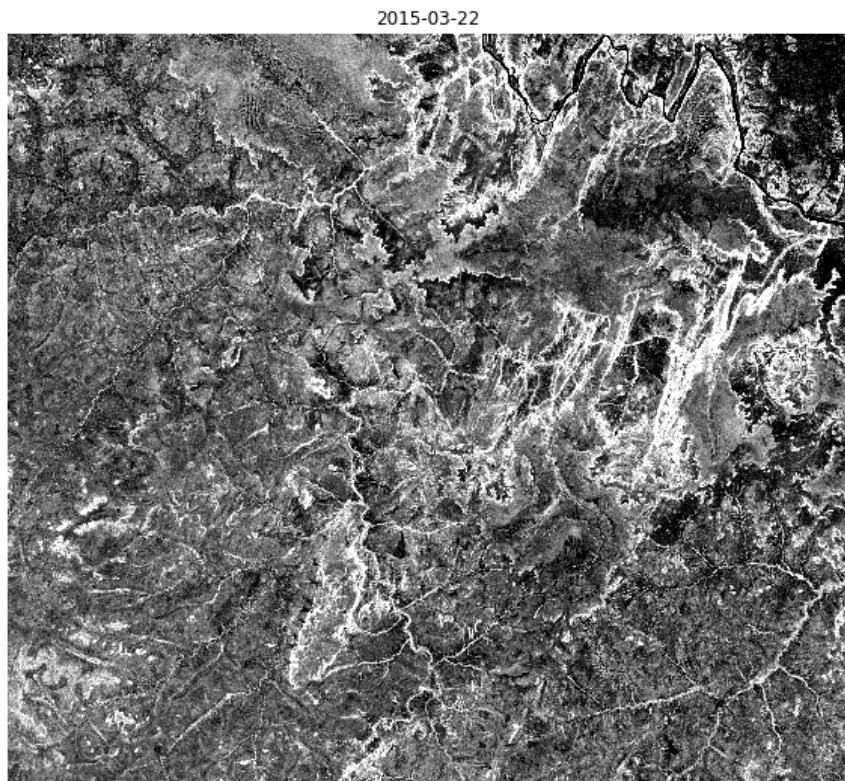
def animate(i):
    ax.set_title("{}".format(tindex[i].date()))
    im.set_data(rasterstack[i])

# Interval is given in milliseconds
ani = matplotlib.animation.FuncAnimation(fig, animate,
                                         frames=rasterstack.shape[0],
                                         interval=400)
```

```
In [12]: rc('animation',embed_limit=40971520.0) # We need to increase the
# limit to show the entire animation
```

```
In [13]: from IPython.display import HTML  
HTML(ani.to_jshtml())
```

Out[13]:



## Plot the global means of the Time Series for the Subset

1. Conversion to power
2. Compute means
3. Convert to dB
4. Plot time series of means

```
In [14]: # 1. Conversion to Power
caldB=-83
calPwr = np.power(10.,caldB/10.)
rasterstack_pwr = np.power(rasterstack,2.)*calPwr
# 2. Compute Means
rs_means_pwr = np.mean(rasterstack_pwr, axis=(1,2))
# 3. Convert to dB
rs_means_dB = 10.*np.log10(rs_means_pwr)
```

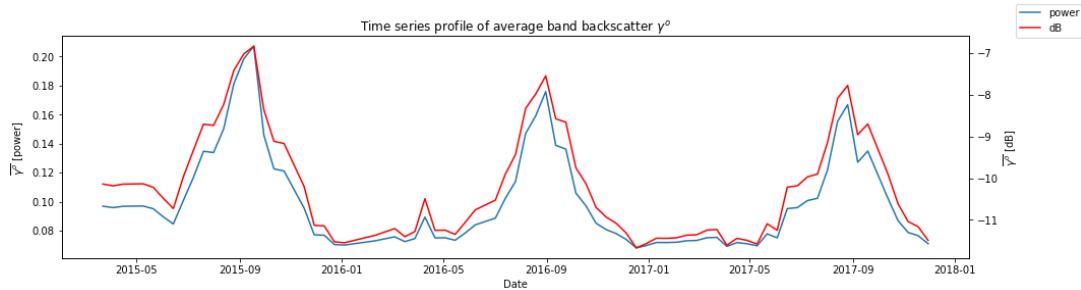
```
In [15]: rs_means_pwr.shape # Check that we got the means over time
```

```
Out[15]: (77,)
```

```
In [16]: # 4. Plot
fig=plt.figure(figsize=(16,4))
ax1=fig.add_subplot(111)
ax1.plot(tindex,rs_means_pwr)
ax1.set_xlabel('Date')
ax1.set_ylabel('$\overline{\gamma^o}$ [power]')

ax2=ax1.twinx()
ax2.plot(tindex,rs_means_dB,color='red')
ax2.set_ylabel('$\overline{\gamma^o}$ [dB]')
fig.legend(['power','dB'],loc=1)
plt.title('Time series profile of average band backscatter $\overline{\gamma^o}$ ')
```

```
Out[16]: Text(0.5, 1.0, 'Time series profile of average band backscatter $\overline{\gamma^o}$ ')
```



```
In [17]: a =pd.Series(rs_means_dB,index=tindex)
```

```
In [18]: #This will print a list of the dates and the respective dB mean values  
a
```

```
Out[18]: 2015-03-22    -10.139979    2016-12-05    -11.314669  
2015-04-03    -10.184575    2016-12-17    -11.671808  
2015-04-15    -10.143337    2016-12-29    -11.566748  
2015-05-09    -10.134364    2017-01-10    -11.438762  
2015-05-21    -10.218999    2017-01-22    -11.441662  
2015-06-02    -10.481450    2017-02-03    -11.427748  
2015-06-14    -10.728489    2017-02-15    -11.367423  
2015-06-26    -9.964857    2017-02-27    -11.354894  
2015-07-08    -9.330598    2017-03-11    -11.246345  
2015-07-20    -8.706461    2017-03-23    -11.229724  
2015-08-01    -8.734129    2017-04-04    -11.605949  
2015-08-13    -8.235253    2017-04-16    -11.442106  
2015-08-25    -7.423883    2017-04-28    -11.490215  
2015-09-06    -7.023914    2017-05-10    -11.580404  
2015-09-18    -6.836782    2017-05-22    -11.092516  
2015-09-30    -8.363434    2017-06-03    -11.248882  
2015-10-12    -9.116455    2017-06-15    -10.215062  
2015-10-24    -9.169089    2017-06-27    -10.183667  
2015-11-17    -10.202442    2017-07-09    -9.969588  
2015-11-29    -11.128629    2017-07-21    -9.901120  
2015-12-11    -11.145837    2017-08-02    -9.140317  
2015-12-23    -11.523537    2017-08-14    -8.085749  
2016-01-04    -11.549477    2017-08-26    -7.777649  
2016-01-28    -11.430446    2017-09-07    -8.956864  
2016-02-09    -11.372730    2017-09-19    -8.700844  
2016-03-04    -11.207707    2017-10-13    -9.901974  
2016-03-16    -11.401895    2017-10-25    -10.611529  
2016-03-28    -11.278982    2017-11-06    -11.038310  
2016-04-09    -10.491783    2017-11-18    -11.161617  
2016-04-21    -11.249865    2017-11-30    -11.492415  
...  
Length: 77, dtype: float64
```

## A two part figure with moving global mean backscatter of the time series in dB

```
In [19]: %%capture
import matplotlib.pyplot as plt
import matplotlib.animation
import numpy as np

fig, (ax1,ax2) = plt.subplots(1,2,figsize=(16,4),gridspec_kw = {'width_ratios':[1, 3]})

vmin=np.percentile(rasterstack.flatten(),5)
vmax=np.percentile(rasterstack.flatten(),95)
im = ax1.imshow(raster0,cmap='gray',vmin=vmin,vmax=vmax)
ax1.set_title("{}".format(tindex[0].date()))
ax1.set_axis_off()

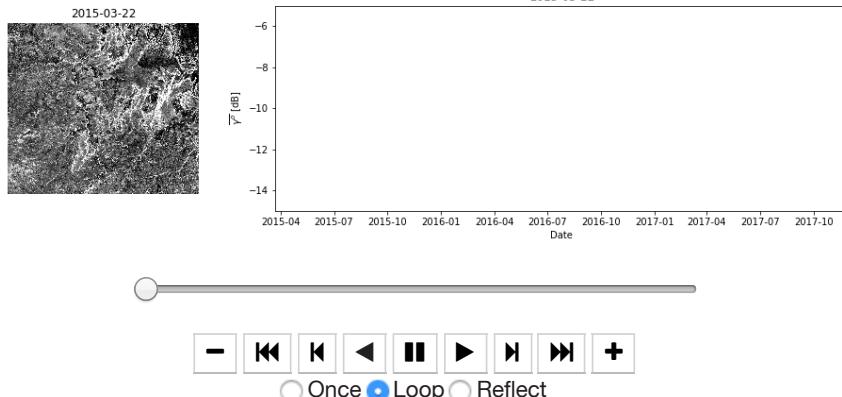
ax2.axis([tindex[0],tindex[-1],rs_means_dB.min(),rs_means_dB.max()])
ax2.set_ylabel('$\overline{\gamma^o}$ [dB]')
ax2.set_xlabel('Date')
ax2.set_ylim((-15,-5))
l, = ax2.plot([],[])

def animate(i):
    ax1.set_title("{}".format(tindex[i].date()))
    im.set_data(rasterstack[i])
    ax2.set_title("{}".format(tindex[i].date()))
    l.set_data(tindex[:i+1],rs_means_dB[:i+1])

# Interval is given in milliseconds
ani = matplotlib.animation.FuncAnimation(fig, animate,
                                         frames=rasterstack.shape[0],
                                         interval=400)
```

```
In [20]: from IPython.display import HTML
HTML(ani.to_jshtml())
```

Out[20]:



## EXERCISE

Modify the animation function to display animation of a single pixel of your choosing.

Bonus: Add a second pixel to the right hand graph.

# SAR Training Workshop for Forest Applications

## PART 3 - Change Detection with Time Series Metrics and Log Ratio Method

Josef Kellndorfer, Ph.D., President and Senior Scientist, Earth Big Data, LLC

Revision date: January 2018

In this chapter we introduce three methods for change detection based on

- Time series metrics 95<sup>th</sup> and 5<sup>th</sup> percentile difference thresholding
- Time series coefficient of variation thresholding
- Log Ratio from two image pairs

### Import Python modules

```
In [1]: import os,sys,gdal  
%matplotlib inline  
import matplotlib.pylab as plt  
import matplotlib.patches as patches # Needed to draw rectangles  
from skimage import exposure # to enhance image display  
import numpy as np  
import pandas as pd
```

### Select the project data set and time series data

#### Louisiana Timber Management Site

```
In [ ]: # SENTINEL-1 TIME SERIES STACK VV from LOUISIANA FOREST MANAGEMENT SITE  
#datapath='/dev/shm/projects/c303nisar/louisiana/15SWRsS1/15SWRsS1-EBD/'  
#imagefile='15SWRsS1_A_vv_0063_A_mtfil.vrt'  
#datefile='15SWRsS1_A_vv_0063_A_mtfil.dates'
```

#### West Africa - Biomass Site

```
In [4]: datapath='/Users/rmuench/Downloads/wa/BIOss1'  
datefile='S32631X398020Y1315440sS1_A_vv_0001_mtfil.dates'  
imagefile='S32631X398020Y1315440sS1_A_vv_0001_mtfil.vrt'  
imagefile_cross='S32631X398020Y1315440sS1_A_vh_0001_mtfil.vrt'  
  
In [5]: os.chdir(datapath)
```

We are defining two helper functions for this task

- **CreateGeoTiff()** to write out images
- **timeseries\_metrics()** to compute various metrics from a time series data stack

```
In [6]: def CreateGeoTiff(Name, Array, DataType, NDV, bandnames=None, ref_image=None,
                       GeoT=None, Projection=None):
    # If it's a 2D image we fake a third dimension:
    if len(Array.shape)==2:
        Array=np.array([Array])
    if ref_image==None and (GeoT==None or Projection==None):
        raise RuntimeWarning('ref_image or settings required.')
    if bandnames != None:
        if len(bandnames) != Array.shape[0]:
            raise RuntimeError('Need {} bandnames. {} given'
                               .format(Array.shape[0],len(bandnames)))
    else:
        bandnames=['Band {}'.format(i+1) for i in range(Array.shape[0])]
    if ref_image!= None:
        refimg=gdal.Open(ref_image)
        GeoT=refimg.GetGeoTransform()
        Projection=refimg.GetProjection()
    driver= gdal.GetDriverByName('GTIFF')
    Array[np.isnan(Array)] = NDV
    DataSet = driver.Create(Name,
                            Array.shape[2], Array.shape[1], Array.shape[0], DataType)
    DataSet.SetGeoTransform(GeoT)
    DataSet.SetProjection(Projection)
    for i, image in enumerate(Array, 1):
        DataSet.GetRasterBand(i).WriteArray( image )
        DataSet.GetRasterBand(i).SetNoDataValue(NDV)
        DataSet.SetDescription(bandnames[i-1])
    DataSet.FlushCache()
    return Name
```

```
In [7]: def timeseries_metrics(raster,ndv=0):
    # Make use of numpy nan functions
    # Check if type is a float array
    if not raster.dtype.name.find('float')>-1:
        raster=raster.astype(np.float32)
    # Set ndv to nan
    if ndv != np.nan:
        raster[np.equal(raster,ndv)]=np.nan
    # Build dictionary of the metrics
    tsmetrics={}
    rperc = np.nanpercentile(raster,[5,50,95],axis=0)
    tsmetrics['mean']=np.nanmean(raster,axis=0)
    tsmetrics['max']=np.nanmax(raster,axis=0)
    tsmetrics['min']=np.nanmin(raster,axis=0)
    tsmetrics['range']=tsmetrics['max']-tsmetrics['min']
    tsmetrics['median']=rperc[1]
    tsmetrics['p5']=rperc[0]
    tsmetrics['p95']=rperc[2]
    tsmetrics['prange']=rperc[2]-rperc[0]
    tsmetrics['var']=np.nanvar(raster,axis=0)
    tsmetrics['cov']=tsmetrics['var']/tsmetrics['mean']
    return tsmetrics
```

## Set the Dates

```
In [8]: # Get the date indices via pandas
dates=open(datefile).readlines()
tindex=pd.DatetimeIndex(dates)
j=1
print('Bands and dates for',imagefile)
for i in tindex:
    print("{:4d} {}".format(j, i.date()),end=' ')
    j+=1
    if j%5==1: print()

Bands and dates for S32631X398020Y1315440sS1_A_vv_0001_mtfil.vrt
  1 2015-03-22   2 2015-04-03   3 2015-04-15   4 2015-05-09   5 2015-05-21
  6 2015-06-02   7 2015-06-14   8 2015-06-26   9 2015-07-08  10 2015-07-20
 11 2015-08-01  12 2015-08-13  13 2015-08-25  14 2015-09-06  15 2015-09-18
 16 2015-09-30  17 2015-10-12  18 2015-10-24  19 2015-11-17  20 2015-11-29
 21 2015-12-11  22 2015-12-23  23 2016-01-04  24 2016-01-28  25 2016-02-09
 26 2016-03-04  27 2016-03-16  28 2016-03-28  29 2016-04-09  30 2016-04-21
 31 2016-05-03  32 2016-05-15  33 2016-05-27  34 2016-06-08  35 2016-07-02
 36 2016-07-14  37 2016-07-26  38 2016-08-07  39 2016-08-19  40 2016-08-31
 41 2016-09-12  42 2016-09-24  43 2016-10-06  44 2016-10-18  45 2016-10-30
 46 2016-11-11  47 2016-11-23  48 2016-12-05  49 2016-12-17  50 2016-12-29
 51 2017-01-10  52 2017-01-22  53 2017-02-03  54 2017-02-15  55 2017-02-27
 56 2017-03-11  57 2017-03-23  58 2017-04-04  59 2017-04-16  60 2017-04-28
 61 2017-05-10  62 2017-05-22  63 2017-06-03  64 2017-06-15  65 2017-06-27
 66 2017-07-09  67 2017-07-21  68 2017-08-02  69 2017-08-14  70 2017-08-26
 71 2017-09-07  72 2017-09-19  73 2017-10-13  74 2017-10-25  75 2017-11-06
 76 2017-11-18  77 2017-11-30
```

## Explore the Images

Below are a couple of plots showing the dataset

Open the image and get dimensions (bands,lines,pixels):

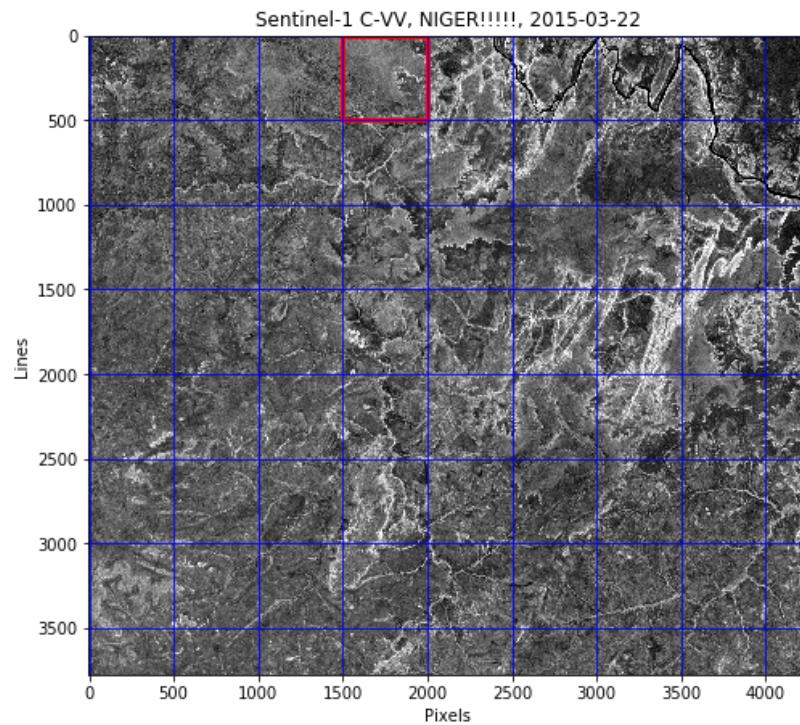
```
In [9]: img=gdal.Open(imagefile)
        img.RasterCount,img.RasterYSize,img.RasterXSize

Out[9]: (77, 3776, 4243)
```

For a managable size we choose a 1000x1000 pixel subset to read the entire data stack. We also convert the amplitude data to power data right away and will perform the rest of the calculations on the power data to be mathematically correct. NOTE: Choose a different xscale/yscale in the subset if you need to.

```
In [10]: subset=(1500,0,500,500)    # (xoff,yoff,xsize,ysize)
bandnbr=1

rasterDN=img.GetRasterBand(bandnbr).ReadAsArray()
fig, ax = plt.subplots(figsize=(8,8))
ax.set_title('Sentinel-1 C-VV, NIGER!!!!, {}'
             .format(tindex[bandnbr-1].date()))
ax.imshow(rasterDN,cmap='gray',vmin=2000,vmax=8000)
ax.grid(color='blue')
ax.set_xlabel('Pixels')
ax.set_ylabel('Lines')
# plot the subset as rectangle
if subset != None:
    _=ax.add_patch(patches.Rectangle((subset[0],subset[1]),
                                     subset[2],subset[3],
                                     fill=False,edgecolor='red',
                                     linewidth=3))
```



```
In [11]: rasterDN=img.ReadAsArray(*subset)
mask=rasterDN==0
CF=np.power(10.,-8.3)
rasterPwr=np.ma.array(np.power(rasterDN,2.)*CF,mask=mask,dtype=np.float32)
# Code below is an example to generate an 8bit scaled dB image
# rasterDB=(10.*np.ma.log10(rasterPwr)+31)/0.15
# rasterDB[rasterDB<1.]=1.
# rasterDB[rasterDB>255.]=255.
# rasterDB=rasterDB.astype(np.uint8)
# rasterDB=rasterDB.filled(0)
```

We make an RGB stack to display the first, center, and last time step as a multi-temporal color composite. The np.dstack results in an array of the form [lines,pixels,bands], which is the format we need for RGB display with matplotlib's imshow() function.

**Note** that numpy array indexing starts with 0, so band 1 is raster[0].

```
In [12]: rgb_bands=(1,int(img.RasterCount/2),img.RasterCount) # first, center, last band
rgb_bands=(1,10,40)
rgb_bands=(18,45,74)
rgb_idx=np.array(rgb_bands)-1 # get array index from bands by subtracting 1
rgb=np.dstack((rasterPwr[rgb_idx[0]],rasterPwr[rgb_idx[1]],rasterPwr[rgb_idx[2]]))
rgb_dates=(tindex[rgb_idx[0]].date(),
           tindex[rgb_idx[1]].date(),tindex[rgb_idx[2]].date())
```

We are also interested in displaying the image enhanced with histogram equalization.

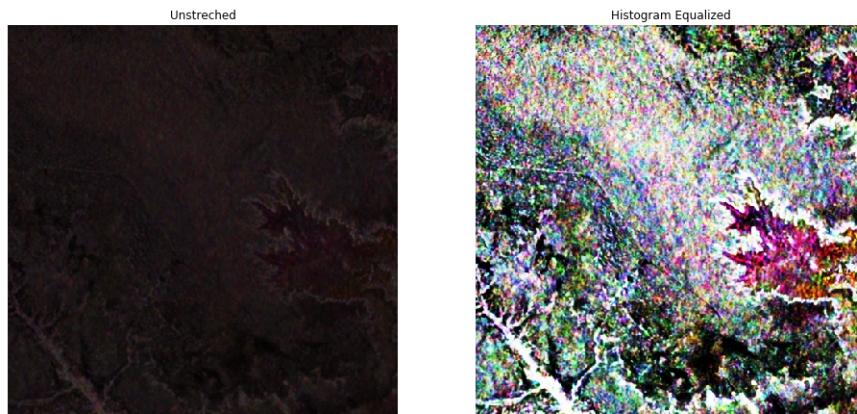
We can use the function `*exposure.equalize_hist()`\* from the skimage.exposure module

```
In [13]: rgb_stretched=rgb.copy()
# For each band we apply the stretch
for i in range(rgb_stretched.shape[2]):
    rgb_stretched[:, :, i] = exposure.\
        equalize_hist(rgb_stretched[:, :, i].data,
                      mask=~np.equal(rgb_stretched[:, :, i].data, 0.))
```

Now let's display the unstretched and histogram equalized images side by side.

```
In [14]: fig,ax = plt.subplots(1,2,figsize=(16,8))
fig.suptitle('Multi-temporal Sentinel-1 backscatter image R:{} G:{} B:{}'
             .format(rgb_dates[0],rgb_dates[1],rgb_dates[2]))
plt.axis('off')
ax[0].imshow(rgb)
ax[0].set_title('Unstretched')
ax[0].axis('off')
ax[1].imshow(rgb_stretched)
ax[1].set_title('Histogram Equalized')
_=ax[1].axis('off')
```

Multi-temporal Sentinel-1 backscatter image R:2015-10-24 G:2016-10-30 B:2017-10-25



# Computation and Visualization of the Time Series Metrics

For the entire time series, we will compute some metrics that will aid us in change detection. For each pixel in the stack we compute:

- Mean
- Median
- Maximum
- Minimum
- Range (Maximum - Minimum)
- 5th Percentile
- 95th Percentile
- PRange (95th - 5th Percentile)
- Variance
- Coefficient of Variation (Variance/Mean)

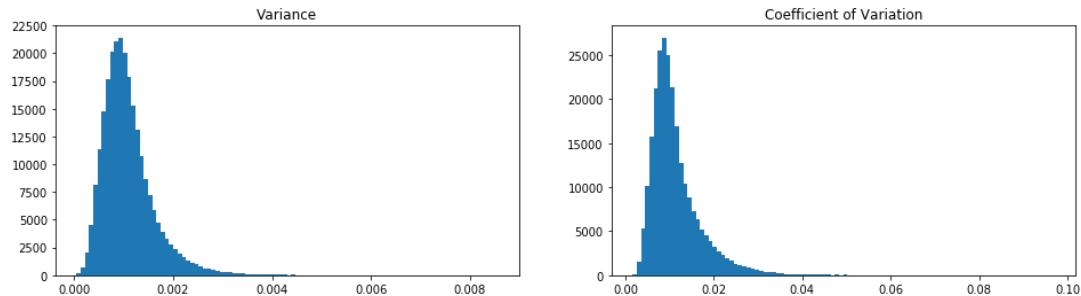
```
In [15]: metrics=timeseries_metrics(rasterPwr.filled(np.nan),ndv=np.nan)
```

```
In [16]: #Print out what the various metrics keys are  
metrics.keys()
```

```
Out[16]: dict_keys(['mean', 'max', 'min', 'range', 'median', 'p5', 'p95', 'prange', 'var', 'cov'])
```

Let's look at the histograms for the time series variance and coefficient of variation to aid displaying those images:

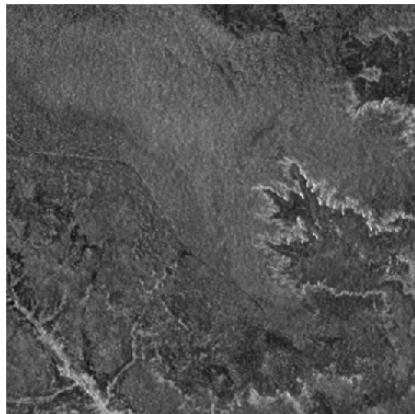
```
In [17]: fig, ax= plt.subplots(1,2,figsize=(16,4))  
ax[0].hist(metrics['var'].flatten(),bins=100)  
ax[1].hist(metrics['cov'].flatten(),bins=100)  
_ =ax[0].set_title('Variance')  
_ =ax[1].set_title('Coefficient of Variation')
```



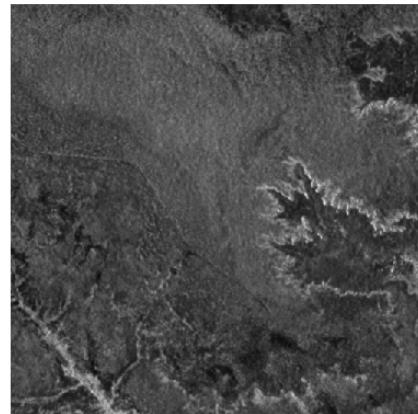
We use thresholds determined from those histograms to set the scaling in the time series visualization. For the backscatter metrics we choose a typical range appropriate for this ecosystem and radar sensor. A typical range is -30 dB (0.0001) to -5.2 dB (0.3).

```
In [18]: # List the metrics keys you want to plot
metric_keys=['mean', 'median', 'max', 'min',
            'p95', 'p5','range', 'prange','var','cov']
fig= plt.figure(figsize=(16,40))
idx=1
for i in metric_keys:
    ax = fig.add_subplot(5,2, idx)
    if i=='var': vmin,vmax=(0.0,0.005)
    elif i == 'cov': vmin,vmax=(0.,0.04)
    else:
        vmin,vmax=(0.0001,0.3)
    ax.imshow(metrics[i],vmin=vmin,vmax=vmax,cmap='gray')
    ax.set_title(i.upper())
    ax.axis('off')
    idx+=1
```

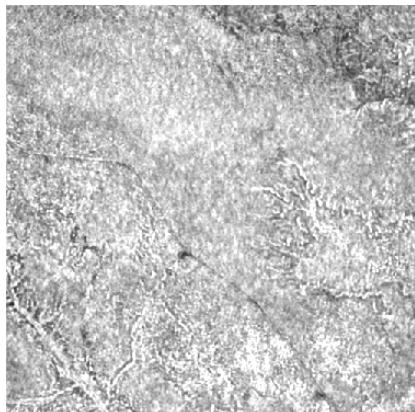
MEAN



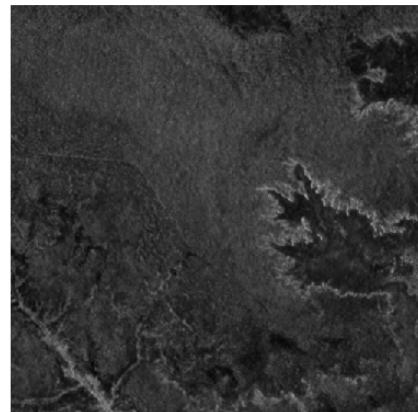
MEDIAN



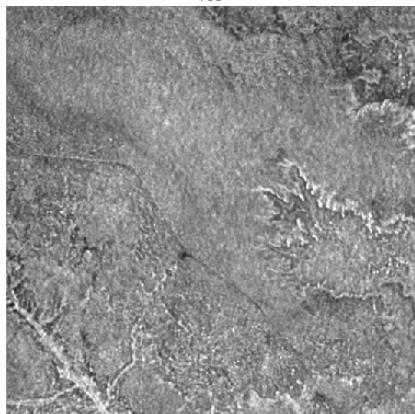
MAX



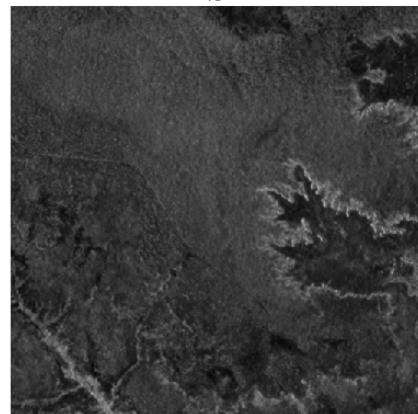
MIN

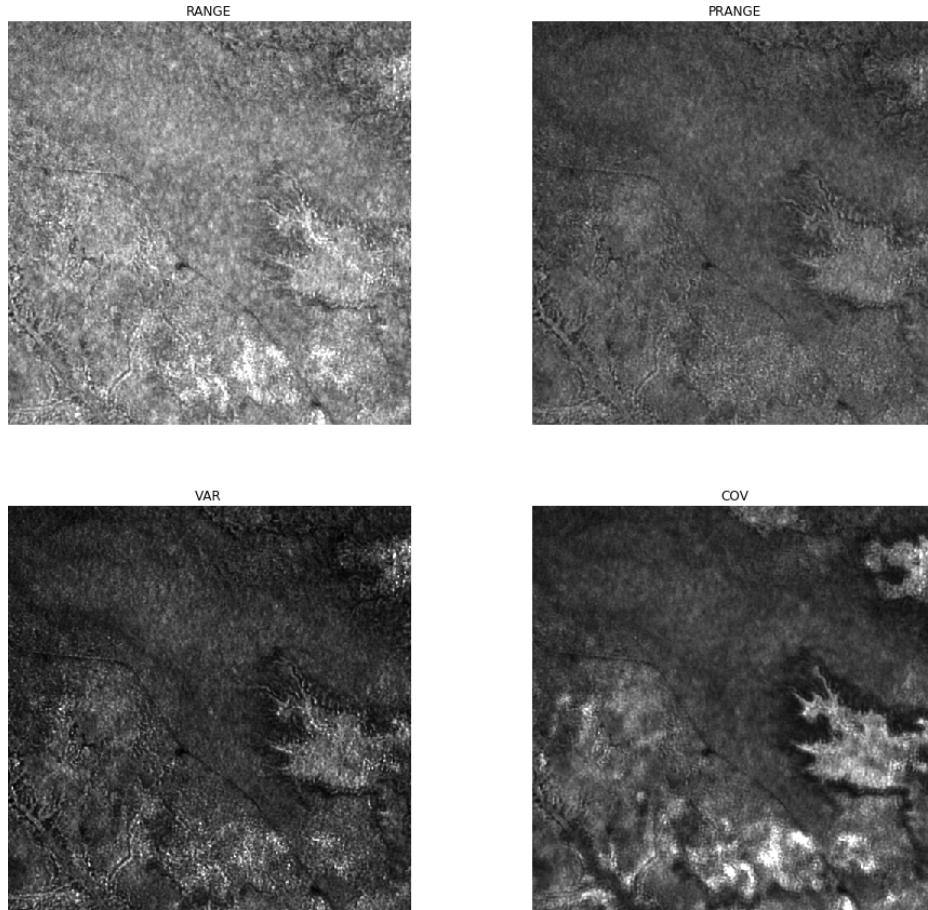


P95



P5



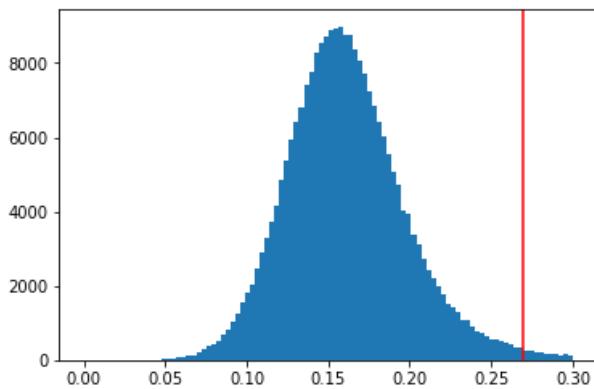


## Change detection with the Percentile Difference Threshold Method

In this method we find thresholds on the 95<sup>th</sup> and 5<sup>th</sup> percentile difference. The advantage to look at percentiles versus maximum minus minimum is that outliers and extrema in the time series are not influencing the result.

For our example, the histogram of the 95<sup>th</sup> and 5<sup>th</sup> percentile difference image looks like this:

```
In [19]: plt.hist(metrics['range'].flatten(), bins=100, range=(0,0.3))
         _=plt.axvline(0.27, color='red')
```

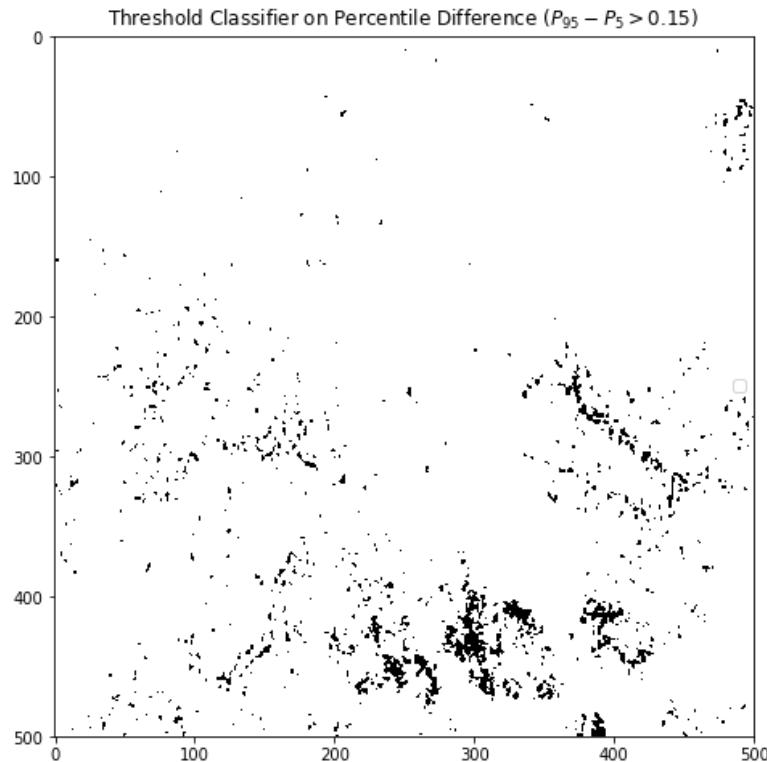


Let's visualize the change pixels (cp) where the 95th - 5th percentile difference in the time series for each pixel (x,y) exceed a threshold  $t$

$$cp_{x,y} = P_{x,y}^{95th} - P_{x,y}^{5th} > t$$

With  $t = 0.15$  the image looks like:

```
In [20]: thres=0.25
plt.figure(figsize=(8,8))
mask=metrics['range']<thres # For display we prepare the inverse mask
maskpdiff=~mask # Store this for later output
plt.imshow(mask,cmap='gray')
plt.legend(['$p_{95} - p_5 > 0.15$'],loc='center right')
_=plt.title('Threshold Classifier on Percentile Difference ($P_{95} - P_5 > 0.15$)')
```



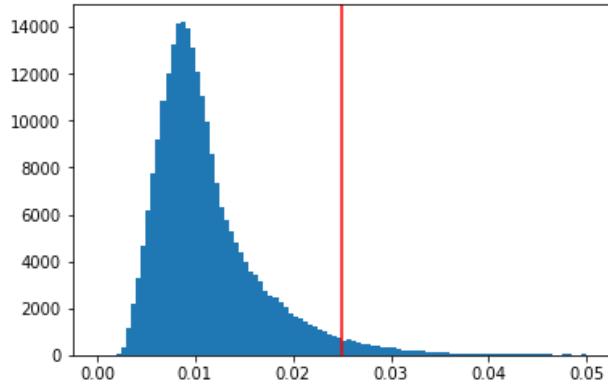
## Change Detection with the Coefficient of Variation Method

We can set a threshold  $t$  for the coefficient of variation image to classify change in the time series:

$$cp_{x,y} = \frac{\sigma_{x,y}^2}{\bar{X}_{x,y}} > t$$

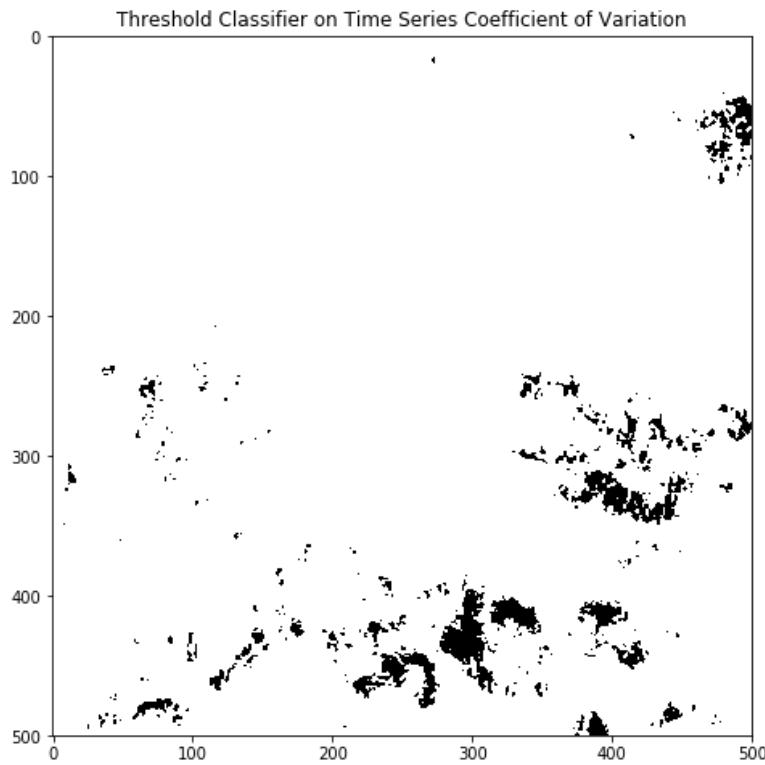
Let's look at the histogram of the coefficient of variation:

```
In [21]: plt.hist(metrics['cov'].flatten(),bins=100,range=(0,0.05))
_=plt.axvline(0.025,color='red')
```



With a threshold  $t=0.01$  the change pixels would look like the following image:

```
In [22]: thres=0.025
mask=metrics['cov'] < thres
maskcv=~mask
plt.figure(figsize=(8,8))
plt.imshow(mask,cmap='gray')
_=plt.title('Threshold Classifier on Time Series Coefficient of Variation')
```



## Change Detection with the Log Ratio Method

We compare two images from the same season in different years. First we look at global means of the backscatter images in the subset building a time series object of acquisition dates and global image means of backscatter.

```
In [23]: tsmean=10*np.log10(np.nanmean(rasterPwr.filled(np.nan),axis=(1,2)))
```

We make a time series object to list the dates, mean backscatter in dB, and band index number for the rasterPwr array:

```
In [24]: ts = pd.Series(tsmean,index=tindex)
for i in range(len(ts)):
    print(i,ts.index[i].date(),ts[i])
```

0	2015-03-22	-9.773781	31	2016-05-15	-11.248089
1	2015-04-03	-9.814333	32	2016-05-27	-10.781347
2	2015-04-15	-9.84827	33	2016-06-08	-10.7717905
3	2015-05-09	-10.075288	34	2016-07-02	-10.622729
4	2015-05-21	-9.987606	35	2016-07-14	-10.262638
5	2015-06-02	-9.835003	36	2016-07-26	-9.969166
6	2015-06-14	-10.412914	37	2016-08-07	-9.227007
7	2015-06-26	-10.64331	38	2016-08-19	-8.372538
8	2015-07-08	-9.98234	39	2016-08-31	-7.8771267
9	2015-07-20	-9.159636	40	2016-09-12	-9.163029
10	2015-08-01	-7.678219	41	2016-09-24	-9.04641
11	2015-08-13	-8.60141	42	2016-10-06	-10.078144
12	2015-08-25	-7.6070075	43	2016-10-18	-10.534364
13	2015-09-06	-7.645421	44	2016-10-30	-11.044583
14	2015-09-18	-6.655918	45	2016-11-11	-11.120414
15	2015-09-30	-8.7717705	46	2016-11-23	-11.056729
16	2015-10-12	-9.348694	47	2016-12-05	-11.187023
17	2015-10-24	-9.547744	48	2016-12-17	-11.514052
18	2015-11-17	-10.2138815	49	2016-12-29	-11.376835
19	2015-11-29	-11.099142	50	2017-01-10	-11.243304
20	2015-12-11	-11.029471	51	2017-01-22	-11.204616
21	2015-12-23	-11.332901	52	2017-02-03	-11.176929
22	2016-01-04	-11.346351	53	2017-02-15	-11.093778
23	2016-01-28	-11.197915	54	2017-02-27	-11.04459
24	2016-02-09	-11.145014	55	2017-03-11	-10.92975
25	2016-03-04	-10.9366045	56	2017-03-23	-10.895084
26	2016-03-16	-11.114582	57	2017-04-04	-11.270055
27	2016-03-28	-11.000681	58	2017-04-16	-11.106432
28	2016-04-09	-10.456753	59	2017-04-28	-11.091718
29	2016-04-21	-11.031124	60	2017-05-10	-11.196309
30	2016-05-03	-11.042203	61	2017-05-22	-10.516581
			62	2017-06-03	-11.056223

```

63 2017-06-15 -10.081734
64 2017-06-27 -10.121447
65 2017-07-09 -10.04177
66 2017-07-21 -9.837778
67 2017-08-02 -9.248106
68 2017-08-14 -8.226735
69 2017-08-26 -8.096363
70 2017-09-07 -9.48612
71 2017-09-19 -9.06922
72 2017-10-13 -10.274279
73 2017-10-25 -10.869721
74 2017-11-06 -11.154692
75 2017-11-18 -11.122526
76 2017-11-30 -11.273689

```

**To compare two dates for change detection with the log ratio approach we pick two dates of relative low backscatter (dry conditions) and from similar times of the year. Two such candidate dates are:**

West Africa / Biomass Site example:

- 2015-11-29 -11.099142 dB (index 19)
- 2017-11-30 -11.273689 dB (index 76)

```
In [26]: # WA biomass
Xr=rasterPwr[19]    # Reference Image
Xi=rasterPwr[76]    # New Image
```

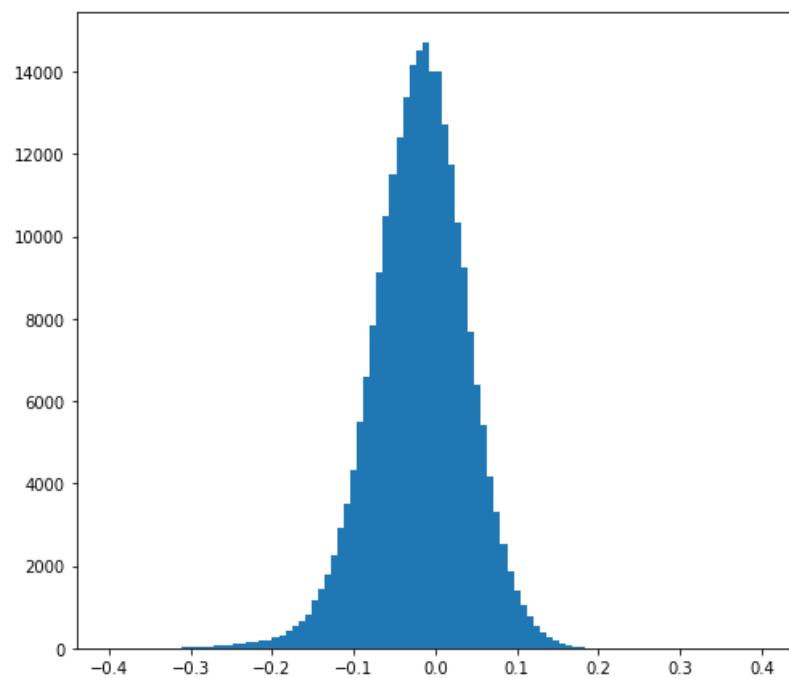
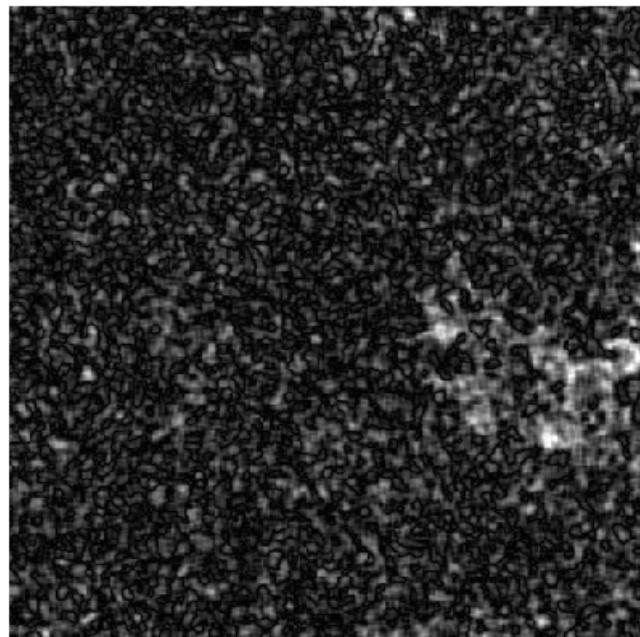
The Log ratio between the images is:

$$r = \log_{10}\left(\frac{X_i}{X_r}\right)$$

```
In [27]: r = np.log10(Xi/Xr)
```

To find a threshold for change, we can display the absolute ration image  $\text{abs}(r)$  and the histogram of  $r$ . We adjust the scale factors for the display to enhance visualization of change areas with largest backscatter change over the time series. Brighter values show larger change.

```
In [28]: # Display r
fig, ax = plt.subplots(2,1,figsize=(8,16))
ax[0].axis('off')
ax[0].imshow(np.abs(r),vmin=0,vmax=0.3,cmap='gray')
_=ax[1].hist(r.flatten(),bins=100,range=(-0.4,0.4))
```



Let's define change pixels as those falling outside the range of **three** times the standard deviation of the ration image  $\sigma_r$  from the image mean  $\bar{r}$ :

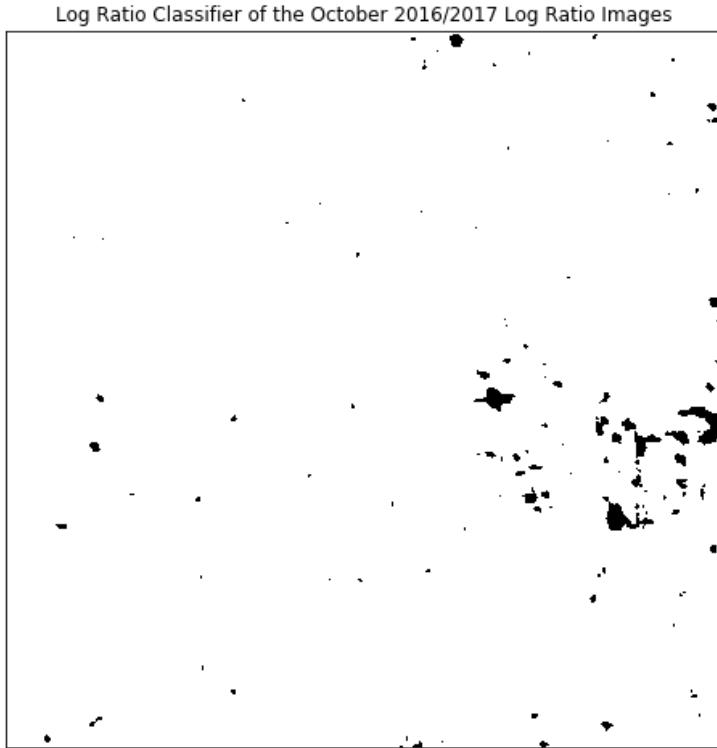
$$cp_{x,y} = (r_{x,y} < \bar{r} - 3\sigma_r) \text{ or } (r_{x,y} > \bar{r} + 3\sigma_r)$$

We are using the numpy masking to set the non-changing pixels inside the range:

```
In [29]: stddev=np.std(r)
thres=3*stddev
mask=np.logical_and(r>-1*thres,r<thres)
masklr=~mask
```

Let's display pixels that fall outside 3 times the standard deviation

```
In [30]: fig,ax = plt.subplots(figsize=(8,16))
ax.imshow(mask,cmap='gray')
ax.xaxis.set_ticks([])
ax.yaxis.set_ticks([])
_=ax.set_title('Log Ratio Classifier of the October 2016/2017 Log Ratio Images')
```



## Write the images to an output file

### Determine output geometry

First, we need to set the correct geotransformation and projection information. We retrieve the values from the input images and adjust by the subset:

```
In [31]: proj=img.GetProjection()
geotrans=list(img.GetGeoTransform())

subset_xoff=geotrans[0]+subset[0]*geotrans[1]
subset_yoff=geotrans[3]+subset[1]*geotrans[5]
geotrans[0]=subset_xoff
geotrans[3]=subset_yoff
geotrans=tuple(geotrans)
geotrans

Out[31]: (428020.0, 20.0, 0.0, 1390960.0, 0.0, -20.0)
```

## Time series metrics images

We use the root of the time series data stack name and append a `tsmetrics.tif` ending as filenames

```
In [32]: # Time Series Metrics as image:
# We make a new subdirectory where we will store the images
dirname=imagefile.replace('.vrt','_tsmetrics2')
os.makedirs(dirname,exist_ok=True)
print(dirname)

S32631X398020Y1315440sS1_A_vv_0001_mtfil_tsmetrics2
```

### Output the individual metrics as GeoTIFF images:

```
In [33]: Names=[] # List to keep track of all the names
for i in metrics:
    # Name, Array, DataType, NDV,bandnames=None,ref_image
    Name=os.path.join(dirname,imagefile.replace('.vrt','_'+str(i)+'.tif'))
    CreateGeoTiff(Name,metrics[i],gdal.GDT_Float32,np.nan,[i],GeoT=geotrans,Projection=proj)
    Names.append(Name)
```

### Build a Virtual Raster Table on the Metrics GeoTIFF images

To tie the images in to one new raster stack of time series metrics we build a virtual raster table with all the metrics.

Trick: Use `''.join(Names)` to build one long string of names separated by a space as input to `gdalbuildvrt`

```
In [34]: cmd='gdalbuildvrt -separate -overwrite -vrtnodata nan '+'\
dirname+'.vrt '+' '.join(Names)
# print(cmd)
os.system(cmd)

Out[34]: 0
```

```
In [35]: os.getcwd()

Out[35]: '/Users/rmuench/Downloads/wa/BIOss1'
```

```
In [36]: print('Time Series Metrics VRT File:\n',dirname+'.vrt')

Time Series Metrics VRT File:
S32631X398020Y1315440sS1_A_vv_0001_mtfil_tsmetrics2.vrt
```

## Change Images from the three methods

We are going to write one three-band GeoTIFF output file that stores the results from the three classifiers

```
In [37]: imangename=imagefile.replace('.vrt','_thresholds.tif')
bandnames=['Percentile','COV','Log Ratio']
Array=np.array([maskpdiff,maskcv,masklr])
CreateGeoTiff(imangename,Array,gdal.GDT_Byte,0,bandnames,GeoT=geotrans,Projection=
proj)

Out[37]: 'S32631X398020Y1315440sS1_A_vv_0001_mtfil_thresholds.tif'
```

This image can now be loaded into QGIS or similar programs and only the detected layers should show.

## Conclusion

Thresholds for the three methods are site dependent and need to be identified with calibration data or visual post-classification interpretation, and can subsequently be adjusted to maximize classification accuracy. Also, some methods will have advantages in different scenarios.

At the Earth Big Data SEPO Processor we actually transform many of the time series metrics data types back to lower volume storage models, e.g. 16 bit scaled amplitudes. See the EBD Data Guide below:

[https://github.com/EarthBigData/openSAR/blob/master/doc/EBD\\_DataGuide.md](https://github.com/EarthBigData/openSAR/blob/master/doc/EBD_DataGuide.md)

[https://github.com/EarthBigData/openSAR/blob/master/doc/EBD\\_DataGuide.pdf](https://github.com/EarthBigData/openSAR/blob/master/doc/EBD_DataGuide.pdf)

## Exercises

- Change the threshold and band choices in this notebook to see the effects on detected changes.
- Load masks on the with QGIS and compare the detected areas with your time series plots and image data in QGIS.
- Look at the effect of using cross-polarized versus like-polarized polarizations

# SAR Training Workshop for Forest Applications

## PART 4 - SAR Time Series Change Point Detection

Josef Kellndorfer, Ph.D., President and Senior Scientist, Earth Big Data, LLC

Revision date: January 2019

In this chapter we introduce the advanced concepts of change point detection in time series. One of the goals of change detection for forest applications is to identify disturbance over an observation period and the timing of events. Many tools for change point detection stem from the financial sector and are available today with different complexities. In this workbook we will analyze time series signatures from SAR with emphasis on forest time series. We will start by exploring time series at pixel levels and will work up to a change point detection scenario with image based analysis.

```
In [1]: # Importing relevant python packages
import pandas as pd
import gdal
import numpy as np
import time,os

# For plotting
%matplotlib inline
import matplotlib.pylab as plt
import matplotlib.patches as patches

font = {'family' : 'monospace',
        'weight' : 'bold',
        'size'   : 18}
plt.rcParams['font', **font]
```

## Set Project Directory and Filenames

### West Africa - Biomass Site

```
In [ ]: # datadirectory='c401servir/wa/BIOssS1/'
# datefile='S32631X398020Y1315440ssS1_A_vv_0001_mtfil.dates'
# imagefile='S32631X398020Y1315440ssS1_A_vv_0001_mtfil.vrt'
```

### West Africa - Niamey Deforestation Site

```
In [2]: datadirectory='/Users/rmuench/Downloads/wa/cra/'
datefile='S32631X402380Y1491460ssS1_A_vv_0001_A_mtfil.dates'
imagefile='S32631X402380Y1491460ssS1_A_vv_0001_A_mtfil.vrt'
```

### West Africa - Dam Site

```
In [ ]: # datadirectory='/dev/shm/projects/c401servir/wa/DAMssS1/'
# datefile='S32631X232140Y1614300ssS1_A_vh_0001_A_mtfil.dates'
# imagefile='S32631X232140Y1614300ssS1_A_vh_0001_A_mtfil.vrt'
```

## HKH Site

```
In [ ]: # datadirectory='hk/h/time_series/S32644X696260Y3052060ss1-EBD'  
# datefile='S32644X696260Y3052060ss1_D_vv_0092_mtfil.dates'  
# imagefile='S32644X696260Y3052060ss1_D_vv_0092_mtfil.vrt'  
# imagefile_cross='S32644X696260Y3052060ss1_D_vh_0092_mtfil.vrt'
```

```
In [3]: # Switch to the data directory  
os.chdir(datadirectory)
```

## Acquisition Dates

Read from the Dates file the dates in the time series and make a pandas date index

```
In [4]: dates=open(datefile).readlines()  
tindex=pd.DatetimeIndex(dates)  
j=1  
print('Bands and dates for',imagefile)  
for i in tindex:  
    print("{:4d} {}".format(j, i.date()),end=' ')  
    j+=1  
    if j%5==1: print()
```

```
Bands and dates for S32631X402380Y1491460ss1_A_vv_0001_A_mtfil.vrt  
 1 2015-04-03    2 2015-11-17    3 2015-11-29    4 2015-12-11    5 2015-12-23  
 6 2016-01-04    7 2016-01-28    8 2016-02-09    9 2016-03-04    10 2016-03-16  
11 2016-03-28   12 2016-04-09   13 2016-04-21   14 2016-05-03   15 2016-05-15  
16 2016-05-27   17 2016-06-08   18 2016-07-02   19 2016-07-14   20 2016-07-26  
21 2016-08-07   22 2016-08-19   23 2016-08-31   24 2016-09-12   25 2016-09-24  
26 2016-10-06   27 2016-10-18   28 2016-10-30   29 2016-11-11   30 2016-11-23  
31 2016-12-05   32 2016-12-17   33 2016-12-29   34 2017-01-10   35 2017-01-22  
36 2017-02-03   37 2017-02-15   38 2017-02-27   39 2017-03-11   40 2017-03-23  
41 2017-04-04   42 2017-04-16   43 2017-04-28   44 2017-05-10   45 2017-05-22  
46 2017-06-03   47 2017-06-15   48 2017-06-27   49 2017-07-09   50 2017-07-21  
51 2017-08-02   52 2017-08-14   53 2017-08-26   54 2017-09-07   55 2017-09-19  
56 2017-10-13   57 2017-10-25   58 2017-11-06   59 2017-11-18   60 2017-11-30
```

## Image data

Get the time series raster stack from the entire training data set.

```
In [5]: rasterstack=gdal.Open(imagefile).ReadAsArray()
```

## Data Pre-Processing

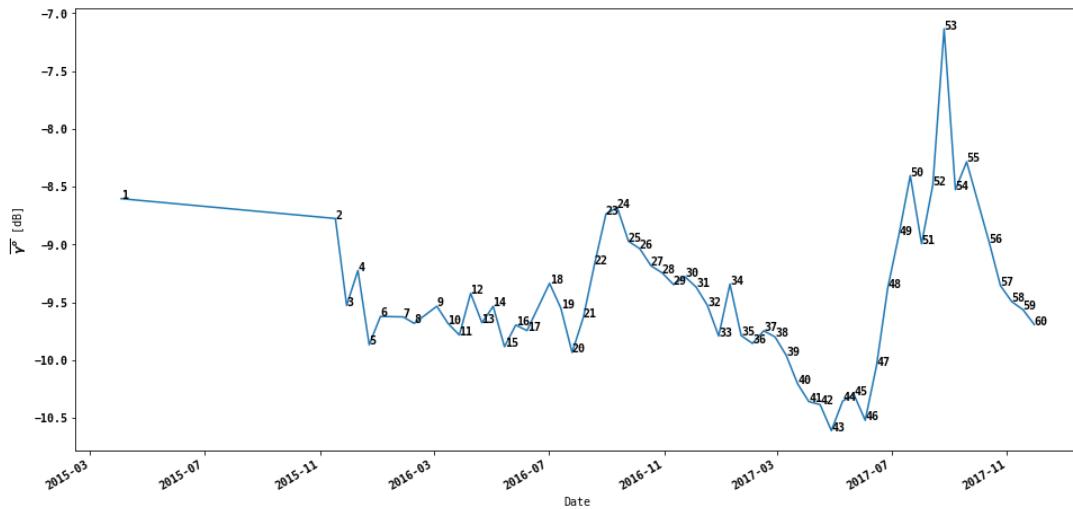
### Plot the global means of the Time Series

1. Conversion to power
2. Compute means
3. Convert to dB
4. Make a pandas time series
5. Plot time series of means

```
In [6]: # 1. Conversion to Power
caldB=-83
calPwr = np.power(10.,caldB/10.)
rasterstack_pwr = np.power(rasterstack,2.)*calPwr
# 2. Compute Means
rs_means_pwr = np.mean(rasterstack_pwr, axis=(1,2))
# 3. Convert to dB
rs_means_dB = 10.*np.log10(rs_means_pwr)
```

```
In [7]: # 4. Make a pandas time series object
ts = pd.Series(rs_means_dB, index=tindex)
```

```
In [8]: # 5. Use the pandas plot function of the time series object to plot
# Put band numbers as data point labels
plt.figure(figsize=(16,8))
ts.plot()
xl = plt.xlabel('Date')
yl = plt.ylabel('$\overline{\gamma^o}$ [dB]')
for xyb in zip(ts.index,rs_means_dB,range(1,len(ts)+1)):
    plt.annotate(xyb[2],xy=xyb[0:2])
```



## EXERCISE

Look at the global means plot and determine from the tindex array at which dates you see maximum and minimum values. Are relative peaks associated with seasons?

# Generate Time Series for Point Locations or Subsets

In python we can use the matrix slicing rules (Like Matlab) to obtain subsets of the data. For example to pick one pixel at a line/pixel location and obtain all band values, use:

```
[:,line,pixel] notation.
```

Or, if we are interested in a subset at an offset location we can use:

```
[:,yoffset:(yoffset+yrange),xoffset:(xoffset+xrange)]
```

In the section below we will learn how to generate time series plots for point locations (pixels) or areas (e.g. a 5x5 window region). To show individual bands, we define a `showImage` function which incorporates the matrix slicing from above.

```
In [9]: def showImage(rasterstack,tindex,bandnbr,subset=None,vmin=None,vmax=None):
    '''Input:
       rasterstack stack of images in SAR power units
       tindex time series date index
       bandnbr bandnumber of the rasterstack to dissplay'''
    fig = plt.figure(figsize=(16,8))
    ax1 = fig.add_subplot(121)
    ax2 = fig.add_subplot(122)

    # If vmin or vmax are None we use percentiles as limits:
    if vmin==None: vmin=np.percentile(rasterstack[bandnbr-1].flatten(),5)
    if vmax==None: vmax=np.percentile(rasterstack[bandnbr-1].flatten(),95)

    ax1.imshow(rasterstack[bandnbr-1],cmap='gray',vmin=vmin,vmax=vmax)
    ax1.set_title('Image Band {} {}'.format(bandnbr,tindex[bandnbr-1].date()))
    if subset== None:
        bands,ydim,xdim=rasterstack.shape
        subset=(0,0,xdim,ydim)

    ax1.add_patch(patches.Rectangle((subset[0],subset[1]),subset[2],subset[3],fill=False,edgecolor='red'))
    ax1.xaxis.set_label_text('Pixel')
    ax1.yaxis.set_label_text('Line')

    ts_pwr=np.mean(rasterstack[:,subset[1]:subset[1]+subset[3]],
                    subset[0]:subset[0]+subset[2]],axis=(1,2))
    ts_dB=10.*np.log10(ts_pwr)
    ax2.plot(tindex,ts_dB)
    ax2.yaxis.set_label_text('$\gamma^o$ [dB]')
    ax2.set_title('$\gamma^o$ Backscatter Time Series')
    # Add a vertical line for the date where the image is displayed
    ax2.axvline(tindex[bandnbr-1],color='red')

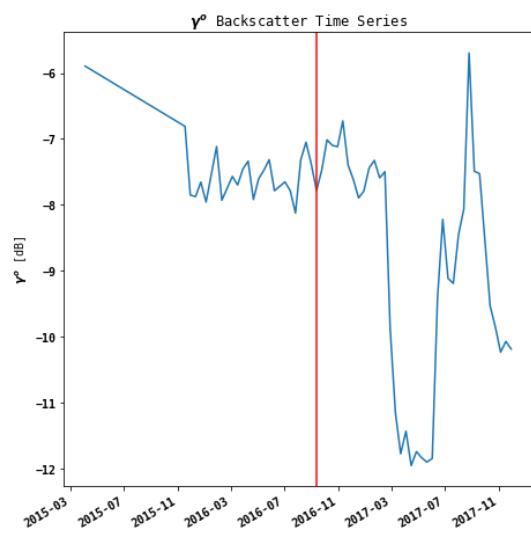
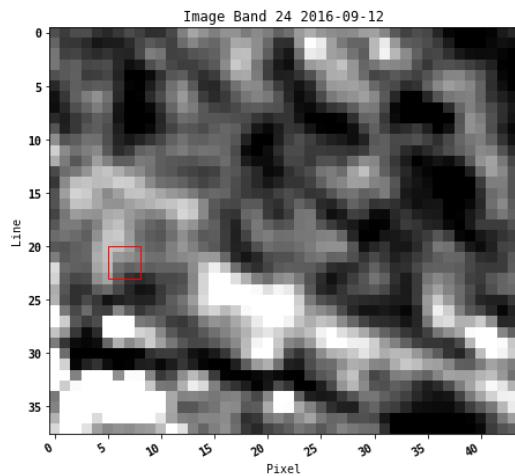
    fig.autofmt_xdate()
```

## Exercise

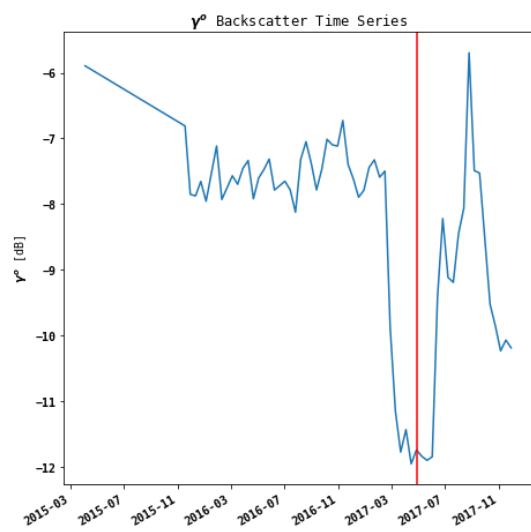
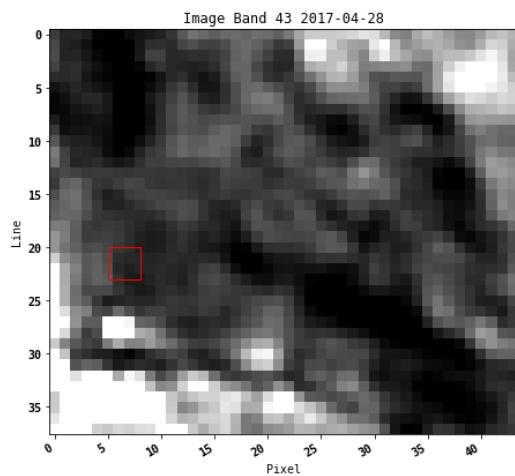
Compare band 24 and band 43 visually

```
In [10]: bandnbr=24 #  
subset=[5,20,3,3]  
# subset=[30,15,3,3]  
# subset=[12,10,3,3]
```

```
In [11]: showImage(rasterstack_pwr,tindex,bandnbr,subset)
```



```
In [12]: bandnbr=43  
showImage(rasterstack_pwr,tindex,bandnbr,subset)
```



## EXERCISE

For subset (5,20,3,3):

1. What are the dates where backscatter falls below - 11 dB?
2. Compute the gradients (backscatter difference between two consecutive dates).
3. What is the largest gradient of backscatter drop between two consecutive dates?
4. What are the dates associated with this gradient (before and after)?

### Helper function to generate a time series object

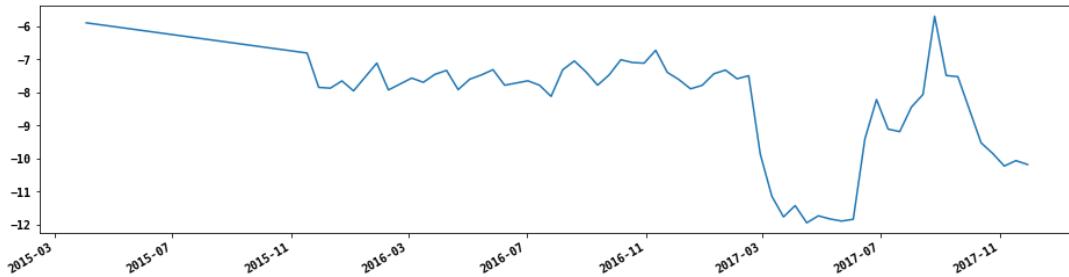
```
In [13]: def timeSeries(rasterstack_pwr,tindex,subset,ndv=0.):  
    # Extract the means along the time series axes  
    # raster shape is time steps, lines, pixels.  
    # With axis=1,2, we average lines and pixels for each time  
    # step (axis 0)  
    raster=rasterstack_pwr.copy()  
    if ndv != np.nan: raster[np.equal(raster,ndv)]=np.nan  
    ts_pwr=np.nanmean(raster[:,subset[1]:(subset[1]+subset[3]),  
                        subset[0]:(subset[0]+subset[2])],axis=(1,2))  
    # convert the means to dB  
    ts_dB=10.*np.log10(ts_pwr)  
    # make the pandas time series object  
    ts = pd.Series(ts_dB,index=tindex)  
    # return it  
    return ts
```

Using the timeSeries(...) function to make a time series object for the chosen subset:

```
In [14]: ts = timeSeries(rasterstack_pwr,tindex,subset)
```

Plot the object:

```
In [15]:_=ts.plot(figsize=(16,4)) #_= is a trick to suppress more output.
```



### ENTER YOUR CODE HERE

```
In [ ]: # 1. What are the dates where backscatter falls below - 11 dB?
```

```
In [ ]: # 2. Compute the gradients (backscatter difference  
# between two consecutive dates.
```

```
In [ ]: # 3. What is the largest gradient of backscatter drop  
# between two consecutive dates?
```

```
In [ ]: # What are the dates associated with this gradient  
# (before and after)
```

**Question:** Can you field verify that change occurred at this location between these two dates?

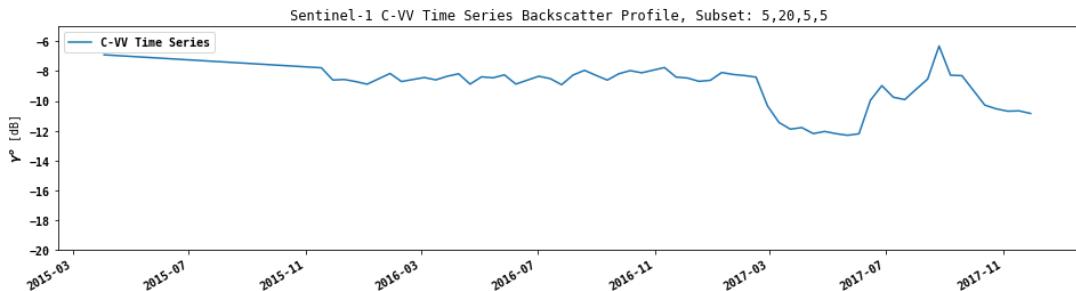
## Seasonal Subsets of time series records

Let's expand upon SAR time series analysis. Often it is desirable to subset time series by season or months to compare with similar conditions of a previous year's observation. For example, in analyzing C-Band backscatter data, it might be useful to limit comparative analysis to dry season observations only as soil moisture might confuse signals during the wet seasons. In this section we will expand upon the concepts of subsetting time series along the time axis. We will make use of the pandas datetime index tools:

- Month
- Day of year

First we extract the time series again for an area at the subset location (5,20,5,5). We then convert the pandas time series to a pandas DataFrame to allow for more processing options. We also label the data value column as 'g0' for gamma0:

```
In [16]: subset=(5,20,5,5)  
ts = timeSeries(rasterstack_pwr,tindex,subset)  
tsdf = pd.DataFrame(ts,index=ts.index,columns=[ 'g0' ])  
  
# Plot  
ylim=(-20,-5)  
tsdf.plot(figsize=(16,4))  
plt.title('Sentinel-1 C-VV Time Series Backscatter Profile, Subset: 5,20,5,5 ')  
plt.ylabel('$\gamma^0$ [dB]')  
plt.ylim(ylim)  
_=plt.legend(['C-VV Time Series'])
```

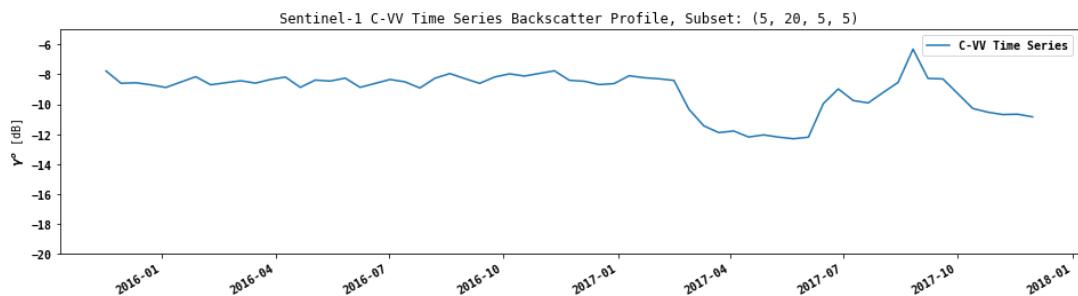


### Start the time series in November 2015

We can use the pandas index parameters like month to make seasonal subsets

```
In [17]: tsdf_sub1=tsdf[tsdf.index>'2015-11-01']

# Plot
tsdf_sub1.plot(figsize=(16,4))
plt.title('Sentinel-1 C-VV Time Series Backscatter Profile, Subset: {}'.format(subset))
plt.ylabel('$\gamma^o$ [dB]')
plt.ylim(ylim)
_=plt.legend(["C-VV Time Series"])
```



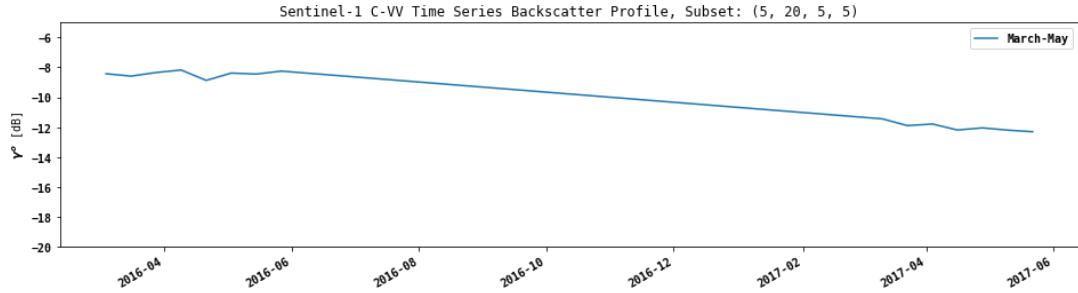
### Subset by months:

We can make use of pandas DateTimeIndex object `index.month` and numpy's `logical_and` function to subset a time series easily by month.

#### March to May data only

```
In [18]: tsdf_sub2=tsdf_sub1[
    np.logical_and(tsdf_sub1.index.month>=3,tsdf_sub1.index.month<=5)

# Plot
fig, ax = plt.subplots(figsize=(16,4))
tsdf_sub2.plot(ax=ax)
plt.title('Sentinel-1 C-VV Time Series Backscatter Profile, Subset: {}'.format(subset))
plt.ylabel('$\gamma^o$ [dB]')
plt.ylim(ylim)
_=plt.legend(["March-May"])
```

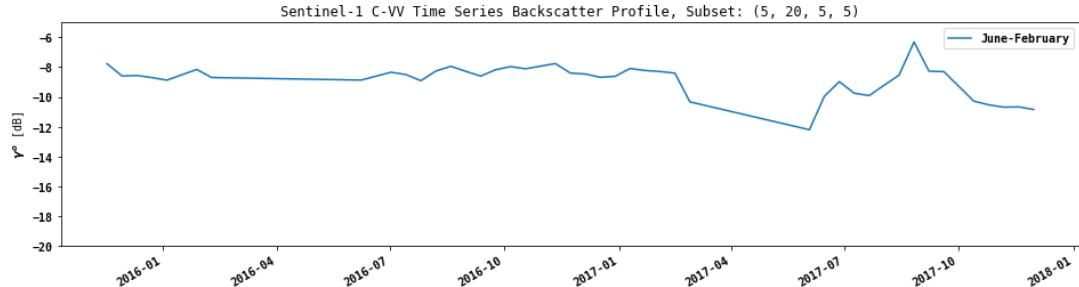


#### All other months

Using numpy's boolean `invert` function, we can invert a selection and in this example get to all other months:

```
In [19]: tsdf_sub3=tsdf_sub1[np.invert(
    np.logical_and(tsdf_sub1.index.month>=3,tsdf_sub1.index.month<=5))]

# Plot
fig, ax = plt.subplots(figsize=(16,4))
tsdf_sub3.plot(ax=ax)
plt.title('Sentinel-1 C-VV Time Series Backscatter Profile, Subset: {}'
          .format(subset))
plt.ylabel('$\gamma^o$ [dB]')
plt.ylim(ylim)
_=plt.legend(["June-February"])
```

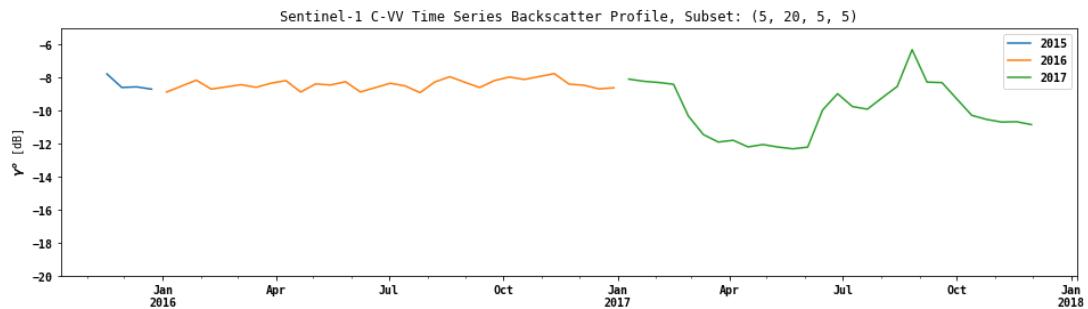


### Group time series by year to compare average backscatter values

```
In [20]: ts_sub_by_year = tsdf_sub1.groupby(pd.Grouper(freq="Y"))
```

```
In [21]: fig, ax = plt.subplots(figsize=(16,4))
for label, df in ts_sub_by_year:
    df.g0.plot(ax=ax, label=label.year)
plt.legend()
# ts_sub_by_year.plot(ax=ax)
plt.title('Sentinel-1 C-VV Time Series Backscatter Profile, Subset: {}'
          .format(subset))
plt.ylabel('$\gamma^o$ [dB]')
plt.ylim(ylim)
```

```
Out[21]: (-20, -5)
```



## Make a pivot table to group year and sort by day of year for plotting overlapping time series

First we add two coluns to the data frame:

- Day of year (doy)
- Year

```
In [22]: # Add doy
tsdf_sub1 = tsdf_sub1.assign(doy=tsdf_sub1.index.dayofyear)
# Add year
tsdf_sub1 = tsdf_sub1.assign(year=tsdf_sub1.index.year)
```

Then a pivot table gets created which has day of year as the index and years as columns:

```
In [23]: piv=pd.pivot_table(tsdf_sub1,index=['doy'],columns=['year'],values=['g0'])
# Set the names for the column indices
piv.columns.set_names(['g0','Year'],inplace=True)
print(piv.head(10))
print('...\\n',piv.tail(10))
```

	g0	g0	g0
Year	2015	2016	2017
doy			
4	NaN	-8.874602	NaN
10	NaN	NaN	-8.091206
22	NaN	NaN	-8.222770
28	NaN	-8.155600	NaN
34	NaN	NaN	-8.294136
40	NaN	-8.695752	NaN
46	NaN	NaN	-8.402759
58	NaN	NaN	-10.330054
64	NaN	-8.426312	NaN
70	NaN	NaN	-11.441220
...			
	g0	g0	g0
Year	2015	2016	2017
doy			
321	-7.774510	NaN	NaN
322	NaN	NaN	-10.665520
328	NaN	-8.395135	NaN
333	-8.594952	NaN	NaN
334	NaN	NaN	-10.840596
340	NaN	-8.461259	NaN
345	-8.560352	NaN	NaN
352	NaN	-8.681982	NaN
357	-8.698992	NaN	NaN
364	NaN	-8.615916	NaN

```
In [24]: piv.columns.set_names(['g0','year'],inplace=True)
```

As we can see, there are NaN (Not a Number) values on the days in a year where no acquisition took place. Now we use time weighted interpolation to fill the dates for all the observations in any given year. For **time weighted interpolation** to work we need to create a dummy year as a date index, perform the interpolation, and reset the index to the day of year. This is accomplished with the following steps:

```
In [25]: # Add fake dates for year 100 to enable time sensitive interpolation
# of missing values in the pivot table
year_doy = ['2100-{}'.format(x) for x in piv.index]
y100_doy=pd.DatetimeIndex(pd.to_datetime(year_doy,format='%Y-%j'))

# make a copy of the piv table and add two columns
piv2=piv.copy()
piv2=piv2.assign(d100=y100_doy) # add the fake year dates
piv2=piv2.assign(doy=piv2.index) # add doy as a column to replace as index later
again

# Set the index to the dummy year
piv2.set_index('d100',inplace=True,drop=True)

# PERFORM THE TIME WEIGHTED INTERPOLATION
piv2 = piv2.interpolate(method='time') # TIME WEIGHTED INTERPOLATION!

# Set the index back to day of year.
piv2.set_index('doy',inplace=True,drop=True)
```

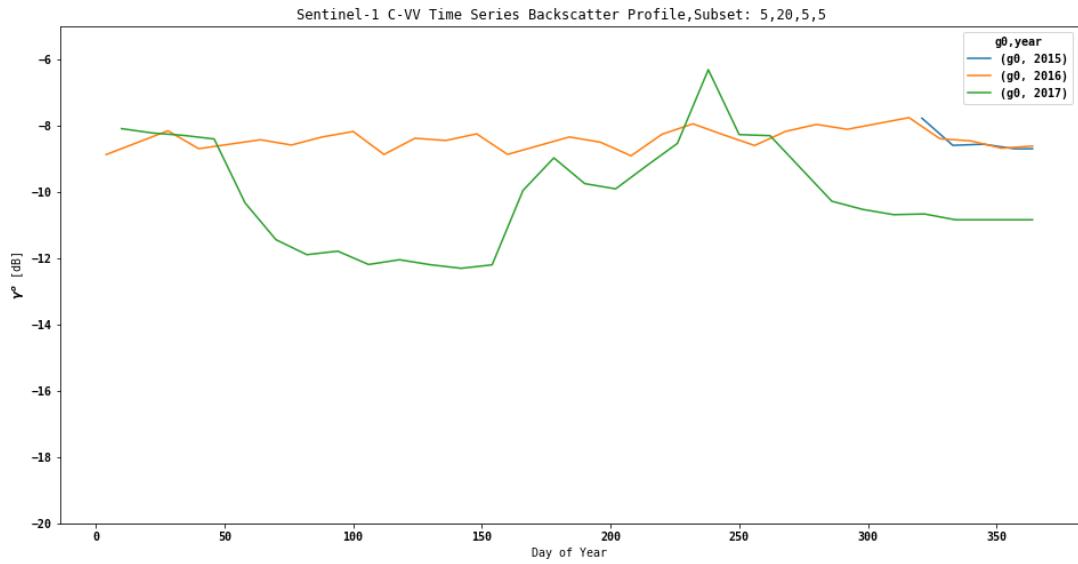
Let's inspect the new pivot table and see whether we interpolated the NaN values where it made sense:

```
In [26]: print(piv2.head(10))
print('...\\n',piv2.tail(10))

g0      g0
year  2015      2016      2017
doy
4      NaN -8.874602      NaN
10     NaN -8.694852 -8.091206
22     NaN -8.335351 -8.222770
28     NaN -8.155600 -8.258453
34     NaN -8.425676 -8.294136
40     NaN -8.695752 -8.348448
46     NaN -8.628392 -8.402759
58     NaN -8.493672 -10.330054
64     NaN -8.426312 -10.885637
70     NaN -8.506059 -11.441220
...
g0      g0
year  2015      2016      2017
doy
321    -7.774510 -8.023862 -10.667306
322    -7.842880 -8.076901 -10.665520
328    -8.253101 -8.395135 -10.753058
333    -8.594952 -8.422687 -10.826007
334    -8.592069 -8.428197 -10.840596
340    -8.574769 -8.461259 -10.840596
345    -8.560352 -8.553227 -10.840596
352    -8.641225 -8.681982 -10.840596
357    -8.698992 -8.654455 -10.840596
364    -8.698992 -8.615916 -10.840596
```

Now we can plot the time series data with overlapping years

```
In [27]: piv2.plot(figsize=(16,8))
plt.title('Sentinel-1 C-VV Time Series Backscatter Profile,\nSubset: 5,20,5,5 ')
plt.ylabel('$\gamma^o$ [dB]')
plt.xlabel('Day of Year')
_=plt.ylim(ylim)
```



# Change Detection on the Time Series Data

We can now analyze the time series for change. We will discuss two approaches:

1. Year-to-year differencing of the subsetted time series
2. Cumulative Sum based change detection

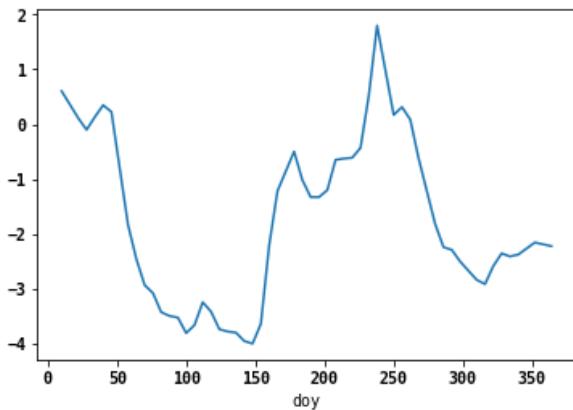
```
In [28]: # Difference between years
# Set a dB change threshold
thres=3
```

```
In [29]: diff1716 = (piv2.g0[2017]-piv2.g0[2016])
```

## Year-to-Year Change Detection

We compute the differences between the interpolated time series and look for change with a threshold value.

```
In [30]:_=diff1716.plot('line')
```



```
In [31]: thres_exceeded = diff1716[abs(diff1716) > thres]
thres_exceeded
```

```
Out[31]: doy
76    -3.081974
82    -3.430615
88    -3.499177
94    -3.527523
100   -3.810760
106   -3.665967
112   -3.248034
118   -3.420952
124   -3.741039
130   -3.782975
136   -3.803250
142   -3.956439
148   -4.003891
154   -3.642117
dtype: float64
```

From the three\_exceeded dataframe we can infer the first date at which the threshold was exceeded. We would label that as a change point. As an additional criteria for labeling a change point, one can also consider the number of observations after identification of a change point where backscatter differed from the year before. If only one or two observations differed from the year before this could be considered an outlier. Additionally, one can introduce smoothing operations with the interpolation

### EXERCISE:

Work through the workbook again with selection of a different point and determine if it is a change point.

## Cumulative Sums for Change Detection

Another approach to detect change in regularly acquired data is employing cumulative sums. Changes are determined against mean observations of time series. A full explanation and examples from the financial sector can be found at <http://www.variation.com/cpa/tech/changepoint.html>

### Time Series and Means

First let's consider a time series and its mean observation. We look at two full years of observations from Sentinel-1 data for an area where we suspect change. In the following we consider  $X$  as a time series

$$X = (X_1, X_2, \dots, X_n)$$

with

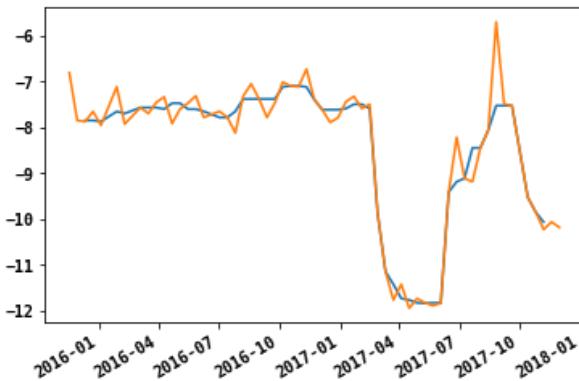
- $X_i$  SAR backscatter at time  $i = 1, \dots, n$
- $n$  number of observations in the time series

```
In [32]: subset=(5,20,3,3)
#subset=(12,5,3,3)
ts1 = timeSeries(rasterstack_pwr,tindex,subset)
X = ts1[ts1.index>'2015-10-31']
```

## Filtering the time series for outliers

It is advantageous in noisy SAR time series data like C-Band data to filter on the time axis. Pandas offers a "rolling" function for these purposes. With that function we can choose, for example, a median filter along the time axis. Below is an example of a median filter for an observation filters the time series when the observation before and after a time stamps are part of the filter.

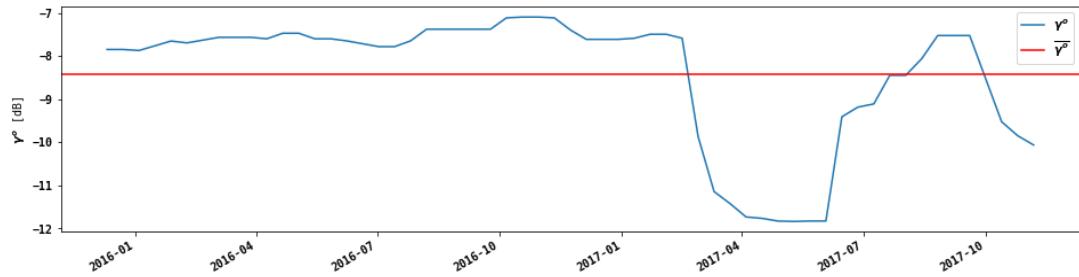
```
In [33]: Xr=X.rolling(5,center=True).median()
Xr.plot()
_=X.plot()
```



Let's plot the time series and it's mean over the time span

```
In [34]: X=Xr # Uncomment if rolling mean is wanted for further computation  
Xmean = X.mean()
```

```
In [35]: fig,ax=plt.subplots(figsize=(16,4))  
X.plot()  
plt.ylabel('$\gamma^o$ [dB]')  
ax.axhline(Xmean,color='red')  
_=plt.legend(['$\gamma^o$', '$\overline{\gamma}^o$'])
```



Let's determine the residuals of the time series against the mean

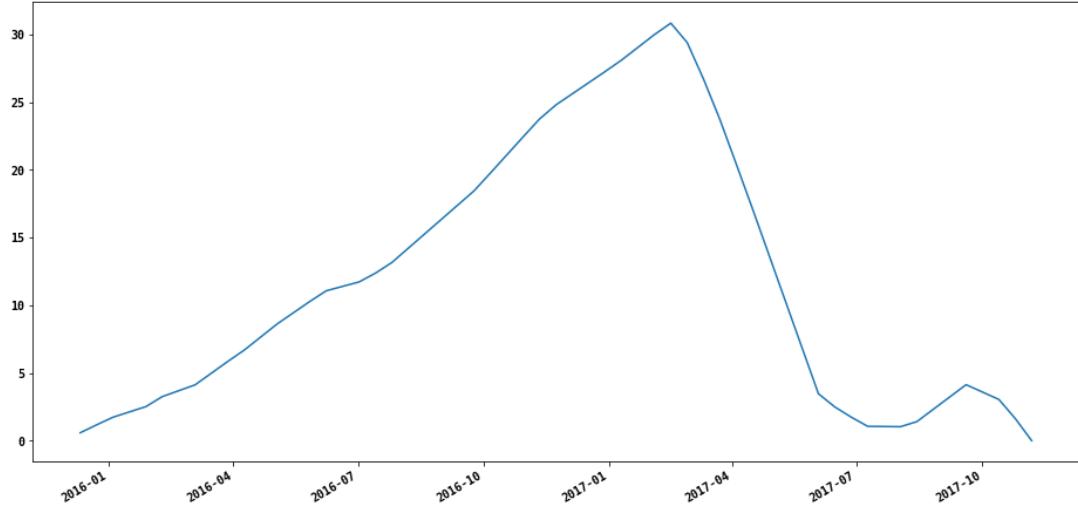
$$R = X_i - \bar{X}$$

```
In [36]: R = X - Xmean
```

Now we compute the cumulative sum of the residuals and plot it:

$$S = \sum_1^n R_i$$

```
In [37]: S = R.cumsum()
         _=S.plot(figsize=(16,8))
```



An estimator for the magnitude of change is given as the difference between the maximum and minimum value of  $S$

$$S_{DIFF} = S_{MAX} - S_{MIN}$$

```
In [38]: Sdiff=S.max() - S.min()
         Sdiff
```

```
Out[38]: 30.847062820803558
```

A candidate change point is identified from the  $S$  curve at the time where  $S_{MAX}$  is found:

$$T_{CP_{before}} = T(S_i = S_{MAX})$$

with

- $T_{CP_{before}}$  Timestamp of last observation before change
- $S_i$  Cumulative Sum of  $R$  with  $i = 1, \dots, n$
- $n$  Number of observations in the time series

The first observation after change occurred ( $T_{CP_{after}}$ ) is then found as the first observation in the time series following  $T_{CP_{before}}$ .

For our example time series X these points are:

```
In [39]: t_cp_before = S[S==S.max()].index[0]
         print('Last date before change: {}'.format(t_cp_before.date()))
```

```
Last date before change: 2017-02-15
```

```
In [40]: t_cp_after = S[S.index > t_cp_before].index[0]
         print('First date after change: {}'.format(t_cp_after.date()))
```

```
First date after change: 2017-02-27
```

## Bootstrapping the cumulative sums by randomly reordering the time series

We can determine if an identified change point is indeed a valid detection by randomly reordering the time series and comparing the various S curves. During bootstrapping we count how many times the  $S_{DIFF}$  values are greater than  $S_{DIFF_{random}}$  of the identified change point. A confidence level  $CL$  is computed as:

$$CL = \frac{N_{GT}}{N_{bootstraps}}$$

with

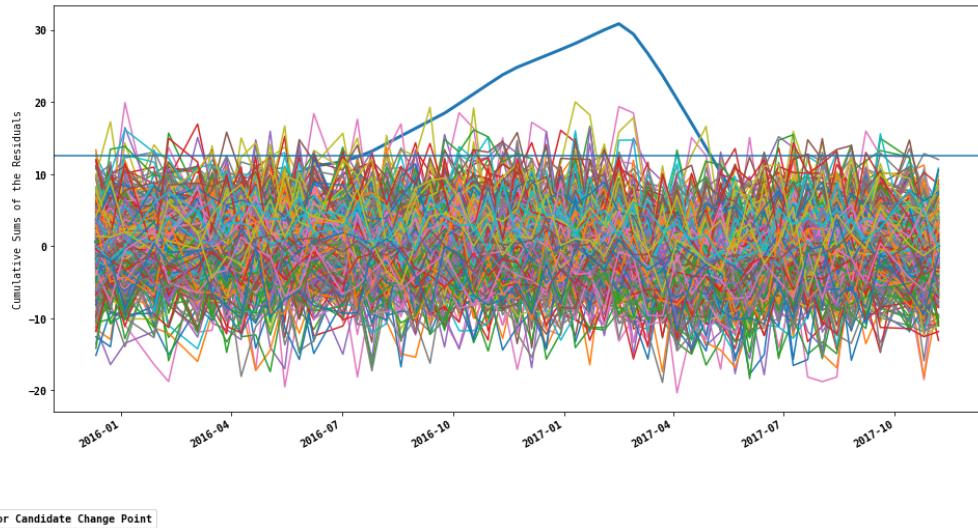
- $N_{GT}$  Number of times  $S_{DIFF} > S_{DIFF_{random}}$
- $N_{bootstraps}$  Number of bootstraps randomizing  $R$

Another metric for the significance of a change point is 1 minus the ratio of the mean of the  $S_{DIFF_{random}}$  values and  $S_{DIFF}$ . The closer this value is approaching 1, the more significant the change point:

$$CP_{significance} = 1 - \left( \frac{\sum_{b=1}^{N_{bootstraps}} S_{DIFF_{random,b}}}{N_{bootstraps}} \right) / S_{DIFF}$$

The python code to conduct the boot strapping, including visualization of the S curves is below:

```
In [41]: n_bootstraps=500 # bootstrap sample size
fig,ax = plt.subplots(figsize=(16,8))
S.plot(ax=ax,linewidth=3)
ax.set_ylabel('Cumulative Sums of the Residuals')
fig.legend(['S Curve for Candidate Change Point'],loc=3)
Sdiff_random_sum=0
Sdiff_random_max=0 # to keep track of the maximum Sdiff of the
# bootstrapped sample
n_Sdiff_gt_Sdiff_random=0 # to keep track of the maximum Sdiff of the
# bootstrapped sample
for i in range(n_bootstraps):
    Rrandom = R.sample(frac=1) # Randomize the time steps of the residuals
    Srandom = Rrandom.cumsum()
    Sdiff_random=Srandom.max()-Srandom.min()
    Sdiff_random_sum += Sdiff_random
    if Sdiff_random > Sdiff_random_max:
        Sdiff_random_max = Sdiff_random
    if Sdiff > Sdiff_random:
        n_Sdiff_gt_Sdiff_random += 1
    Srandom.plot(ax=ax)
_=ax.axhline(Sdiff_random_sum/n_bootstraps)
```



```
In [42]: CL = 1.*n_Sdiff_gt_Sdiff_random/n_bootstraps
print('Confidence Level for change point {} percent'.format(CL*100.))
```

Confidence Level for change point 100.0 percent

```
In [43]: CP_significance = 1. - (Sdiff_random_sum/n_bootstraps)/Sdiff
print('Change point significance metric: {}'.format(CP_significance))
```

Change point significance metric: 0.5910152452854309

Another useful metric to determine strength of a change point is the normalized integral  $S_{ni}$  of the absolute values of the S curve:

$$S_{normintegral} = \frac{\int_{i=1}^n \frac{abs(S_i)}{\max abs(S)}}{n}$$

```
In [44]: # NaN's to be excluded in the computation
S_ni=(S.abs()/S.abs().max()).cumsum().max()/len(S[S != np.nan])
print('Normalized Integral of cumulative sum: {}'.format(S_ni))
```

Normalized Integral of cumulative sum: 0.3741739515908098

## EXERCISE

Conduct the change point analysis for different subsets in the traning data

### Selection of threshold values

$CL$  and  $CP_{significance}$  can be used as threshold values for the acceptance or rejection of a candidate threshold. These values are to some degree specific to a SAR sensor and environmental conditions. E.g. L-Band SAR has a more pronounced decrease in backscatter after forest disturbance and logging, whereas C-Band can have more ambiguous signals. Also moisture regime changes, e.g. with snow cover, freeze/thaw conditions or dry/wet season changes have an influence on the time series signal. For example El Nino years can suggest changes solely due to different wetting and dryup conditions pertinent to a particular year. For this reason other techniques can be added to the SAR time series analysis. Two techniques can readily be thought of:

- Subsetting of time series by seasons
- Detrending time series with global image means

If year-to-year comparison is the focus, the first approach likely leads to subsets that are too small for meaningful cumulative sum change point detection. The approach of interannual differencing as discussed above likely performs better.

In the following we explore the approach to detrend the data with global image means.

### De-trending time series with global image means

The idea of de-trending time series with global image means should prepare time series for a somewhat more robust change point detection as global image time series anomalies stemming calibration or seasonal trends are removed prior to time series analysis. This de-trending needs to be performed with large subsets so real change is not influencing the image statistics.

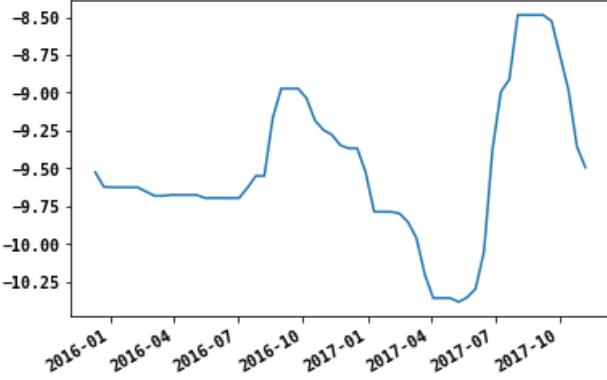
NOTE: For our small subset, we will see some of these effects.

Let's start by building a global image means time series:

```
In [45]: means_pwr = np.mean(rasterstack_pwr, axis=(1,2))
means_db = 10.*np.log10(means_pwr)
gm_ts = pd.Series(means_db, index=tindex)
gm_ts=gm_ts[gm_ts.index > '2015-10-31'] # filter dates
gm_ts=gm_ts.rolling(5,center=True).median()
```

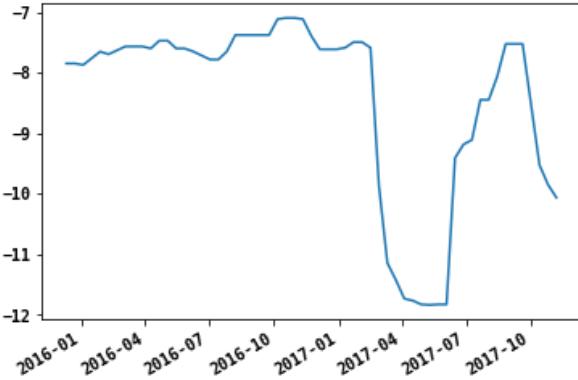
```
In [46]: gm_ts.plot()
```

```
Out[46]: <matplotlib.axes._subplots.AxesSubplot at 0x128525518>
```



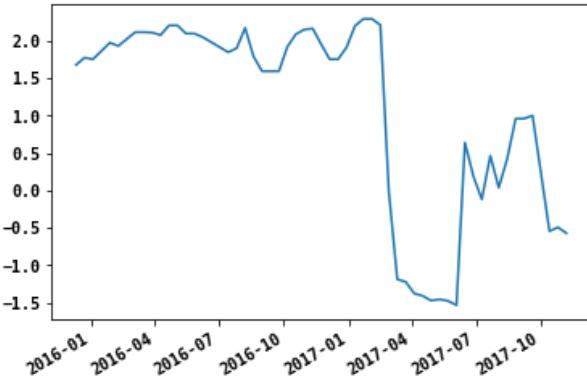
```
In [47]: X.plot()
```

```
Out[47]: <matplotlib.axes._subplots.AxesSubplot at 0x12854aa90>
```



```
In [48]: Xd=X-gm_ts  
Xmean=Xd.mean()  
Xd.plot()
```

```
Out[48]: <matplotlib.axes._subplots.AxesSubplot at 0x128576278>
```

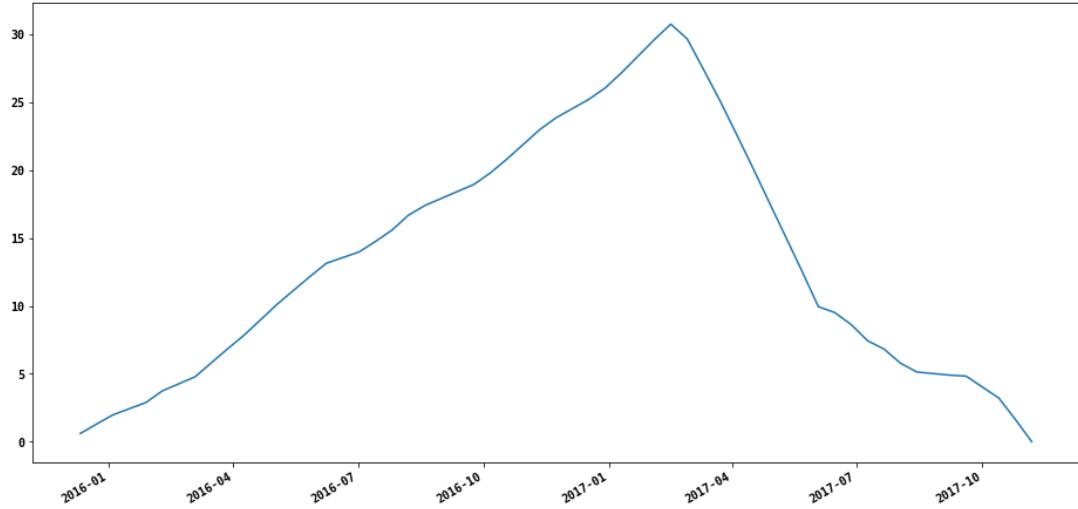


```
In [49]: R = Xd - Xmean
```

Now we compute the cumulative sum of the residuals and plot it:

$$S = \sum_1^n R_i$$

```
In [50]: S = R.cumsum()  
_ = S.plot(figsize=(16,8))
```



An estimator for the magnitude of change is given as the difference between the maximum and minimum value of S

$$S_{DIFF} = S_{MAX} - S_{MIN}$$

```
In [51]: Sdiff=S.max() - S.min()  
Sdiff
```

```
Out[51]: 30.765903052062484
```

A candidate change point is identified from the  $S$  curve at the time where  $S_{MAX}$  is found:

$$T_{CP_{before}} = T(S_i = S_{MAX})$$

with

- $T_{CP_{before}}$  Timestamp of last observation before change
- $S_i$  Cumulative Sum of R with  $i = 1, \dots, n$
- $n$  Number of observations in the time series

The first observation after change occurred ( $T_{CP_{after}}$ ) is then found as the first observation in the time series following  $T_{CP_{before}}$ .

For our example time series X these points are:

```
In [52]: t_cp_before = S[S==S.max()].index[0]
print('Last date before change: {}'.format(t_cp_before.date()))
```

Last date before change: 2017-02-15

```
In [53]: t_cp_after = S[S.index > t_cp_before].index[0]
print('First date after change: {}'.format(t_cp_after.date()))
```

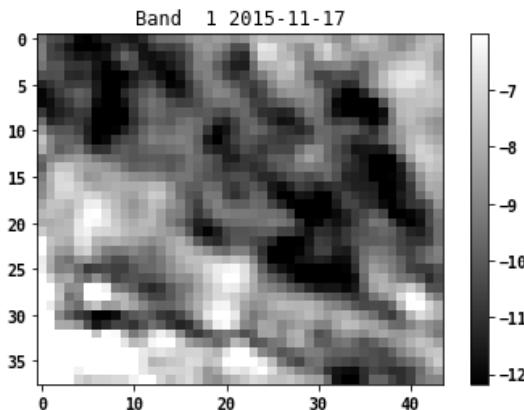
First date after change: 2017-02-27

## Cumulative Sum Change Detection for the entire image

With numpy arrays we can apply the concept of cumulative sum change detection analysis effectively on the entire image stack. We take advantage of array slicing and axis-based computing in numpy. Axis 0 is the time domain in our raster stacks

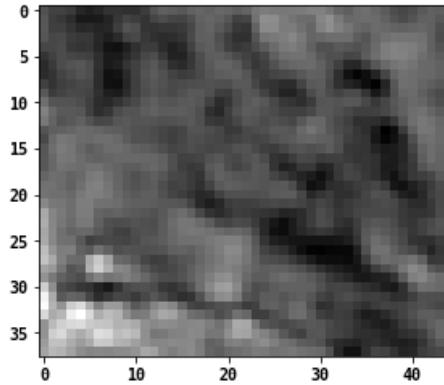
```
In [58]: # Can do this in power or dB scale
X = rasterstack_pwr
# Filter out the first layer ( Dates >= '2015-11-1')
X_sub=X[1:,:,:]
tindex_sub=tindex[1:]
X= 10.*np.log10(X_sub) # Uncomment to test dB scale
```

```
In [59]: plt.figure()
#Indicate the band number
bandnbr=0
vmin=np.percentile(X[bandnbr],5)
vmax=np.percentile(X[bandnbr],95)
plt.title('Band {} {}'.format(bandnbr+1,tindex_sub[bandnbr].date()))
plt.imshow(X[0],cmap='gray',vmin=vmin,vmax=vmax)
_=plt.colorbar()
```



```
In [60]: Xmean=np.mean(X,axis=0)
plt.figure()
plt.imshow(Xmean,cmap='gray')
```

```
Out[60]: <matplotlib.image.AxesImage at 0x128a80908>
```

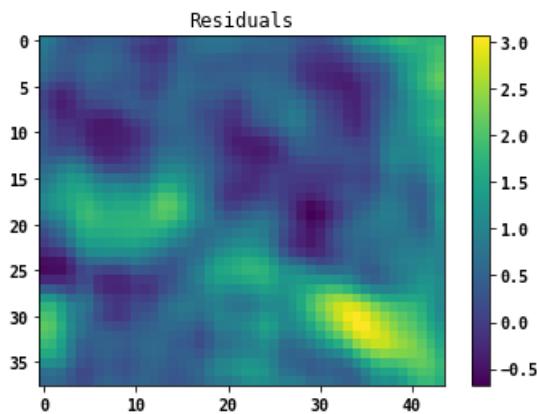


```
In [61]: #  
X.shape
```

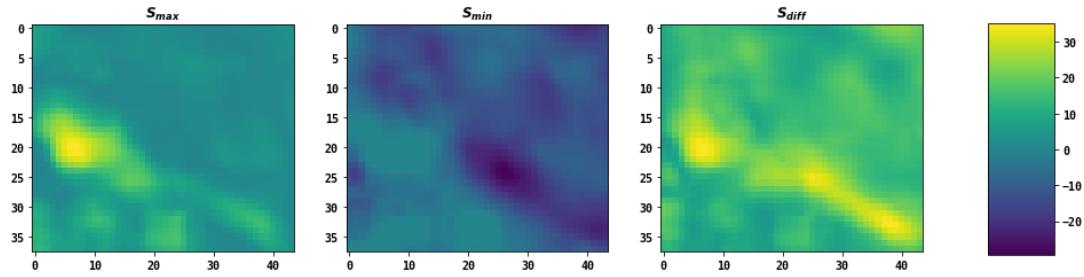
```
Out[61]: (59, 38, 44)
```

```
In [62]: R=X-Xmean
```

```
In [63]: #Create an image that spatially displays the residuals (R)
plt.imshow(R[0])
plt.title('Residuals')
_=plt.colorbar()
```



```
In [64]: S = np.cumsum(R, axis=0)
Smax= np.max(S, axis=0)
Smin= np.min(S, axis=0)
Sdiff=Smax-Smin
fig,ax=plt.subplots(1,3,figsize=(16,4))
vmin=Smin.min()
vmax=Smax.max()
p=ax[0].imshow(Smax,vmin=vmin,vmax=vmax)
ax[0].set_title('Smax')
ax[1].imshow(Smin,vmin=vmin,vmax=vmax)
ax[1].set_title('Smin')
ax[2].imshow(Sdiff,vmin=vmin,vmax=vmax)
ax[2].set_title('Sdiff')
fig.subplots_adjust(right=0.8)
cbar_ax = fig.add_axes([0.85, 0.15, 0.05, 0.7])
cbar_ax = fig.colorbar(p,cax=cbar_ax)
```



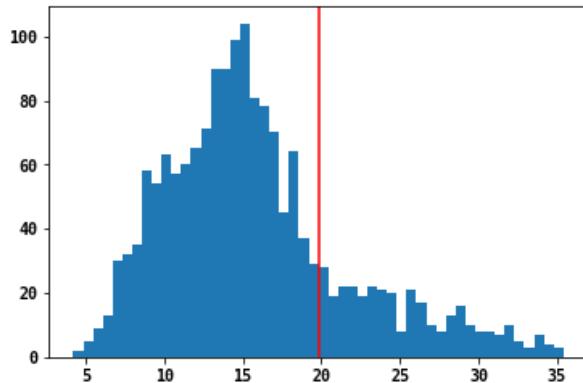
## Mask $S_{\text{diff}}$ with a priori threshold for expected change

If we have an assumption as to how much actual change we expect in the image, we can threshold  $S_{\text{diff}}$  to reduce computation of the bootstrapping. For land cover change we would not expect more than 5-10% change in a landscape. So, if the test region is reasonably large, setting a threshold for expected change to 10% would be appropriate. Thus we can set a mask with the 90th percentile of the histogram of  $S_{\text{diff}}$ . In our example we'll start out with a very conservative threshold of 50%.

The histogram for  $S_{\text{diff}}$  is shown below:

```
In [65]: #Display the Sdiff histogram
precentile=50
fig,ax=plt.subplots()
h=ax.hist(Sdiff.flatten(),bins=50)
thres=np.percentile(h[1],50)
print('At the {}% percentile, the threshold value is {:.2f}'.format(precentile,
thres))
_=ax.axvline(thres,color='red')
```

At the 50% percentile, the threshold value is 19.82

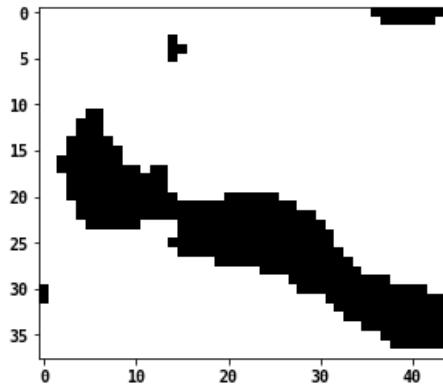


At the 50% percentile, the threshold value is \_\_\_\_ (printed above the histogram)

Using this threshold, we can visualize the candidate changepoints:

```
In [66]: Sdiffmask=Sdiff

```



Now we can filter our Residuals and perform bootstrapping analysis on these data. We make use of numpy masked arrays for this purpose.

```
In [67]: Rmask = np.broadcast_to(Sdiffmask,R.shape)
```

```
In [68]: Rmasked = np.ma.array(R,mask=Rmask)
```

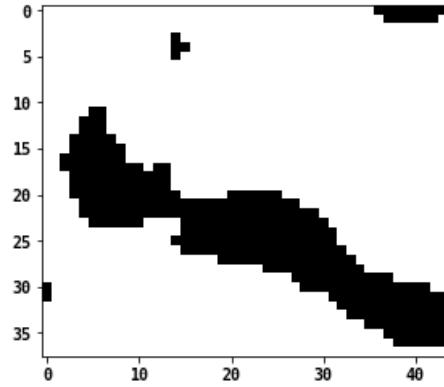
On the masked time series stack of residuals we can compute the cumulative sums:

```
In [69]: Smasked = np.ma.cumsum(Rmasked, axis=0)
```

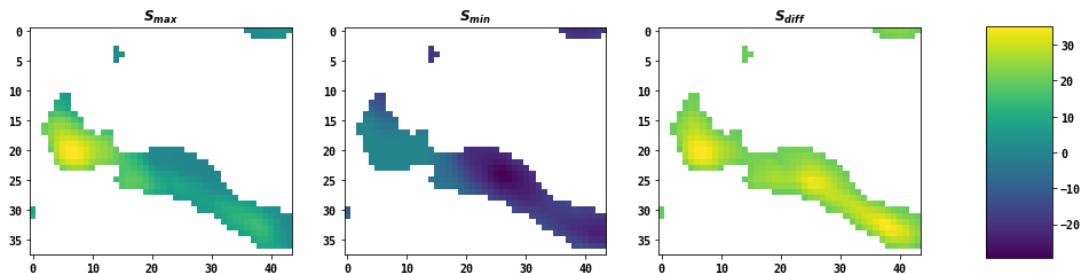
$S_{MAX}$ ,  $S_{MIN}$ ,  $S_{DIFF}$  can also be computed on the masked arrays :

```
In [70]: plt.imshow(Rmasked.mask[0], cmap='gray')
```

```
Out[70]: <matplotlib.image.AxesImage at 0x127cb1390>
```



```
In [71]: Smasked = np.ma.cumsum(Rmasked, axis=0)
Smax = np.ma.max(Smaskd, axis=0)
Smin = np.ma.min(Smaskd, axis=0)
Sdiff = Smax - Smin
fig, ax = plt.subplots(1, 3, figsize=(16, 4))
vmin = Smin.min()
vmax = Smax.max()
p = ax[0].imshow(Smax, vmin=vmin, vmax=vmax)
ax[0].set_title('S_{max}')
ax[1].imshow(Smin, vmin=vmin, vmax=vmax)
ax[1].set_title('S_{min}')
ax[2].imshow(Sdiff, vmin=vmin, vmax=vmax)
ax[2].set_title('S_{diff}')
fig.subplots_adjust(right=0.8)
cbar_ax = fig.add_axes([0.85, 0.15, 0.05, 0.7])
cbar = fig.colorbar(p, cax=cbar_ax)
```

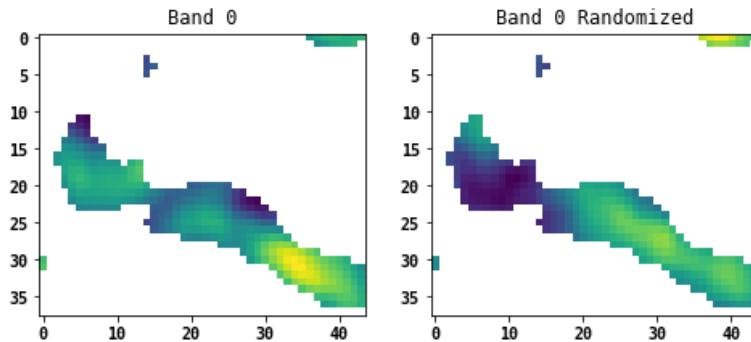


## Bootstrapping over the masked change point candidates

We can now perform the bootstrapping analysis over the not masked out values. For efficient computing we permute the index of the time axis.

```
In [72]: random_index=np.random.permutation(Rmasked.shape[0])
Rrandom=Rmasked[random_index,:,:,:]

fig,ax=plt.subplots(1,2,figsize=(8,4))
ax[0].imshow(Rmasked[0])
ax[0].set_title('Band 0')
ax[1].imshow(Rrandom[0])
_=ax[1].set_title('Band 0 Randomized')
```



```
In [73]: Smasked_max=np.ma.max(Smasked, axis=0)
```

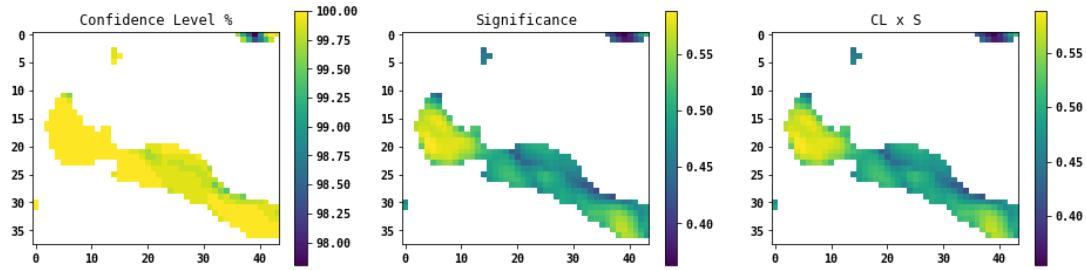
Below is the numpy based implementation of the bootstrapping over all pixels. Note the efficient implementation using numpy masked arrays.

```
In [74]: n_bootstraps=1000 # bootstrap sample size

# to keep track of the maximum Sdiff of the bootstrapped sample:
Sdiff_random_max = np.ma.copy(Smasked_diff)
Sdiff_random_max[-Sdiff_random_max.mask]=0
# to compute the Sdiff sums of the bootstrapped sample:
Sdiff_random_sum = np.ma.copy(Smasked_diff)
Sdiff_random_sum[-Sdiff_random_max.mask]=0
# to keep track of the count of the bootstrapped sample
n_Sdiff_gt_Sdiff_random = np.ma.copy(Smasked_diff)
n_Sdiff_gt_Sdiff_random[~n_Sdiff_gt_Sdiff_random.mask]=0
for i in range(n_bootstraps):
    # For efficiency, we shuffle the time axis index and use that
    # to randomize the masked array
    random_index=np.random.permutation(Rmasked.shape[0])
    # Randomize the time step of the residuals
    Random = Rmasked[random_index,:,:,:]
    Random = np.ma.cumsum(Random, axis=0)
    Random_max=np.ma.max(Random, axis=0)
    Random_min=np.ma.min(Random, axis=0)
    Sdiff_random=Random_max-Random_min
    Sdiff_random_sum += Sdiff_random
    Sdiff_random_max[np.ma.greater(Sdiff_random,Sdiff_random_max)]=\
    Sdiff_random[np.ma.greater(Sdiff_random,Sdiff_random_max)]
    n_Sdiff_gt_Sdiff_random[np.ma.greater(Smasked_diff,Sdiff_random)] += 1
```

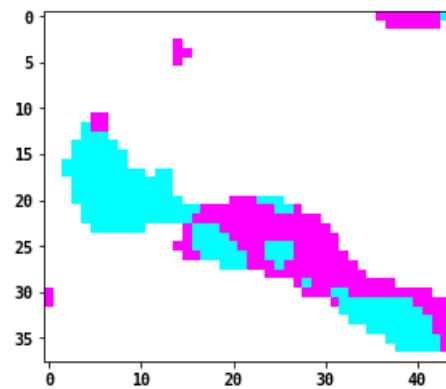
Now we can compute for all pixels the confidence level  $CL$ , the change point significance metric  $CP_{\{significance\}}$  and the product of the two as our confidence metric for identified changepoints.

```
In [75]: CL = n_Sdiff_gt_Sdiff_random/n_bootstraps
CP_significance = 1. - (Sdiff_random_sum/n_bootstraps)/Sdiff
#Plot
fig,ax=plt.subplots(1,3,figsize=(16,4))
a = ax[0].imshow(CL*100)
fig.colorbar(a,ax=ax[0])
ax[0].set_title('Confidence Level %')
a = ax[1].imshow(CP_significance)
fig.colorbar(a,ax=ax[1])
ax[1].set_title('Significance')
a = ax[2].imshow(CL*CP_significance)
fig.colorbar(a,ax=ax[2])
ax[2].set_title('CL x S')
```



Now if we were to set a threshold of 0.5 for the product as identified change map would look like the following figure:

```
In [76]: cp_thres=0.5
In [77]: plt.imshow(CL*CP_significance < cp_thres,cmap='cool')
Out[77]: <matplotlib.image.AxesImage at 0x126bf49e8>
```



Our last step is the identification of the change points is to extract the timing of the change. We will produce a raster layer that shows the band number of this first date after detected change. We will make use of the numpy indexing scheme. First, we create a combined mask of the first threshold and the identified change points after the bootstrapping. For this we use the numpy "mask\_or" operation.

```
In [78]: # make a mask of our change points from the new threshold and the previous mask
cp_mask=np.ma.mask_or(CL*CP_significance<cp_thres,CL.mask)
# Broadcast the mask to the shape of the masked S curves
cp_mask2 = np.broadcast_to(cp_mask,Smasked.shape)
# Make a numpy masked array with this mask
CPraster = np.ma.array(Smasked.data,mask=cp_mask2)
```

To retrieve the dates of the change points we find the band indices in the time series along the time axis where the maximum of the cumulative sums was located. Numpy offers the "argmax" function for this purpose.

```
In [79]: CP_index= np.ma.argmax(CPraster, axis=0)
change_indices = list(np.unique(CP_index))
change_indices.remove(0)
print(change_indices)
# Look up the dates from the indices to get the change dates
alldates=tindex[tindex>'2015-10-31']
change_dates=[str(alldates[x+1].date()) for x in change_indices]
print(change_dates)

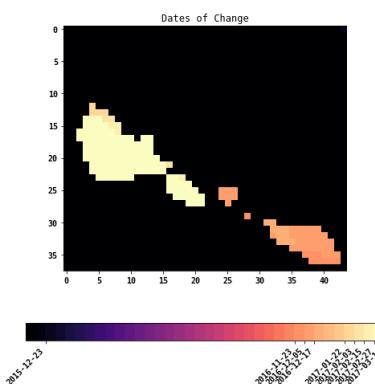
[2, 27, 28, 29, 32, 33, 34, 35, 36]
['2015-12-23', '2016-11-23', '2016-12-05', '2016-12-17', '2017-01-22', '2017-02-03', '2017-02-15', '2017-02-27', '2017-03-11']
```

Lastly, we visualize the change dates by showing the CP\_index raster and label the change dates.

```
In [80]: ticks=change_indices
ticklabels=change_dates

cmap=plt.cm.get_cmap('magma',ticks[-1])
fig, ax = plt.subplots(figsize=(8,8))
cax = ax.imshow(CP_index,interpolation='nearest',cmap=cmap)
# fig.subplots_adjust(right=0.8)
# cbar_ax = fig.add_axes([0.85, 0.15, 0.05, 0.7])
# fig.colorbar(p,cax=cbar_ax)

ax.set_title('Dates of Change')
# cbar = fig.colorbar(cax,ticks=ticks)
cbar=fig.colorbar(cax,ticks=ticks,orientation='horizontal')
_cbar.ax.set_xticklabels(ticklabels, size=10, rotation=45, ha='right')
```



## Secondary Change Points

After detection of a change point in the time series we can split the series in before and after change subsets. For forest degradation or deforestation detection for example this could apply when over the course of a multi-year time series selective logging precedes a clearing event or conversion of a logged plot to agriculture or regrowth, which show typically different time series profiles of radar backscatter. The approach to detect secondary change points would be to repeat analysis of the time series split into before and after change point detection.

## Conclusion

Pandas and numpy are powerful open source scripting tools to implement change point detection on large data stacks. For image based analysis numpy offers more efficient implementations compared to pandas, whereas pandas is more powerful in date time processing, e.g. time-weighted interpolation.

## Solutions

```
In [ ]: # 1.  
ts[ts<-11].index  
  
In [ ]: # 2.  
gradient_lag1 = ts.diff(1)  
gradient_lag1.plot()  
  
In [ ]: # 3.  
gradient_lag1.min()  
  
In [ ]: # 4.  
gradient_lag1[gradient_lag1==gradient_lag1.min()]  
  
In [ ]: before = gradient_lag1[gradient_lag1==gradient_lag1.min()].index[0]  
before  
  
In [ ]: after=tindex[tindex>before][0]  
after
```

# SAR Training Workshop for Forest Applications

## PART 5 - SAR/Optical (NDVI) Time Series Analysis

Josef Kellndorfer, Ph.D., President and Senior Scientist, Earth Big Data, LLC

Revision date: January 2019

In this chapter we compare time series data of C-band Backscatter and Landsat 8 Normalized Difference Vegetation Index (NDVI) over a forested site in Southern Niger.

```
In [1]: # Importing relevant python packages
import pandas as pd
import gdal
import numpy as np
import time,os
from skimage import exposure # to enhance image display

# For plotting
%matplotlib inline
import matplotlib.pylab as plt
import matplotlib.patches as patches
import matplotlib.cm as cm

font = {'family' : 'monospace',
        'weight' : 'bold',
        'size'   : 18}
plt.rc('font',**font)

# Define a helper function for a 4 part figure with backscatter, NDVI and False
Color Infrared
def ebd_plot(bandnbrs):
    fig,ax=plt.subplots(2,2,figsize=(16,16))
    # Bands for sentinel and landsat:
    # Sentinel VV
    sentinel_vv=img_handle[0].GetRasterBand(bandnbrs[0]).ReadAsArray(*subset_sen
tinel)
    sentinel_vv=20.*np.log10(sentinel_vv)-83 # Covert to dB
    # Sentinel VH
    sentinel_vh=img_handle[1].GetRasterBand(bandnbrs[0]).ReadAsArray(*subset_sen
tinel)
    sentinel_vh=20.*np.log10(sentinel_vh)-83 # Covert to dB
    # # Landsat False Color InfraRed
    r=img_handle[5].GetRasterBand(bandnbrs[1]).ReadAsArray(*subset_landsat)/1000
0.
    g=img_handle[4].GetRasterBand(bandnbrs[1]).ReadAsArray(*subset_landsat)/1000
0.
    b=img_handle[3].GetRasterBand(bandnbrs[1]).ReadAsArray(*subset_landsat)/1000
0.
    fcir=np.dstack((r,g,b))
    for i in range(fcir.shape[2]):
        fcir[:, :, i] = exposure.\_
            equalize_hist(fcir[:, :, i],
            mask=~np.equal(fcir[:, :, i],-.9999))
```

```

# Landsat NDVI
landsat_ndvi=img_handle[2].GetRasterBand(bandnbrs[1]).ReadAsArray(*subset_landsat)
mask=landsat_ndvi==9999
landsat_ndvi = landsat_ndvi/10000. # Scale to real NDVI value
landsat_ndvi[mask]=np.nan
svv = ax[0][0].imshow(sentinel_vv,cmap='jet',vmin=np.nanpercentile(sentinel_vv,5),
                      vmax=np.nanpercentile(sentinel_vv,95))
cb = fig.colorbar(svv,ax=ax[0][0],orientation='horizontal')
cb.ax.set_title('C-VV $\gamma^{\circ}$ [dB]')
svh = ax[0][1].imshow(sentinel_vh,cmap='jet',vmin=np.nanpercentile(sentinel_vh,5),
                      vmax=np.nanpercentile(sentinel_vh,95))
cb = fig.colorbar(svh,ax=ax[0][1],orientation='horizontal')
cb.ax.set_title('C-VH $\gamma^{\circ}$ [dB]')

nvmin=np.nanpercentile(landsat_ndvi,5)
nvmax=np.nanpercentile(landsat_ndvi,95)

# nvmin=-1
# nvmax=1
nax = ax[1][0].imshow(landsat_ndvi,cmap='jet',vmin=nvmin,
                      vmax=nvmax)
cb = fig.colorbar(nax,ax=ax[1][0],orientation='horizontal')
cb.ax.set_title('NDVI')

fc= ax[1][1].imshow(fcir)
# cb = fig.colorbar(fc,cmap=cm.gray,ax=ax[1][1],orientation='horizontal')
# cb.ax.set_title('False Color Infrared')

ax[0][0].axis('off')
ax[0][1].axis('off')
ax[1][0].axis('off')
ax[1][1].axis('off')
ax[0][0].set_title('Sentinel-1 C-VV {}'.format(stindex[bandnbrs[0]-1].date()))
ax[0][1].set_title('Sentinel-1 C-VH {}'.format(stindex[bandnbrs[0]-1].date()))
ax[1][0].set_title('Landsat-8 NDVI {}'.format(ltindex[bandnbrs[1]-1].date()))
ax[1][1].set_title('Landsat-8 False Color IR {}'.format(ltindex[bandnbrs[1]-1].date()))
fig.suptitle('Sentinel-1 Backscatter and Landsat NDVI and FC IR',size=16)

```

# Set Project Directory and Filenames

## West Africa - Biomass Site

```
In [2]: datadirectory='C:\\\\Users\\\\loaner.SERVIRLOAN-5057.001\\\\Downloads\\\\BIOss1'  
#datadirectory='/dev/shm/projects/c401servir/wa/BIOss1'  
sentinel1_datefile='S32631X398020Y1315440sS1_A_vv_0001_mtfil.dates'  
sentinel1_imagefile='S32631X398020Y1315440sS1_A_vv_0001_mtfil.vrt'  
sentinel1_imagefile_cross='S32631X398020Y1315440sS1_A_vh_0001_mtfil.vrt'  
landsat8_ndvi='landsat/L8_192_052_NDVI.vrt'  
landsat8_b3='landsat/L8_192_052_B3.vrt'  
landsat8_b4='landsat/L8_192_052_B4.vrt'  
landsat8_b5='landsat/L8_192_052_B5.vrt'  
landsat8_datefile='landsat/L8_192_052_NDVI.dates'
```

```
In [3]: # Switch to the data directory  
os.chdir(os.path.join(datadirectory))
```

## Acquisition Dates

Read from the Dates file the dates in the time series and make a pandas date index

```
In [4]: sdates=open(sentinel1_datefile).readlines()  
stindex=pd.DatetimeIndex(sdates)  
j=1  
print('Bands and dates for',sentinel1_imagefile)  
for i in stindex:  
    print("{:4d} {}".format(j, i.date()),end=' ')  
    j+=1  
    if j%5==1: print()
```

Bands and dates for S32631X398020Y1315440sS1_A_vv_0001_mtfil.vrt					
1	2015-03-22	2	2015-04-03	3	2015-04-15
6	2015-06-02	7	2015-06-14	8	2015-06-26
11	2015-08-01	12	2015-08-13	13	2015-08-25
16	2015-09-30	17	2015-10-12	18	2015-10-24
21	2015-12-11	22	2015-12-23	23	2016-01-04
26	2016-03-04	27	2016-03-16	28	2016-03-28
31	2016-05-03	32	2016-05-15	33	2016-05-27
36	2016-07-14	37	2016-07-26	38	2016-08-07
41	2016-09-12	42	2016-09-24	43	2016-10-06
46	2016-11-11	47	2016-11-23	48	2016-12-05
51	2017-01-10	52	2017-01-22	53	2017-02-03
56	2017-03-11	57	2017-03-23	58	2017-04-04
61	2017-05-10	62	2017-05-22	63	2017-06-03
66	2017-07-09	67	2017-07-21	68	2017-08-02
71	2017-09-07	72	2017-09-19	73	2017-10-13
76	2017-11-18	77	2017-11-30	74	2017-10-25
					75 2017-11-06

```
In [5]: ldates=open(landsat8_datefile).readlines()
ltindex=pd.DatetimeIndex(ldates)
j=1
print('Bands and dates for',landsat8_ndvi)
for i in ltindex:
    print("{:4d} {}".format(j, i.date()),end=' ')
    j+=1
    if j%5==1: print()
```

Bands and dates for landsat/L8_192_052_NDVI.vrt				
1	2015-01-13	2	2015-01-29	3 2015-02-14
6	2015-04-19	7	2015-05-05	8 2015-05-21
11	2015-07-08	12	2015-07-24	13 2015-08-09
16	2015-09-26	17	2015-10-12	18 2015-10-28
21	2015-12-15	22	2015-12-31	23 2016-01-16
26	2016-03-04	27	2016-03-20	28 2016-04-05
31	2016-05-23	32	2016-06-08	33 2016-06-24
36	2016-08-11	37	2016-08-27	38 2016-09-12
41	2016-10-30	42	2016-11-15	43 2016-12-01
46	2017-01-18	47	2017-02-03	48 2017-02-19
51	2017-04-08	52	2017-04-24	49 2017-03-07
				50 2017-03-23

## Projection and Georeferencing Information of the SAR and Optical Time Series Data Stacks

For processing of the imagery in this notebook we generate a list of image handles and retrieve projection and georeferencing information. We print out the retrieved information.

```
In [6]: imagelist=[sentinel1_imagefile,sentinel1_imagefile_cross,landsat8_ndvi,landsat8_b3,landsat8_b4,landsat8_b5]
geotrans=[]
proj=[]
img_handle=[]
xsize=[]
ysize=[]
bands=[]
for i in imagelist:
    img_handle.append(gdal.Open(i))
    geotrans.append(img_handle[-1].GetGeoTransform())
    proj.append(img_handle[-1].GetProjection())
    xsize.append(img_handle[-1].RasterXSize)
    ysize.append(img_handle[-1].RasterYSize)
    bands.append(img_handle[-1].RasterCount)
# for i in proj:
#     print(i)
# for i in geotrans:
#     print(i)
# for i in zip(['C-VV','C-VH','NDVI','B3','B4','B5'],bands,ysize,xsize):
#     print(i)
```

## Display SAR and NDVI Images

First, depending on the capacity of the computer we might want to define a subset. We will choose the subset in the raster extension of the Sentinel-1 Image and use the geotransformation information to extract the corresponding subset in the Landsat Image. We assume that the images have the same upper left coordinate. Then we can compute the offsets and extent in the Landsat image as follows:

$$x_{cal} = \frac{xres_{sentinel-1}}{xres_{landsat}}$$

$$y_{cal} = \frac{yres_{sentinel-1}}{yres_{landsat}}$$

We can use these calibration factors to get the landsat subset as follows:

- $xoff_{landsat} = xoff_{sentinel-1} \times x_{cal}$
- $yoff_{landsat} = yoff_{sentinel-1} \times y_{cal}$
- $xsize_{landsat} = xsize_{sentinel-1} \times x_{cal}$
- $ysize_{landsat} = ysize_{sentinel-1} \times y_{cal}$

(xoffset,yoffset,xsize,ysize)

```
In [7]: subset_sentinel=None
subset_sentinel=(570,40,500,500) # Adjust or comment out if you don't want a subset
if subset_sentinel == None:
    subset_sentinel=(0,0,img_handle[0].RasterXSize,img_handle[0].RasterYSize)
    subset_landsat=(0,0,img_handle[2].RasterXSize,img_handle[2].RasterYSize)
else:
    xoff,yoff,xsize,ysize=subset_sentinel
    xcal=geotrans[0][1]/geotrans[2][1]
    ycal=geotrans[0][5]/geotrans[2][5]
    subset_landsat=(int(xoff*xcal),int(yoff*ycal),int(xsize*xcal),int(ysize*ycal))
)
print('Subset Sentinel-1',subset_sentinel,'Subset Landsat ',subset_landsat)

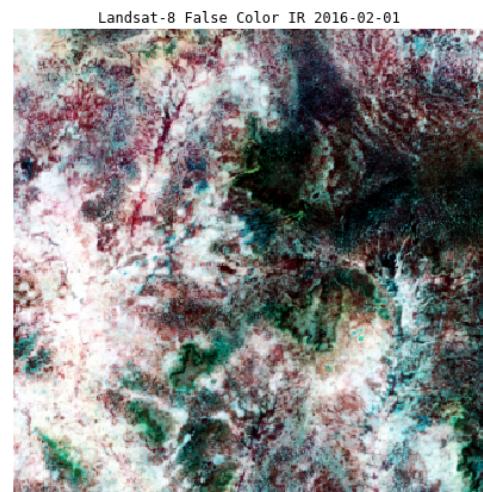
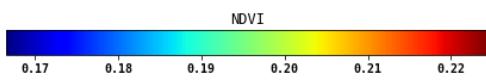
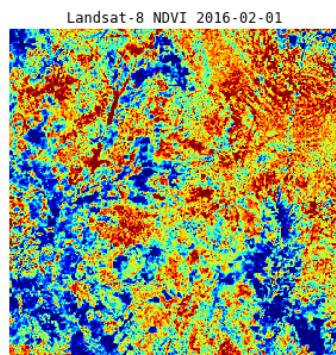
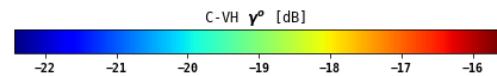
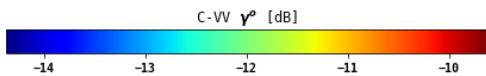
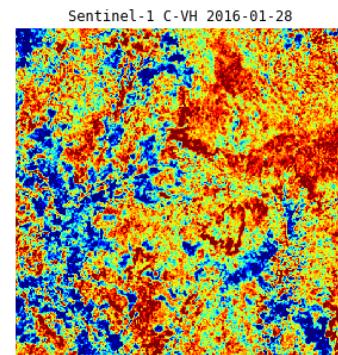
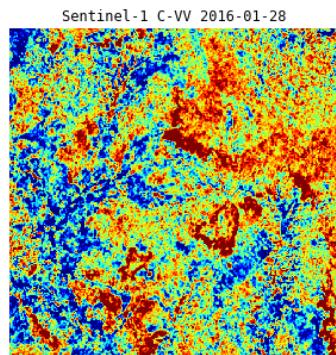
Subset Sentinel-1 (570, 40, 500, 500)
Subset Landsat (380, 26, 333, 333)
```

Now we can pick the bands and plot the Sentinel-1 and Landsat NDVI images of the subset. Change the band numbers to the bands we are interested in.

## Dry Season Plot

```
In [8]: # Dry season plot  
bandnbrs=(24,24)  
ebd_plot(bandnbrs)
```

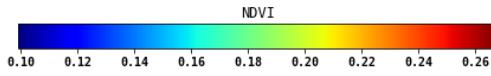
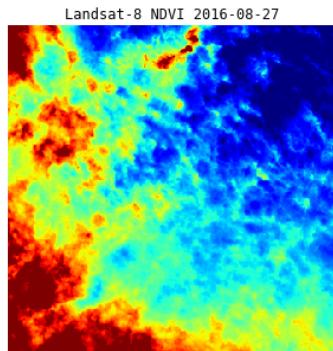
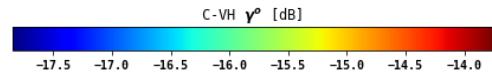
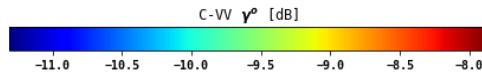
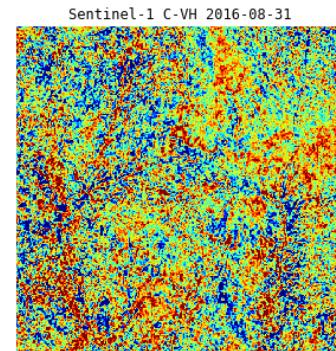
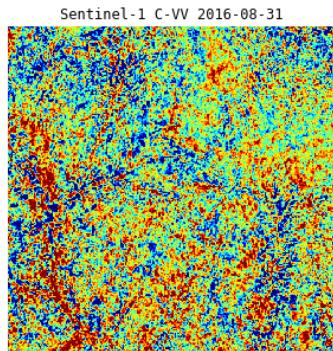
Sentinel-1 Backscatter and Landsat NDVI and FC IR



## Wet Season Plot

```
In [9]: # Wet season plot  
bandnbrs=(40,37)  
ebd_plot(bandnbrs)
```

Sentinel-1 Backscatter and Landsat NDVI and FC IR



In the figure above, for band 24 of Sentinel-1 and 24 of NDVI, which was acquired three days after the Sentinel-1 image, there is an inverse relationship. Where Sentinel-1 exhibits low backscatter, NDVI shows relatively higher NDVI. What are the reasons for this in this environment?

## Exercise

Pick different bands to compare. Look at the list of the dates for SAR data and Landsat data acquisitions in the above. One good option is to compare bands from the dry and wet seasons 2016.

# Time Series Profiles of C-Band Backscatter and NDVI

We compute the image means of each time step in the time series stack and plot them together.

## Prepare Sentinel-1 and NDVI Data stacks

### Sentinel time series stack

```
In [10]: caldB=-83
calPwr = np.power(10.,caldB/10.)

s_ts=[]
for idx in (0,1):
    means=[]
    for i in range(bands[idx]):
        rs=img_handle[idx].GetRasterBand(i+1).ReadAsArray(*subset_sentinel)
        # 1. Conversion to Power
        rs_pwr=np.power(rs,2.)*calPwr
        rs_means_pwr = np.mean(rs_pwr)
        rs_means_dB = 10.*np.log10(rs_means_pwr)
        means.append(rs_means_dB)
    s_ts.append(pd.Series(means,index=stindex))
```

### Landsat NDVI time series stack

```
In [11]: means=[]
idx=2
for i in range(bands[idx]):
    r=img_handle[idx].GetRasterBand(i+1).ReadAsArray(*subset_landsat)
    means.append(r[r!= -9999].mean()/10000.)
l_ts=pd.Series(means,index=ltindex)
```

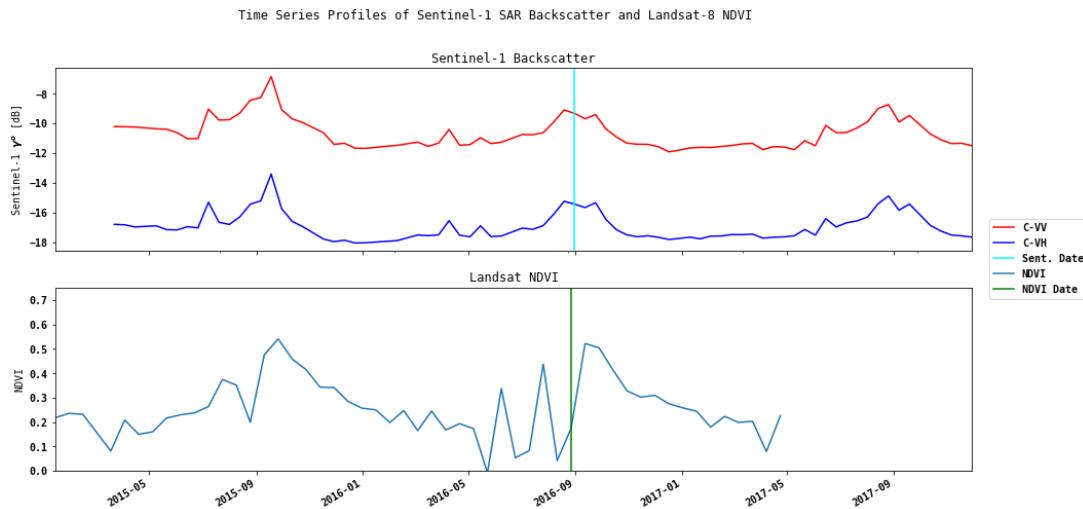
## Joint Plot of SAR Backscatter and NDVI of Image Subset Means

Now we plot the time series of the SAR backscatter and NDVI values scaled to the same time axis. We also show the time stamps for the images we display above.

```
In [12]: fig, ax = plt.subplots(2,1,figsize=(16,8))
# ax1.plot(s_ts.index,s_ts.values, 'r-')
s_ts[0].plot(ax=ax[0],color='red',label='C-VV',xlim=(min(min(ltindex),min(stindex)),
                                                    max(max(ltindex),max(stindex
x))))
s_ts[1].plot(ax=ax[0],color='blue',label='C-VH')
ax[0].set_xlabel('Date')
ax[0].set_ylabel('Sentinel-1 $\gamma^o$ [dB]')

# Make the y-axis label, ticks and tick labels match the line color. ax1.set_ylabel('exp', color='b')
# ax1.tick_params('y', colors='b')
# ax[1] = ax1.twinx()
# s_ts.plot(ax=ax[1],sharex=ax[0])
l_ts.plot(ax=ax[1],sharex=ax[0],label='NDVI',xlim=(min(min(ltindex),min(stindex
)),max(max(ltindex),max(stindex
x))),ylim=(0,0.75))
# ax[1].plot(l_ts.index,l_ts.values,color='green',label='NDVI')
ax[1].set_ylabel('NDVI')
ax[0].set_title('Sentinel-1 Backscatter')
ax[1].set_title('Landsat NDVI')

ax[0].axvline(stindex[bandnbrs[0]-1],color='cyan',label='Sent. Date')
ax[1].axvline(ltindex[bandnbrs[1]-1],color='green',label='NDVI Date')
_=fig.legend(loc='center right')
_=fig.suptitle('Time Series Profiles of Sentinel-1 SAR Backscatter and Landsat-8
NDVI ')
# fig.tight_layout()
```



## Comparison of time series profiles at point locations.

We will pick a pixel location in the SAR image, find the corresponding location in the Landsat NDVI stack and plot the joint time series.

First let's pick a pixel location in the SAR image (i.e. the reference image)

We use the geotrans info to find the same location in the Landsat image

```
In [13]: sarloc=(2000,2000)
ref_x=geotrans[0][0]+sarloc[0]*geotrans[0][1]
ref_y=geotrans[0][3]+sarloc[1]*geotrans[0][5]
print('UTM Coordinates      ',ref_x,ref_y)
print('SAR pixel/line      ',sarloc[0],sarloc[1])
target_pixel=round((ref_x-geotrans[2][0])/geotrans[2][1])
target_line=round((ref_y-geotrans[2][3])/geotrans[2][5])
print('Landsat pixel/line   ',target_pixel,target_line)
```

```
UTM Coordinates      438020.0 1350960.0
SAR pixel/line      2000 2000
Landsat pixel/line   1333 1334
```

Read the image data at these locations

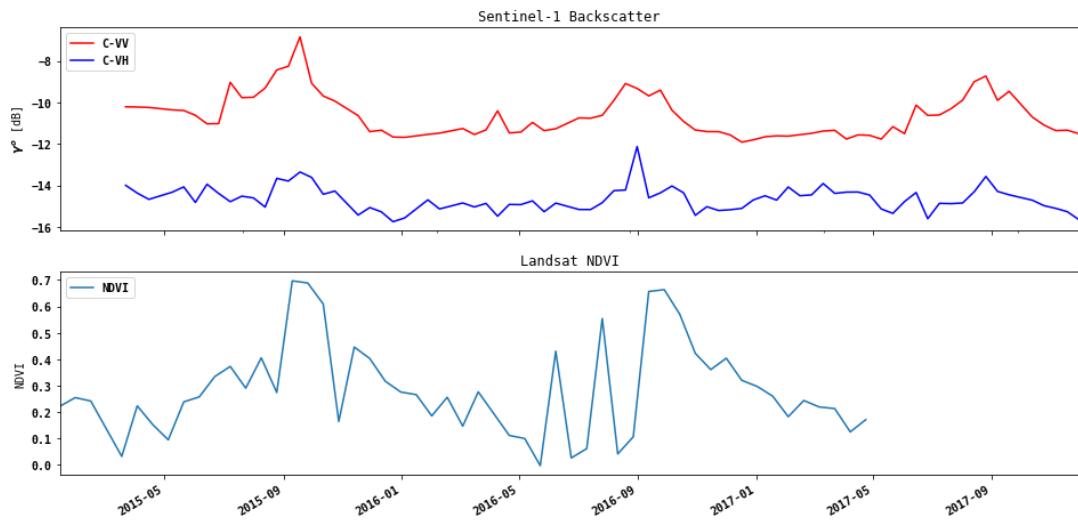
```
In [14]: s_ts_pixel=[]
for idx in (0,1):
    means=[]
    for i in range(bands[idx]):
        rs=img_handle[idx].GetRasterBand(i+1).ReadAsArray(*sarloc,6,6)
        # 1. Conversion to Power
        rs_pwr=np.power(rs,2.)*calPwr
        rs_means_pwr = np.mean(rs_pwr)
        rs_means_dB = 10.*np.log10(rs_means_pwr)
        means.append(rs_means_dB)
    s_ts_pixel.append(pd.Series(means,index=stindex))

means=[]
idx=2
for i in range(bands[idx]):
    r=img_handle[idx].GetRasterBand(i+1).ReadAsArray(target_pixel,target_line,4,
4)
    means.append(np.nanmean(r)/10000.)
l_ts_pixel=pd.Series(means,index=ltindex)
```

Plot the joint time series.

```
In [15]: fig, ax = plt.subplots(2,1,figsize=(16,8))
# ax1.plot(s_ts.index,s_ts.values, 'r-')
s_ts[0].plot(ax=ax[0],color='red',label='C-VV',xlim=(min(min(ltindex),min(stindex)),
max(max(ltindex),max(stindex))))
s_ts_pixel[1].plot(ax=ax[0],color='blue',label='C-VH')
ax[0].set_xlabel('Date')
ax[0].set_ylabel('$\gamma^o$ [dB]')

# Make the y-axis label, ticks and tick labels match the line color. ax1.set_ylabel('exp', color='b')
# ax1.tick_params('y', colors='b')
# ax[1] = ax1.twinx()
# s_ts.plot(ax=ax[1],sharex=ax[0])
l_ts_pixel.plot(ax=ax[1],sharex=ax[0],label='NDVI',xlim=(min(min(ltindex),min(stindex)),
max(max(ltindex),max(stindex))))
# ax[1].plot(l_ts.index,l_ts.values,color='green',label='NDVI')
ax[1].set_ylabel('NDVI')
ax[0].set_title('Sentinel-1 Backscatter')
ax[1].set_title('Landsat NDVI')
_=ax[0].legend(loc='upper left')
_=ax[1].legend(loc='upper left')
# fig.tight_layout()
```



Interpret these time series profiles. While generally the seasonal trends are visible in both like (VV) and cross-polarized (VH) data, and correlate well with the NDVI temporal profile, the cross-polarized response is less pronounced at the example pixel location UTM Coordinates Zone 31N 438020.0 1350960.0.

## EXERCISE

Pick different pixel locations and replot the figure above. Interpret the result with respect to forest, non-forest, deforestation and forest degradation signatures. In your interpretation look for image signals of strong rain events in the SAR data and cloud covered scenes in the Landsat imagery.

# SAR Training Workshop for Forest Applications

## How to Make RGB Composites from Dual-Polarimetric Data

Josef Kellndorfer, Ph.D., President and Senior Scientist, Earth Big Data, LLC

Revision date: January 2019

In this chapter we introduce how to make a three band color composite and save it

### Import Python modules

```
In [1]: import os,sys,gdal  
%matplotlib inline  
import matplotlib.pyplot as plt  
import matplotlib.patches as patches # Needed to draw rectangles  
from skimage import exposure # to enhance image display  
import numpy as np  
import pandas as pd  
  
In [2]: # Select the project data set and time series data  
  
In [3]: # West Africa - Biomass Site  
datapath='Users/rmuench/Downloads/wa/BIOss1/'  
datefile='S32631X398020Y1315440ss1_A_vv_0001_mtfil.dates'  
imagefile_like='S32631X398020Y1315440ss1_A_vv_0001_mtfil.vrt'  
imagefile_cross='S32631X398020Y1315440ss1_A_vh_0001_mtfil.vrt'  
  
In [4]: os.chdir(datapath)
```

We are defining two helper functions for this task

- **CreateGeoTiff()** to write out images
- **dualpol2rgb()** to compute various metrics from a time series data stack

```
In [5]: def CreateGeoTiff(Name, Array, DataType, NDV, bandnames=None, ref_image=None,
                       GeoT=None, Projection=None):
    # If it's a 2D image we fake a third dimension:
    if len(Array.shape)==2:
        Array=np.array([Array])
    if ref_image==None and (GeoT==None or Projection==None):
        raise RuntimeWarning('ref_image or settings required.')
    if bandnames != None:
        if len(bandnames) != Array.shape[0]:
            raise RuntimeError('Need {} bandnames. {} given'
                               .format(Array.shape[0],len(bandnames)))
    else:
        bandnames=[ 'Band {}'.format(i+1) for i in range(Array.shape[0])]
    if ref_image!= None:
        refimg=gdal.Open(ref_image)
        GeoT=refimg.GetGeoTransform()
        Projection=refimg.GetProjection()
    driver= gdal.GetDriverByName('GTIFF')
    Array[np.isnan(Array)] = NDV
    DataSet = driver.Create(Name,
                           Array.shape[2], Array.shape[1], Array.shape[0], DataType)
    DataSet.SetGeoTransform(GeoT)
    DataSet.SetProjection( Projection)
    for i, image in enumerate(Array, 1):
        DataSet.GetRasterBand(i).WriteArray( image )
        DataSet.GetRasterBand(i).SetNoDataValue(NDV)
        DataSet.SetDescription(bandnames[i-1])
    DataSet.FlushCache()
    return Name
```

```
In [6]: def dualpol2rgb(like,cross,sartype='amp',ndv=0):
    CF=np.power(10.,-8.3)
    if np.isnan(ndv):
        mask=np.isnan(cross)
    else:
        mask=np.equal(cross,ndv)

    l = np.ma.array(like,mask=mask,dtype=np.float32)
    c = np.ma.array(cross,mask=mask,dtype=np.float32)

    if sartype=='amp':
        l=np.ma.power(l,2.)*CF
        c=np.ma.power(l,2.)*CF
    elif sartype=='dB':
        l=np.ma.power(10.,l/10.)
        c=np.ma.power(10.,c/10.)
    elif sartype=='pwr':
        pass
    else:
        print('invalid type ',sartype)
        raise RuntimeError

    if sartype=='amp':
        ratio=np.ma.sqrt(l/c)/10
        ratio[np.isinf(ratio.data)]=0.00001
    elif sartype=='dB':
        ratio=10.*np.ma.log10(l/c)
    else:
        ratio=l/c

    ratio=ratio.filled(ndv)

    rgb=np.dstack((like,cross,ratio.data))

    bandnames=( 'Like','Cross','Ratio')
    return rgb,bandnames,sartype

def any2amp(raster,sartype='amp',ndv=0):
    CF=np.power(10.,-8.3)
    mask=raster==ndv

    if sartype=='pwr':
        raster=np.sqrt(raster/CF)
    elif sartype=='dB':
        raster=np.ma.power(10.,(raster+83)/20.)
    elif sartype=='amp':
        pass
    else:
        print('invalid type ',sartype)
        raise RuntimeError

    raster[raster<1]=1
    raster[raster>65535]=65535
    raster[mask]=0
    raster=np.ndarray.astype(raster,dtype=np.uint16)
    return raster
```

## Set the Dates

```
In [7]: # Get the date indices via pandas
dates=open(datefile).readlines()
tindex=pd.DatetimeIndex(dates)
j=1
print('Bands and dates for',imagefile_like)
for i in tindex:
    print("{:4d} {}".format(j, i.date()),end=' ')
    j+=1
    if j%5==1: print()

Bands and dates for S32631X398020Y1315440ss1_A_vv_0001_mtfil.vrt
  1 2015-03-22      2 2015-04-03      3 2015-04-15      4 2015-05-09      5 2015-05-21
  6 2015-06-02      7 2015-06-14      8 2015-06-26      9 2015-07-08     10 2015-07-20
 11 2015-08-01     12 2015-08-13     13 2015-08-25     14 2015-09-06     15 2015-09-18
 16 2015-09-30     17 2015-10-12     18 2015-10-24     19 2015-11-17     20 2015-11-29
 21 2015-12-11     22 2015-12-23     23 2016-01-04     24 2016-01-28     25 2016-02-09
 26 2016-03-04     27 2016-03-16     28 2016-03-28     29 2016-04-09     30 2016-04-21
 31 2016-05-03     32 2016-05-15     33 2016-05-27     34 2016-06-08     35 2016-07-02
 36 2016-07-14     37 2016-07-26     38 2016-08-07     39 2016-08-19     40 2016-08-31
 41 2016-09-12     42 2016-09-24     43 2016-10-06     44 2016-10-18     45 2016-10-30
 46 2016-11-11     47 2016-11-23     48 2016-12-05     49 2016-12-17     50 2016-12-29
 51 2017-01-10     52 2017-01-22     53 2017-02-03     54 2017-02-15     55 2017-02-27
 56 2017-03-11     57 2017-03-23     58 2017-04-04     59 2017-04-16     60 2017-04-28
 61 2017-05-10     62 2017-05-22     63 2017-06-03     64 2017-06-15     65 2017-06-27
 66 2017-07-09     67 2017-07-21     68 2017-08-02     69 2017-08-14     70 2017-08-26
 71 2017-09-07     72 2017-09-19     73 2017-10-13     74 2017-10-25     75 2017-11-06
 76 2017-11-18     77 2017-11-30
```

```
In [8]: # PICK A BAND NUMBER
bandnbr=70
```

Open the image and get dimensions (bands,lines,pixels):

```
In [9]: img_like=gdal.Open(imagefile_like)
img_cross=gdal.Open(imagefile_cross)
# Get Dimensions
print('Likepol ',img_like.RasterCount,img_like.RasterYSize,img_like.RasterXSize)
print('Crosspol',img_cross.RasterCount,img_cross.RasterYSize,img_cross.RasterXSize)

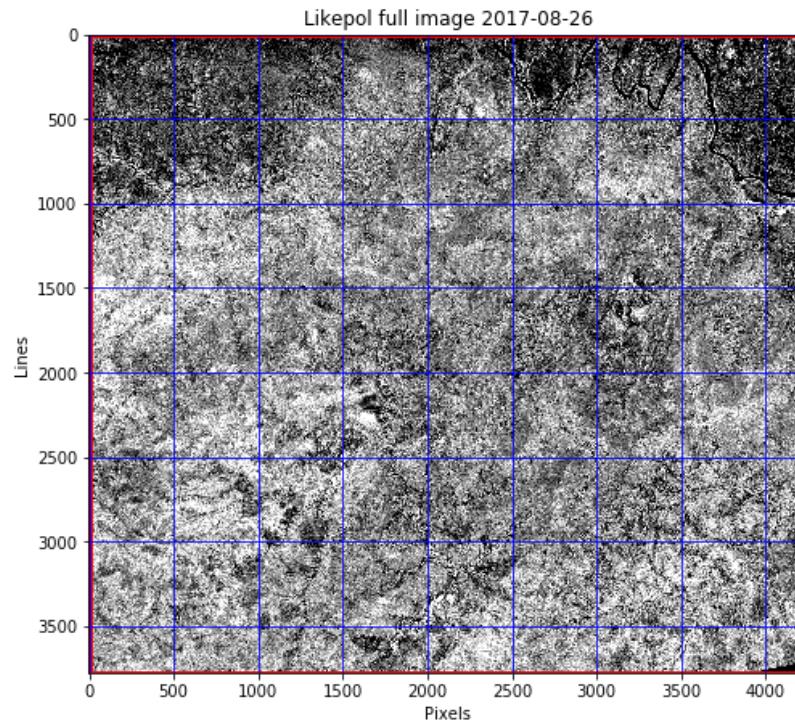
Likepol 77 3776 4243
Crosspol 77 3776 4243
```

For a manageable size we can choose a 500x500 pixel subset to read the entire data stack (commented out). We also convert the amplitude data to power data right away and will perform the rest of the calculations on the power data to be mathematically correct.

**NOTE:** Choose a different xsizes/ysize in the subset if you need to.

```
In [10]: subset=None
#subset=(3500,1000,500,500)    # (xoff,yoff,xsize,ysize)
if subset==None:
    subset=(0,0,img_like.RasterXSize,img_like.RasterYSize)

raster=img_like.GetRasterBand(bandnbr).ReadAsArray()
fig, ax = plt.subplots(figsize=(8,8))
ax.set_title('Likepol full image {}'
             .format(tindex[bandnbr-1].date()))
ax.imshow(raster,cmap='gray',vmin=np.nanpercentile(raster,5),vmax=np.nanpercentile(raster,95))
ax.grid(color='blue')
ax.set_xlabel('Pixels')
ax.set_ylabel('Lines')
# plot the subset as rectangle
if subset != None:
    _=ax.add_patch(patches.Rectangle((subset[0],subset[1]),
                                     subset[2],subset[3],
                                     fill=False,edgecolor='red',
                                     linewidth=3))
```



## Make the RGB like/cross/ratio image

```
In [11]: raster_like=img_like.GetRasterBand(bandnbr).ReadAsArray(*subset)
raster_cross=img_cross.GetRasterBand(bandnbr).ReadAsArray(*subset)
```

We make an RGB stack to display the like,cross, and ratio data as a color composite.

```
In [12]: rgb,bandnames,sartype=dualpol2rgb(raster_like,raster_cross)
```

We are interested in displaying the image enhanced with histogram equalization.

We can use the function `exposure.equalize_hist()` from the skimage.exposure module

```
In [13]: rgb_stretched=np.ndarray.astype(rgb.copy(),'float32')
# For each band we apply the stretch
for i in range(rgb_stretched.shape[2]):
    rgb_stretched[:, :, i] = np.ndarray.astype(exposure.equalize_hist(rgb_stretched[:, :, i]),
                                                'float32')
mask=~np.equal(rgb_stretched[:, :, i], 0), 'float32')
```

```
In [14]: rgb_stretched
```

```
Out[14]: array([[[6.78325057e-01, 9.72052574e-01, 3.90828580e-01],
[8.76864195e-01, 9.84529495e-01, 1.71240658e-01],
[9.83614683e-01, 9.98086393e-01, 2.89953277e-02],
...,
[5.28072240e-03, 1.07818116e-02, 9.95057702e-01],
[6.99684722e-03, 1.37452455e-02, 9.93614554e-01],
[3.84910442e-02, 9.95012820e-02, 9.67868745e-01]],

[[5.00227690e-01, 9.16491270e-01, 5.65926731e-01],
[6.34485126e-01, 9.05717731e-01, 4.34903175e-01],
[8.29794526e-01, 9.68486011e-01, 2.25173086e-01],
...,
[5.95453382e-03, 1.87972672e-02, 9.94502723e-01],
[6.23923307e-03, 1.84695572e-02, 9.94265974e-01],
[2.66998932e-02, 9.82193723e-02, 9.77528811e-01]],

[[1.96703568e-01, 4.68630195e-01, 8.36585879e-01],
[2.68557519e-01, 4.29826617e-01, 7.75374472e-01],
[4.55813646e-01, 7.10059583e-01, 6.08429134e-01],
...,
[1.46300010e-02, 1.38415113e-01, 9.87303555e-01],
[1.15622561e-02, 9.69374627e-02, 9.89897549e-01],
[2.56231278e-02, 1.70091271e-01, 9.78379309e-01]],

...,

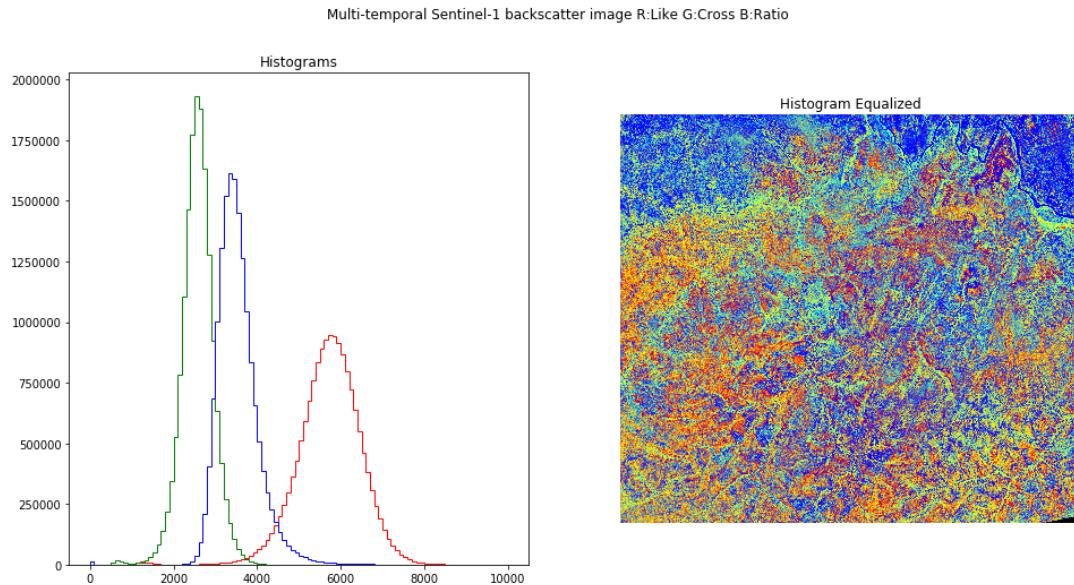
[[5.30907631e-01, 7.67807901e-01, 5.37106931e-01],
[2.64292091e-01, 5.22130668e-01, 7.78782785e-01],
[1.08426727e-01, 3.33118588e-01, 9.10773575e-01],
...,
[1.24921982e-07, 1.87382980e-07, 2.49843964e-07],
[1.24921982e-07, 1.87382980e-07, 2.49843964e-07],
[1.24921982e-07, 1.87382980e-07, 2.49843964e-07]],

[[4.51068103e-01, 8.14454377e-01, 6.13029897e-01],
[2.54053503e-01, 6.20940566e-01, 7.87545741e-01],
[1.29825607e-01, 3.66306335e-01, 8.92777562e-01],
...,
[1.24921982e-07, 1.87382980e-07, 2.49843964e-07],
[1.24921982e-07, 1.87382980e-07, 2.49843964e-07],
[1.24921982e-07, 1.87382980e-07, 2.49843964e-07]],

[[4.22746390e-01, 8.42523038e-01, 6.38080597e-01],
[2.41145849e-01, 6.77982450e-01, 7.98704743e-01],
[1.81992248e-01, 4.34495091e-01, 8.49451780e-01],
...,
[1.24921982e-07, 1.87382980e-07, 2.49843964e-07],
[1.24921982e-07, 1.87382980e-07, 2.49843964e-07],
[1.24921982e-07, 1.87382980e-07, 2.49843964e-07]]], dtype=float32)
```

Now let's display the histograms and equalized image side by side.

```
In [15]: fig,ax = plt.subplots(1,2,figsize=(16,8))
fig.suptitle('Multi-temporal Sentinel-1 backscatter image R:{} G:{} B:{}'
             .format(bandnames[0],bandnames[1],bandnames[2]))
plt.axis('off')
ax[0].hist(rgb[:, :, 0].flatten(),histtype='step',color='red',bins=100,range=(0,1000))
ax[0].hist(rgb[:, :, 1].flatten(),histtype='step',color='green',bins=100,range=(0,10000))
ax[0].hist(rgb[:, :, 2].flatten(),histtype='step',color='blue',bins=100,range=(0,1000))
ax[0].set_title('Histograms')
ax[1].imshow(rgb_stretched)
ax[1].set_title('Histogram Equalized')
_=ax[1].axis('off')
```



## Write the images to an output file

### Determine output geometry

First, we need to set the correct geotransformation and projection information. We retrieve the values from the input images and adjust by the subset:

```
In [16]: proj=img_like.GetProjection()
geotrans=list(img_like.GetGeoTransform())

subset_xoff=geotrans[0]+subset[0]*geotrans[1]
subset_yoff=geotrans[3]+subset[1]*geotrans[5]
geotrans[0]=subset_xoff
geotrans[3]=subset_yoff
geotrans=tuple(geotrans)
geotrans
```

```
Out[16]: (398020.0, 20.0, 0.0, 1390960.0, 0.0, -20.0)
```

## Convert to 16bit Amplitude image

We use the root of the time series data stack name and append a \_ts\_metrics\_.tif ending as filenames

### Build a like/cross/ratio amplitude scaled GeoTIFF images

```
In [17]: outbands=[ ]
for i in range(3):
    outbands.append(any2amp(rgb[:, :, i]))

imagefile_like.replace('_vv_','_lcr_').replace('.vrt','_{}.tif'.format(
dates[bandnbr-1].rstrip()))
bandnames=['Like','Cross','Ratio']
Array=np.array(outbands)
CreateGeoTiff(imagefile_like,Array,gdal.GDT_UInt16,0,bandnames,GeoT=geotrans,Projectio
n=proj)

Out[17]: 'S32631X398020Y1315440sS1_A_lcr_0001_mtfil_20170826.tif'
```

This Image can now be loaded into QGIS or similar programs

## Exercise

Change the bandnbr, generate a new rgb image and export it. Display in QGIS





**DR. PAUL SIQUEIRA** is a Professor of Electrical and Computer Engineering at the University of Massachusetts in Amherst and is a co-Director of the Microwave Remote Sensing Laboratory, which has a long history in the development of Microwave Remote Sensing instrumentation for Earth Science Applications. Professor Siqueira is the Ecosystems Lead for the US NISAR mission, a joint project between NASA and the Indian Research Organization. Professor Siqueira has over 20 years experience in the use of SAR, has been a visiting scientist at the European Commission's Joint Research Center, was as a Senior Engineer at NASA's Jet Propulsion Laboratory and held a Bullard Fellowship at the Harvard Forest.

*The author of this chapter would like to acknowledge the fundamental contributions of Yang Lei in creating the algorithms and background that make up the backbone of this work, and to Tracy Whelen for contributing to the training exercises. We would also like to recognize the Japanese Aerospace Agency's Kyoto and Carbon Initiative which has provided the spaceborne L-band SAR data on top of which the FSH algorithm is based. Finally, we would like to thank SERVIR, USAID, NASA and ADPC for making this interaction a reality.*

# CHAPTER 4

## Forest Stand Height Estimation

Paul Siqueira, Professor of Electrical and Computer Engineering, Microwave Remote Sensing Laboratory, University of Massachusetts, Amherst

### ABSTRACT

The measurement of forest structural characteristics is important for a variety of Monitoring, Reporting, and Verification (MRV) protocols in resource management. One characteristic of particular importance is Forest Stand Height (FSH), or the average height of trees in a forest stand. In this context, FSH can be used as an indicator of the age of a forest stand, plant and animal habitats, and the amount of Above Ground Biomass (AGB) held in the forest stand. FSH can be measured through the use of terrestrial and/or airborne lidar, with airborne lidar being especially useful due to its wide area coverage and direct measurement of forest height. A difficulty with airborne measurements, however, is that while these measurements work well at the tens- to hundreds-of-hectares-level, they are difficult to scale beyond that.

One method for the spatial scaling of FSH is through the use of spaceborne Synthetic Aperture Radar (SAR), especially at L-band repeat-pass Interferometric SAR (InSAR), which can be obtained through repeat observations from ALOS-2 and the future NISAR mission. In this scenario, the measure of InSAR decorrelation can be related to FSH through the use of localized training data obtained from lidar. This chapter focuses on the use of repeat-pass InSAR for FSH estimation, and presents the theory, software, and examples of these methods. Although there is currently a limited availability of L-band SAR from ALOS-2, when NISAR launches in 2021, the presented method of FSH determination can be applied over large regions, especially when initialized using instruments such as the Global Ecosystems Dynamics Investigation Lidar (GEDI) aboard the International Space Station, or other lidar observations.

## 4.1 Theory

### 4.1.1 BACKGROUND AND BASIC CONCEPTS

The average height of trees in a forest stand, or Forest Stand Height (FSH), is an indicator of the age of a forest stand and an important forest structure metric that helps to characterize (1) plant and animal habitats, (2) the history of land use, and (3) the amount of Above Ground Biomass (AGB) held in the forest stand. The size of the forest stand in this context is minimally 1 ha in size, but is generally larger depending on the homogeneity of the forested region. In general, when using remote sensing data to estimate FSH, the smaller the size of the land unit, the less accurate the FSH estimate will be. This is due to basic sampling statistics and estimation errors that are incurred when a statistically varying quantity (such as forest height) is measured remotely.

#### 4.1.1.1 Relating SAR to Forest Stand Height

SAR sensitivity to FSH is based on three funda-

mental SAR properties. These three fundamental properties are discussed below and are illustrated in **Figure 4.1**:

- (1) As the number of scatterers increase within a SAR resolution cell, so does the reflected power. This trend is moderated by the effect of attenuation of signals as they pass through a forested canopy, and is directly related to the saturation effect seen in backscatter to biomass relationships (discussed in **Chapter 5**). Insomuch as the number of scatterers increases with FSH and forest density, observations of the backscatter power from radar can be used as an indirect measure of FSH. This relationship is often obtained through an empirical relationship between the two variables. It should be noted that SAR data can have a number of different polarization combinations, with the simplest being a co-polarized return, such as HH or VV (see **Chapter 2**); followed by dual-polarized, which is a combi-
- nation of one of the co-polarized returns with its cross-polarized counterpart (HH with HV, or VV with VH); and finally, the quad-polarized signature, which is the most complicated as it has all four components (HH, HV, VV, and VH) of the polarimetric scattering matrix. Because of the sensitivity of the cross-polarized signature to the multiple scattering that occurs in vegetated environments, the cross-polarized channels of the backscatter power are most often used for characterizing forest structure.
- (2) In addition to the power measured in a SAR backscatter image, SAR can also very accurately measure the distance to targets. When the height of target is not accurately known, there exists an ambiguity in the geometric relationship between the target and the SAR sensor, principally through the look angle, which is defined as the angle between the nadir direction of the SAR and the vector pointing from the SAR to the target.

When two SAR observations are made, however, this angle can be determined very accurately through some basic trigonometric calculations and indeed can be used for measuring the topography of the Earth through a process known as Interferometric SAR, or InSAR. If the measure of InSAR height can be modeled relative to the bare ground surface, and if the topography of that surface can be determined through other means, then an estimate of the vegetation height can be determined by the difference between the InSAR-measured height and ground surface Digital Elevation Model (DEM).

In places where the topographic height is not well defined (e.g., in a forest canopy where the interferometrically measured height can mean either at the ground surface or the canopy top), a unique interferometric signature arises in which the detected height from the interferometer can be shown to be a random number. Its mean is an extinction weighted average of the radar signal penetration into the canopy. The term “extinction weighted average” refers to the loss of signal strength (extinction) as a radar signal penetrates a forest canopy. Hence, parts at the top of the canopy will contribute more to the backscatter signature than the bottom of the

canopy. This depth of penetration is proportional to the signal wavelength (24 cm for L-band and 5-cm for C-band) and the density of scatterers. For interferometric applications, the vertical distribution of scatterers plays a role in the overall signature, and hence the use of the term “extinction weighted average.” The magnitude of this weighted average is known as the “interferometric coherence,” a normalized value with a range between 0 and 1. InSAR sensitivity to FSH statistics have led to a number of approaches to be explored using spaceborne satellites (e.g., Treuhart & Siqueira 2000, Cloude & Papathanasiou 2001).

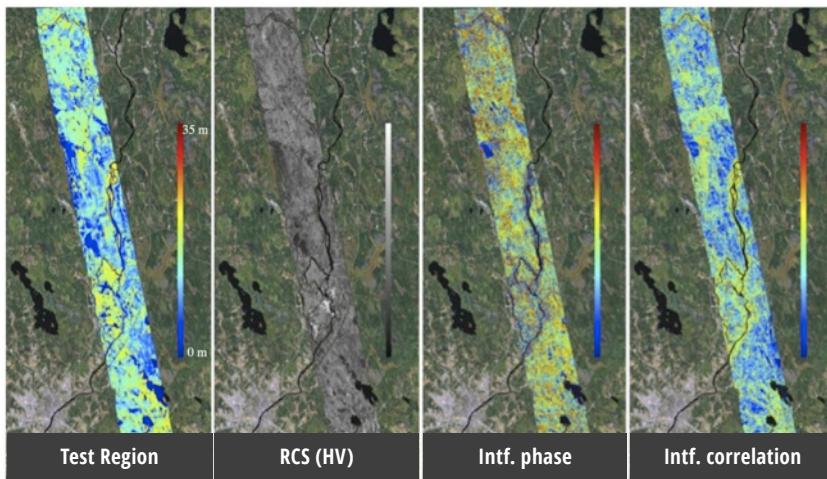
(3) For InSAR to work well over vegetated surfaces in the previously described manner, it is important to make the SAR observations simultaneously, or as close together in time as possible. This is because if the observations are made at different times, the targets within a SAR resolution cell may have moved, and this movement will cause an error in measuring the trigonometric look angle and will create a reduction in the interferometric coherence. This process is known as “temporal decorrelation,” that is, the more that a target changes between observations, the lower the coherence will be.

When an InSAR system makes both observations at the same time (typically requiring two satellites or a single airborne platform with two antennas), it is known as “single-pass InSAR.” Conversely, if the observations of the scene are made at different times, this is called “repeat-pass InSAR.”

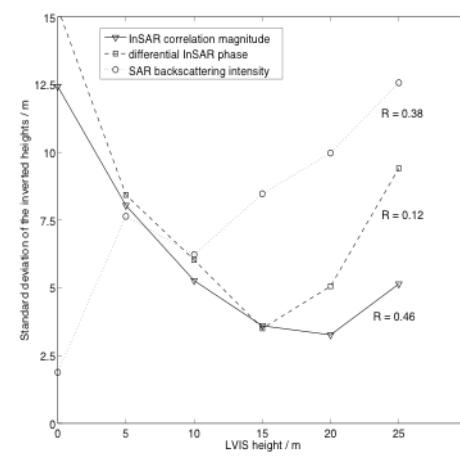
One way FSH can be estimated from repeat-pass InSAR is to measure the amount of temporal decorrelation that has occurred between passes and to make the broad assumption that the taller a tree (or forest stand) is, the more movement that will occur between passes of the satellite. Hence, when the interferometric coherence is measured, it can indirectly (through an empirical relationship) be used to estimate FSH.

As in the case of backscatter to biomass relationships, the cross-polarized channel (HV) of the interferometric coherence is more sensitive to FSH than the co-polarized channels (HH and VV).

Based on the principles highlighted above, a set of algorithms has been created for estimating FSH from InSAR observations. Because most spaceborne SAR systems cannot perform single-pass interferometry, the FSH algorithm relies on the repeat-pass relationship between interferometric coherence and vegetation height.



**Figure 4.1** Illustration of the three principles behind the relationship of SAR measurements to vegetation height. Shown from left to right are (a) a test region located in the U.S. state of Maine imaged by the LVIS lidar sensor, (b) the radar backscatter intensity for the region (grayscale), (c) the height difference between L-band repeat-pass SAR and the ground surface DEM, and (d) a height estimate based on the interferometric correlation. The graphic at the right (e) shows the FSH error relative to the lidar measurement for each of the three SAR methods derived from the cross-polarized signal. It can be seen from the plot that for vegetation heights of less than 10 m, the backscatter intensity is most accurate. For vegetation taller than 10 m, the InSAR coherence proves to be more accurate.



#### 4.1.1.2 Mission Platforms for Estimating Forest Stand Height

At the time of this writing, and algorithm development, for spaceborne applications with a global extent, there are only two SAR systems with single-pass interferometry. One of these is the Shuttle Radar Topography Mission (SRTM), a C-band InSAR flown on board NASA's space shuttle for an 11-day mission in February 2000. The other is TanDEM-X, flown by the German Space Agency (DLR), where data were collected by two co-orbiting satellites at X-band in 2010 and made mostly available through commercial arrangements. Because both satellites operate with wavelengths less than 10 cm, the signals from SRTM and TanDEM-X do not penetrate far into the canopy, and without a model for the ground surface DEM, will have difficulty estimating FSH.

Upon implementation, a significant source of error in estimating coherence is related to thermal noise. As the amount of backscatter power received from a target decreases, an increased proportion of the coherence measurement is related to the signal that remains. In the case of a radar system, the residual signal not originating from the target itself is considered thermal noise (or simply instrument noise). Since bare surfaces (especially smooth surfaces) do not have a strong backscatter signal, the error in measuring interferometric coherence is large. Hence, the error in FSH estimation increases with decreasing values of vegetation height. For this reason, the best estimate of FSH made from repeat-pass interferometry is made from a combination of SAR backscatter power and InSAR coherence. For this reason, the approach described here can be referred to as a combined SAR/InSAR estimation of FSH.

With respect to theory, a final note should be made about this method's sensitivity to the observing SAR's wavelength. For most terrestrial remote sensing systems, wavelengths range between ~1 m (P-band) to ~3 cm (X-band) (for more information on SAR wavelengths, see [Chapter 2, Section 2.3.1](#)). Because vegetation structures are on the order of some tens of centimeters, forest vegetation is often best observed using P- and L-bands (~24 cm). For L-band SAR, only the Japanese Aerospace Exploration

Agency's (JAXA's) JERS-1 and ALOS-1 and -2 satellites are available, but are limited due to their observing strategy and data distribution policy. The European Space Agency's (ESA's) Sentinel-1a and -1b satellites that operate at C-band (5 cm) are potential resources, but are limited for repeat-pass InSAR because of the short wavelength and dominance of temporal decorrelation over vegetated targets.

This leaves the capacity for estimating FSH on a global basis to future satellite systems. Of these, there are three upcoming missions that may fill this need:

- (1) The Argentinian Space Agency's (CONAE's) L-band SAOCOM mission that was launched in 2018. The observing plan and data availability for this mission are currently not known.
- (2) ESA's P-band Biomass mission, which will launch in the 2021–2022 timeframe. This will be a first-of-its-kind spaceborne P-band repeat-pass InSAR.
- (3) NASA and the Indian Space Research Organization's (ISRO's) L-band and S-band (10 cm) NISAR mission, which will launch in late 2021 or early 2022. Data will be freely available and have global coverage at L-band.

With the NISAR mission in mind, and understanding the C-band wavelength limitations of ESA's Sentinel-1 data, prototyping of FSH algorithms have concentrated on L-band using geographically limited ALOS data as a proxy.

#### 4.1.1.3 Additional Theoretical and Applied Background

To learn more about the FSH algorithm and to access Python-based scripts for executing the algorithms described here, refer to the following journal articles:

- An introductory paper on the topic:

*Lei, Y., P. Siqueira, "Estimation of Forest Height Using Spaceborne Repeat-Pass L-Band InSAR Correlation Magnitude over the US State of Maine," Rem. Sens., 6(11), 10252-10285, 2014.*

- An automated method for mosaicking FSH data and minimizing errors

*Lei, Y., P. Siqueira, "An Automatic Mosaicking Algorithm for*

*the Generation of a Large-Scale Forest Height Map Using Spaceborne Repeat-Pass InSAR Correlation Magnitude," Rem. Sens., 7(5), 5639-5659, 2015.*

- An article describing the theory behind the approach

*Lei, Y., P. Siqueira, R. Treuhaft, "A physical scattering model of repeat-pass InSAR correlation for vegetation," Wvs. Rand. Cmpx. Med., 27(1), 129-152, 2017.*

- Application of FSH and Repeat-pass InSAR for Forest disturbance detection

*Lei, Y., R. Lucas, P. Siqueira, M. Schmidt, and R. Treuhaft, "Detection of forest disturbance with spaceborne repeat-pass SAR interferometry," IEEE Trans. Geosci. Rem. Sens., 56(4), 2424-2439, Apr 2018.*

- Statistical evaluation of the FSH algorithm over a wide area

*Lei, Y., P. Siqueira, N. Torbick, M. Ducey, D. Chowdhury, and W. Salas, "Generation of large-scale moderate-resolution forest height mosaic with spaceborne repeat-pass SAR interferometry and lidar," To be published IEEE Trans. Geosci. Rem. Sens., 34 pp., 2019.*

## 4.1.2 PROCESSING TECHNIQUES

InSAR data processing for FSH estimation requires either raw satellite data that have been downlinked but not processed, or SAR data that have been processed into Single Look Complex (SLC) imagery that is appropriate for forming interferograms. If the user has access to SLCs directly, then it is recommended to begin from there. If only the raw data are available, then some additional processing is necessary. One advantage to beginning with raw data is that the output formats of the interferograms and ancillary data are assembled in such a way as to make it easy to follow-on the processing with additional steps implemented to estimate FSH.

Software for processing raw data into SLCs can be obtained both commercially and through open source licensing agreements. Of the open source licensing processors, there are two that have been used for processing raw ALOS data into SLCs and then into FSH estimates. These are ROI\_PAC (Repeat Orbit Interfer-

ometry PACkage) and ISCE (InSAR Scientific Computing Environment). In this document, ROI\_PAC is used because it has completed its development lifetime and is somewhat easier to obtain than ISCE. At the time of this writing, ISCE continues to be developed, whereas ROI\_PAC is not. With this in mind, the scripts that estimate FSH from SLCs have been designed to work with both ROI\_PAC and ISCE.

It is important to note that while Python programs can be run in Windows, Mac OS X, and Unix environments, processing raw data into interferograms using the methods described requires a Unix or Linux environment. For this reason, it is assumed that the reader has access to these types of computing capabilities and is familiar with operating inside of them.

The following sections describe three steps: (1) downloading and processing ALOS data, (2) staging of ground validation data (necessary for establishing empirical relationships between SAR backscatter power and interferometric coherence to forest height), and (3) running the FSH algorithms. Users starting with SLC data may begin at the second step.

#### 4.1.2.1 Processing ALOS Data

To understand SAR data processing for FSH estimation, it is helpful to refer to a particular software so that the user can conceptualize the steps necessary to process SAR data. This section begins with a short description on how to obtain and install the ROI\_PAC software.

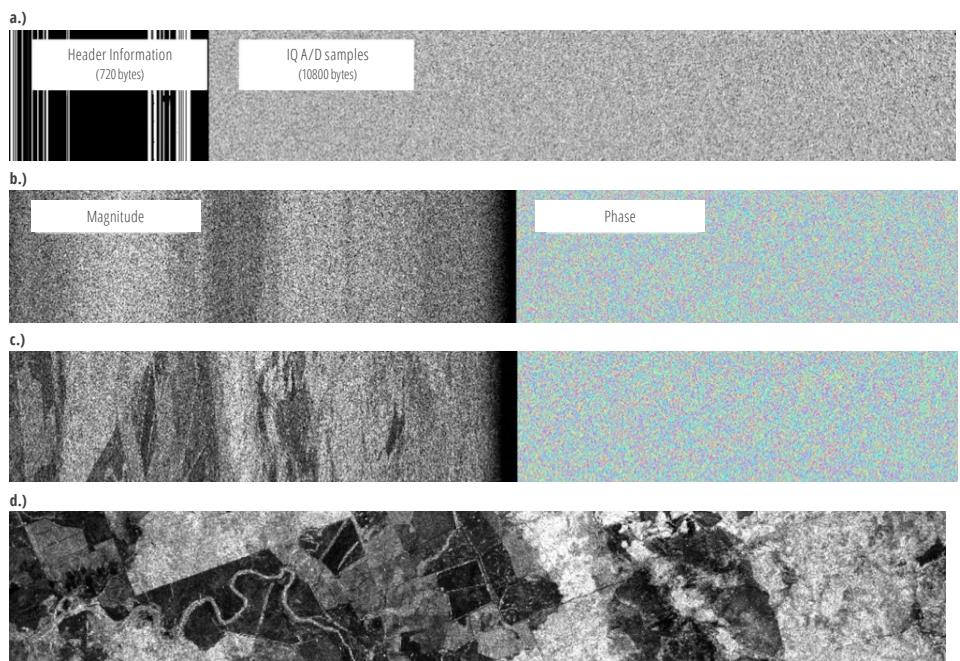
##### 4.1.2.1.1 Installing and Testing ROI\_PAC

In this work, the ROI\_PAC processing software can be obtained in TGZ (i.e., gzipped TAR) format at [http://www.openchannelfoundation.org/projects/ROI\\_PAC](http://www.openchannelfoundation.org/projects/ROI_PAC). To fully install the ROI\_PAC software, it is also necessary to have available a Fortran compiler (e.g., gfortran) and the FFTW library. Additional details for the installation of ROI\_PAC software can be found at <http://roipac.org/cgi-bin/moin.cgi/Installation>.

The ROI\_PAC software distribution comes with a test dataset that can be processed by ROI\_PAC to test the software installation. The details of this test processing can be found in the ROI\_PAC installation subdirectory fullpath/contrib/multtest.sh, where fullpath refers to the folder that the ROI\_PAC installation archive is unzipped.



**Figure 4.2** Processing chain for SAR data showing the steps that occur in the transition of a SAR image from raw data into processed data. Interferometric analysis should be done at Level 1.1. Level 2.0 refers to data that have been multi-looked, corrected for terrain effects, etc. Level 3.0 data (not shown) refers to data that have been interpreted in some way, either through classification or parameter estimation. Note that the different level numbering specified in the headings of the processing steps may vary from space agency to space agency.



**Figure 4.3** The four steps of processing ALOS SAR data beginning from (a) raw samples from the satellite, (b) range compression, (c) azimuth compression resulting in an SLC, and (d) projection into map coordinates (Level 2.0). Shown in parts (b) and (c) is the signal phase used in interferometry for determining topographic height and coherence.

##### 4.1.2.1.2 SAR Processing

Processing SAR data from raw digital values retrieved from the satellite into what ultimately becomes SAR imagery can be a detailed and complex process. In the processing of SAR data, corrections are made to account for the motion of the satellite and for the image projection effects that arise from the atmosphere, viewing geometry, and topography of the Earth. A summary of the basic steps executed in processing are shown in **Figure 4.2**. An illustration of SAR data as they are processed from raw imagery into map-projected ground-range (i.e., Level 2.0) is shown in **Figure 4.3**.

#### 4.1.2.2 Staging of Ground Validation Data

FSH ground validation data is an important component of the data processing methodology necessary for converting interferometric SAR data into FSH estimates. The FSH algorithms are implemented such that they can ingest geographically explicit data of measured (either ground-based or lidar) forest heights through the GeoTIFF format. The following subsection provides a review of the ground validation data necessary for the running of FSH.

##### 4.1.2.2.1 Types of Ground Validation Input

There are two types of ancillary ground validation data that are necessary for completing the specification of the empirical models used for the estimation of FSH from SAR data: (1) a Forest/Non-Forest (FNF)

mask that indicates where in the image the estimates should be exercised, and (2) a map of locations where forest height has been previously determined and will be used by the FSH algorithm for the training of the empirical models.

The FNF map can be derived from a number of sources or created independently by the user. Examples of external data sources that can be used to derive an FNF mask are (1) JAXA's FNF mask, (2) the U.S. National LandCover Dataset (NLCD), and (3) ESA's CCI Landcover (formerly GlobCover). From datasets such as these, a determination can be made where forests are situated and hence, where it is desired to estimate FSH. The contents of the FNF mask should be such that all regions where FSH should be estimated have a value of 0, and all regions where FSH should not be estimated have a value of 1. An example of this classification is shown in **Figure 4.4(a)**.

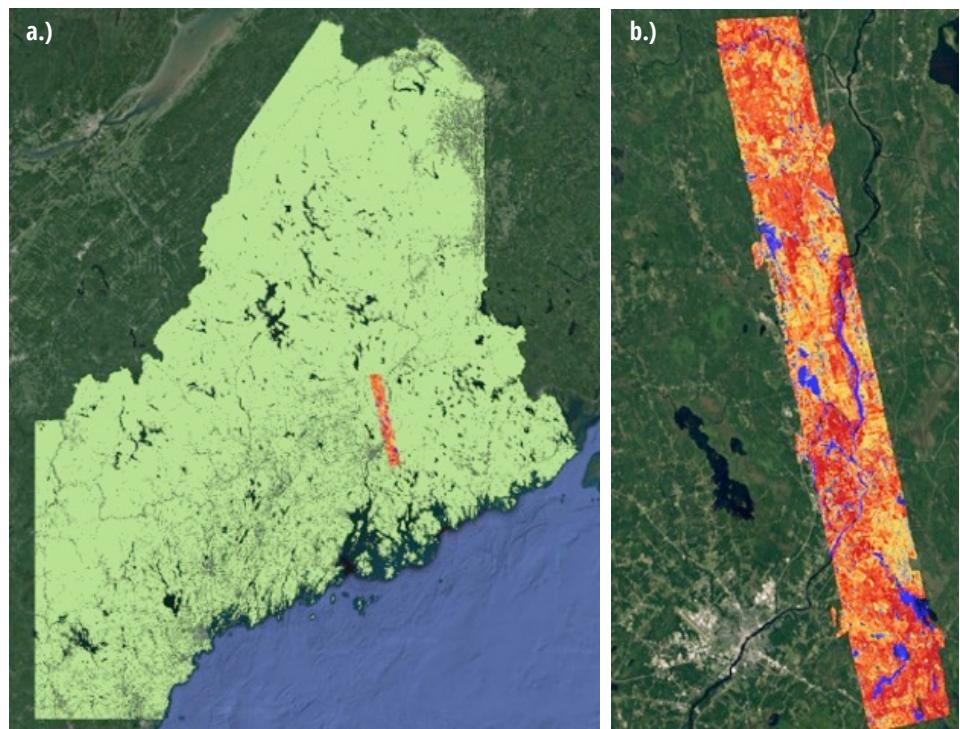
#### 4.1.2.2 Use of Lidar for Forest Stand Height Model Development

To determine values for the empirical models that relate radar backscatter power and interferometric coherence to FSH, some independent measure of forest height is necessary. Because of its ability to acquire accurate measurements of vegetation height over an extended geographic region, lidar is a preferred method for determining the coefficients that parameterize these models. An example of lidar data for a region in Maine, U.S., is shown in **Figure 4.4(b)**, which was derived from the Laser Vegetation Imaging Sensor (LVIS) operated by NASA's Goddard Space Flight Center.

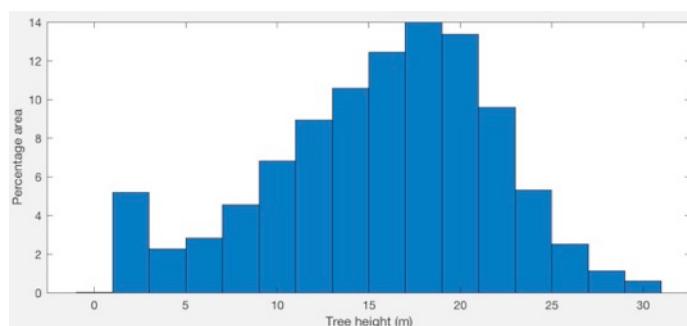
The LVIS data in **Figure 4.4** show vegetation height gridded into 30-m pixels, converted into a GeoTIFF format, and visualized using QGIS software. A resolution of 30 m was selected for this example because it is commensurate with the LVIS spot size of 25 m and the multi-looked resolution of the L-band SAR data. An example of the distribution of LVIS-estimated tree heights is shown in **Figure 4.5**.

#### 4.1.2.3 Alternative Methods for Estimating Forest Height

If lidar data are not available, then another form of independent forest height measurement over the training area needs to be identified or created. Since



**Figure 4.4** Examples of ground validation input for the FSH algorithm: (a) Optical image overlain with the FNF mask (green areas indicate regions that will be estimated for FSH), and (b) image of Laser Vegetation Imaging Sensor- (LVIS-) derived vegetation height, where blue indicates zero height, and dark red indicates the maximum height of 25 m. Data such as these are used for determining coefficients for the empirical models that relate the radar backscatter and interferometric coherence to vegetation height.



**Figure 4.5** Histogram of lidar-derived tree heights used for the training of empirical models of FSH. The spatial resolution of the LVIS data used here is 30 m.

the FSH estimator is only accurate to the 3- to 5-m level, a simple solution would be to perform a land-cover classification of a region using optical data. Stands of different ages and species composition will have different heights, which can be estimated from the ground to the same accuracy as the FSH. During the development and testing of the FSH algorithm described here, this approach was used at times. However, the results have been somewhat mixed in terms of success.

As a final approach, it should be noted that freely available satellite resources of lidar data are either

available or soon to become available. Notable among these are ICESAT-1 and -2, as well as the upcoming NASA GEDI mission.

#### 4.1.2.3 Running Forest Stand Height Algorithms

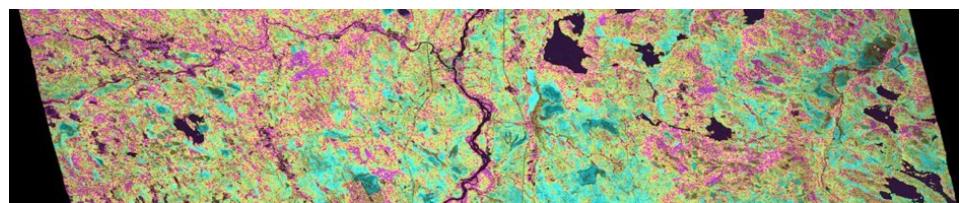
In order to run the FSH algorithms, it is assumed that the first two steps of the process described in **Section 4.1.2** have been accomplished: (1) the creation or obtaining of SLCs and (2) the obtaining of an FNF mask and vegetation height ground validation data. Once these

two steps have been accomplished, the data should be organized in a file structure such that individual folders hold results from individual interferograms between two dates (the SLCs and ancillary data for individual scenes (frames) and orbit (path) numbers). For any one frame and path number, there may exist multiple interferograms, related to multiple repeat-pass combinations of data from two different dates. These interferograms should be stored in subdirectories with the naming convention: `int_date1_date2`. Scenes from differing frames and paths can be interferometrically processed in order to create an estimate of FSH over an extended geographic region.

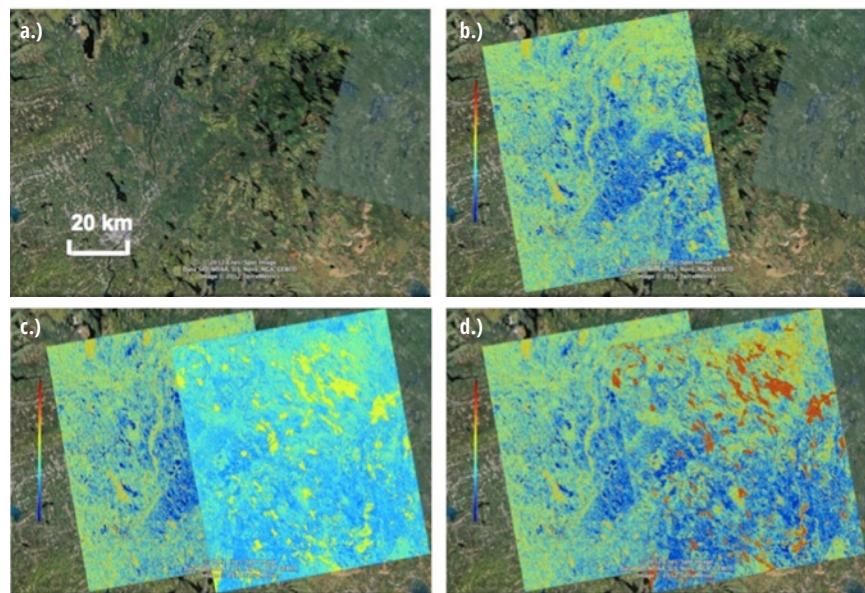
The interferogram subdirectories will hold all of the data and information necessary for creating and documenting interferograms made for an observation on two specific dates (date1 and date2). For ROI\_PAC-processed data, the most important file looks like `geo_date1-date2_2rlks.cor` and `geo_date1-date2_2rlks.cor.rsc`. The resource (“.rsc”) file is a text file that has information on the location and size of the geolocated correlation data held in `geo_date1-date2_2rlks.cor`. The format of the correlation file is known as “sample-interleaved,” or an RMG format file. An image of interferometric coherence (color) overlain on a geo-referenced image of radar backscatter cross-polarized power is shown in **Figure 4.6**.

Because radar data is organized in terms of orbits and scenes, in order to make a map of FSH over an extended geographic region, it is necessary to mosaic the images. While the process of mosaicking can be done either before or after the estimation of FSH, it is best to do so beforehand to take advantage of the overlap region between images in adjacent paths. In these regions, while the value of the coherence magnitude may vary due to the fact that the observations (and image pairs) have occurred from different orbits (and hence different dates), the overlap regions can be used to correct for these temporal differences and to adjust the coefficients for the empirical relationships of the SAR products to estimates of FSH. An example of this process is shown in **Figure 4.7**.

Once the data have been organized into directories of scenes described by their individual row and path numbers, and the interferograms have been examined



**Figure 4.6** A combined image of interferometric coherence (color) and cross-polarized backscatter power (brightness). The interferometric coherence in this image ranges between 0.1 (magenta) and 0.6 (cyan). Regions of low interferometric coherence are likely due to the presence of vegetation.



**Figure 4.7** Example of FSH/coherence equalization through the use of overlapping image regions: (a) optical image in central Maine, (b) an estimate of FSH for this region (color scale on the left extends from blue (0 m) to red (35 m)) where lidar data were available from LVIS, (c) an unconstrained estimate of FSH from an adjacent satellite pass, and (d) a corrected estimate of FSH for both scenes included in the mosaic. Color scale for all figures is the same (from Lei & Siqueira 2014).

to determine which SLC pairs yield data with the highest coherence (i.e., the least amount of temporal decorrelation), there remains the task of creating what is known as a “flag file” and a “link file.”

In this context, the *flag file* is a listing of all of the interferograms to be used in creating the region-wide mosaic of FSH. In the case discussed here, there are three such row/path combinations that will create a three-scene mosaic of FSH located in central Maine. The middle of the three scenes overlaps with the LVIS data discussed in **Section 4.1.2**, and all scenes are within

the region where identification of FNF is used for determining geographic locations where the FSH algorithm will be applied. An example of the contents of a flag file (in text format) is at the bottom of the page:

In this example, the first column of numbers indicates the interferogram number, the second column is the root file name of the data that forms the interferogram, the third and fourth columns are the dates that the data were collected for the interferometric pairs, the fifth and sixth columns give the satellite path and orbit numbers (respectively), and the last

001	890_120_20070727_HV_20070911_HV	070727	070911	890	120	HV
002	890_119_20070710_HV_20071010_HV	070710	071010	890	119	HV
003	890_118_20070808_HV_20070923_HV	070808	070923	890	118	HV

column indicates the polarization of the data.

A question that may arise when looking at the flag file is if it is possible to use other polarization combinations (e.g., HH and/or VV) for the determination of FSH. Indeed, such combinations were tested in the development stages of the algorithm, and it was determined that cross-polarized data worked best because of its higher sensitivity to volume scattering than co-polarized data. Similarly, other polarization combinations that emphasize the volume scattering return over surface scattering components (such as the circular polarization combination of LR) would be equally appropriate for the algorithm. If only co-polarized data are available, however, then it is generally preferable to have HH-polarization over VV, and then to move forward with the FSH algorithm, with the expectation that both accuracy and sensitivity will be reduced.

The *link file* mentioned above provides information on which files are expected to have some degree of geographic overlap and hence be used in propagating the coefficients of FSH. While many files may have such a geographic overlap—and that, indeed, this overlap can be automatically calculated—a separate link file is desired so that links can be added or broken as necessary in order to account for the varying quality of data in the overlap region used to estimate the coefficients (e.g., a scene with a particularly high degree of temporal decorrelation can be removed from the link list). A simple example of the text-formatted link file is as follows:

```
2 1
2 3
```

This indicates that image 2 is connected to image 1, and that image 2 is also connected to image 3 (and also that images 1 and 3 are not connected).

In this context, the high degree of temporal decorrelation referenced in the previous paragraph indicates those situations in which the temporal decorrelation is large enough to obliterate any information content in the repeat-pass interferogram. Such is the case when the average interferometric correlation magnitude for a scene falls in the range of 0 to 0.5.

Once these files are created and put into place, the FSH set of scripts can be run by calling it in the command line and passing arguments that indicate the

various input file names as well as ancillary information. An example of a call to the FSH algorithm call is

```
python forest_stand_height.py <# scenes> <# edges> <start scene #> <# iterations> <link filename> <flag filename> <lidar heights file> <forest/non-forest file> <directory of input/output files> <list of output formats>
--flag_proc=0
```

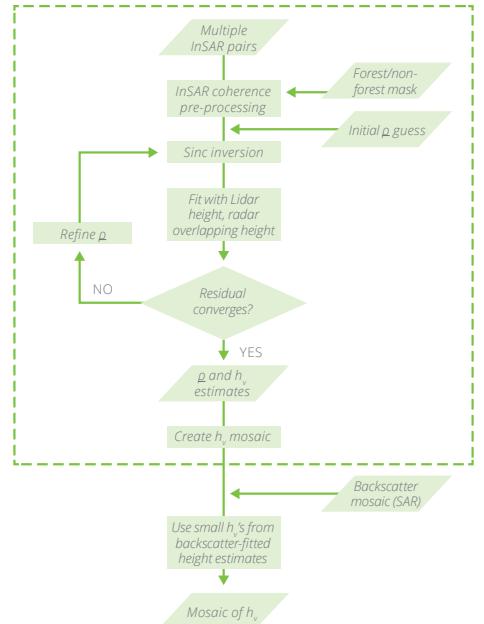
In the last line of the FSH algorithm call, the list of output formats should be in quotes, and can contain one or all of the following: “***tif kml gif mat json***”. In other words, output formats can be created for any of these options. Further, the command option **--flag\_proc 0** indicates that the input data has been processed into SLCs by the ROI\_PAC algorithm (as opposed to processing by ISCE, which should have a value of 1 instead).

#### 4.1.3 ALGORITHM DEVELOPMENT

The FSH algorithm described in **Sections 4.1.1 and 4.1.2** are based on a combination of empirical relationships between cross-polarized radar backscatter power and interferometric coherence. Through the development of the algorithm, and following analysis such as that shown in **Figure 4.1**, it has been determined that FSH values below 10 m should be determined by the backscatter power relationship, and values above this threshold should be determined by interferometric coherence. In order to determine if this threshold has been met, the interferometric coherence version of FSH is first computed, and in regions where that is determined to be below the threshold value, the backscatter power empirical relationship is used.

A block diagram for this approach is given in **Figure 4.8**. In the diagram, parallelograms refer to inputs and outputs of the algorithm. Rectangles are steps in the processing, and a diamond is a point of evaluation. Also, in the diagram, the variable  $h_v$  refers to the value of FSH, and  $\rho = [S_{\text{scene}} \ C_{\text{scene}}]$  is the set of two values per scene that parameterize the model that relates temporal decorrelation to the vegetation height (**Sec. 4.1.3.2**) (see Lei et al. 2019).

In order to gain some appreciation of the simplicity of the relationships described above, it is valuable to specify what these equations are. A more detailed ex-



**Figure 4.8** Block diagram for the processing of FSH.

planation of this approach, complete with equations and a statistical examination, can be found in Lei et al. (2019).

##### 4.1.3.1 Relationship of Backscatter to Forest Stand Height

The backscatter power, after correcting for topographic and other geometric effects, is written as

$$\gamma^0 = A \left( 1 - e^{-Bh_v} \right), \quad (4.1)$$

where  $\gamma^0$  is the terrain-corrected form of radar cross section (e.g., see Small 2011),  $h_v$  is the vegetation height, and the coefficients  $A$ ,  $B$ , and  $C$  are determined in the FSH algorithm using a least-squares fit between the backscatter power and the vegetation height provided by the ground validation and/or overlap data between scenes. Sample values for these coefficients that have been automatically determined by the FSH algorithm are  $A = 0.11$ ,  $B = 0.0622$ , and  $C = 1.0143$ .

A common issue with the relationship of backscatter to vegetation characteristics is that above a certain threshold of biomass, there is no longer a sensitivity of increasing  $\gamma^0$  to increasing biomass. This saturation effect is wavelength-dependent. At L-band, an accepted value for the saturation limit is for 100 tons of biomass/hectare. Under the assumption that a rela-

tionship exists between vegetation height and biomass, whatever that may be, there is a similar saturation effect that occurs between  $\rho^0$  and  $h_v$ . This is reflected in the exponential relationship shown at the beginning of this section. While the saturation limits of sensitivity of  $\rho^0$  to  $h_v$  are less well-characterized, nominal estimates for these values provided in **Table 4.1**.

Note that values listed in the table are nominal values only and are strongly dependent on the biome, stand age, soil moisture and surface roughness.

When working in a specific region, however, the *correct approach* for determining these limits is to make a plot of radar backscatter as a function of lidar-derived vegetation heights.

In this section, backscatter power is referred to as  $\rho^0$ . This backscatter power is a measure of the power that the radar receives from a particular region on the Earth's surface that is reflecting energy back to the radar. These values are stored on the satellite or airborne platform in digital values that are related to the power recorded by the radar. After processing to put the data into ground coordinates, and to perform aperture synthesis (a critical part of SAR processing), these values are transformed by the processor and provided to the user either as Digital Numbers (DN values) or in terms of calibrated radar backscatter power, either in units of  $\rho^0$  or  $\rho^0$ , depending on the level of processing employed. The term "calibration" refers to correcting the radar power returns for gains that are internal to the radar system and processing chain and making all measurements proportional to the transmitted power. Values of  $\rho^0$  are calibrated in terms of the range coordinate of the radar system and have been normalized for the size of the area reflecting the energy back to the system (larger areas will reflect more energy). The units of  $\rho^0$  are in  $\text{m}^{-2}$ . The radar cross section  $\sigma$  is not normalized for this area and is in units of  $\text{m}^2$ . When a DEM is used and the value of the radar cross section is adjusted to account for the intercepted surface area in the direction of radar viewing, this is what is termed  $\rho^0$  and is the form of radar cross section most appropriate for quantitative analysis (Small 2011).

#### 4.1.3.2 Relationship of Interferometric Coherence to Forest Stand Height

The interferometric coherence is derived from

BAND	WAVELENGTH	FSH	HV BACKSCATTER SATURATION LEVEL
X- (10 GHz)	3 cm	10 cm	-10 dB
C- (5.4 GHz)	5.6 cm	1 m	-12 dB
L- (1.2 GHz)	24 cm	10 m	-13 dB

**Table 4.1** Nominal estimates for HV backscatter saturation levels for typical SAR wavelengths.

the interferometric correlation, which is the normalized geometric average between two complex images. Mathematically, the interferometric correlation  $\gamma$  is defined as

$$\gamma = \frac{\langle E_1 E_2^* \rangle}{\sqrt{\langle |E_1|^2 \rangle \langle |E_2|^2 \rangle}}, \quad (4.2)$$

where  $E_1$  and  $E_2$  are the complex values of radar cross sections observed by the SAR satellite and delivered as SLCs, the brackets indicate averaging over multiple looks, and \* indicates a complex conjugation. Note that the  $\gamma$  defined in the interferometric correlation expression above is not the same as the  $\rho^0$  specified for the terrain-corrected value of radar cross section described in **Section 4.1.3.1**. When an image is referred to as an interferogram, it indicates an image of  $\gamma$  as specified previously. This correlation is complex-valued, with its magnitude (the coherence) varying between 0 and 1, and the phase between 0 and  $2\pi$ . A signal with low correlation will have a coherence close to 0 and a random phase. A signal with a high correlation will have a coherence close to 1 and a well-determined phase that is related to the viewing geometry.

A number of factors contribute to the general value of the interferometric correlation:

- The geometric correlation due to incidence angle and projection effects,  $\gamma_{\text{geom}}$
- The correlation related the proportion of noise in the receive system,  $\gamma_{\text{SNR}}$
- The correlation related to the interferometric baseline and the volume scattering of the target,  $\gamma_{\text{vol}}$
- The temporal correlation (or decorrelation, as the case may be),  $\gamma_{\text{temp}}$

The net effect of all of these sources of decorrelation multiplied by one another make up the total observed correlation  $\gamma$  as described previously by

Eq. (4.2):

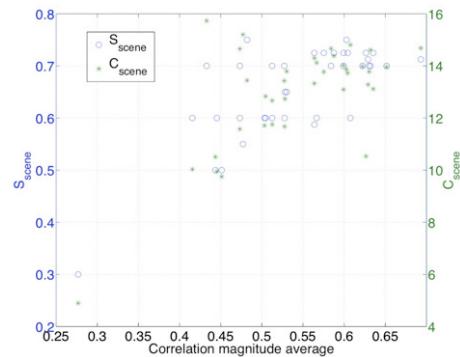
$$\gamma = \gamma_{\text{geom}} \cdot \gamma_{\text{SNR}} \cdot \gamma_{\text{vol}} \cdot \gamma_{\text{temp}}. \quad (4.3)$$

When a satellite doing repeat-pass interferometry and has an orbital repeat that minimizes the orbital distance between repeat-orbits, the condition exists known as "zero-baseline interferometry," which is the case for most repeat-pass SAR systems. In such cases, the contribution of the volumetric decorrelation  $\gamma_{\text{vol}}$  to the total correlation is minimal; hence, the best way for relating interferometric correlation to FSH is through the temporal decorrelation signature, which is a statistical-empirical relationship by its nature.

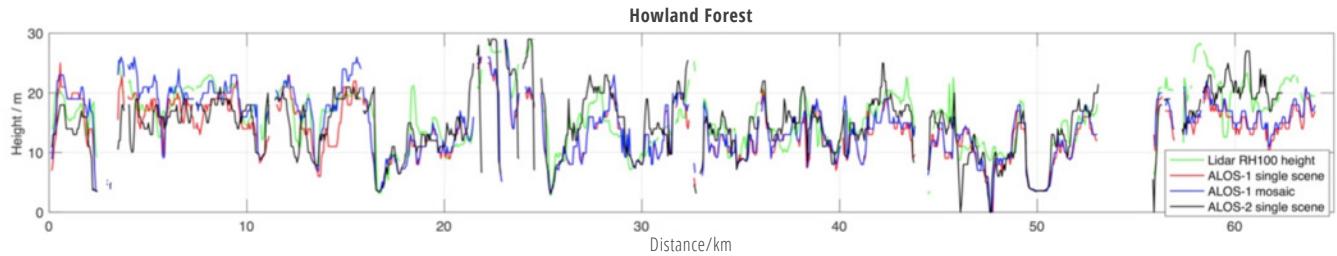
In the FSH algorithm, the combination of volume and temporal correlation (or coherence),  $|\gamma_{\text{v\&t}}| = |\gamma_{\text{vol}} \gamma_{\text{temp}}|$ , is related to the vegetation height  $h_v$  by the empirical equation:

$$|\gamma_{\text{v\&t}}| = S_{\text{scene}} \cdot \text{sinc}\left(\frac{h_v}{C_{\text{scene}}}\right), \quad (4.4)$$

where the coefficients of  $S_{\text{scene}}$  and  $C_{\text{scene}}$  are scene-wide coefficients (i.e., have only one value for the entire radar scene) determined using a least-squares fit to the ground validation data and/or overlap regions between neighboring interferograms (e.g., Lei et al. 2019). Typical



**Figure 4.9** Typical values for the model coefficients of  $S_{\text{scene}}$  and  $C_{\text{scene}}$  used by the FSH algorithm for relating vegetation height to the interferometric coherence.



**Figure 4.10** A spatial comparison between lidar-derived tree height (RH100) from NASA’s LVIS instrument, and the FSH approach using either the ALOS-1 or -2 sensors.

values of  $S_{\text{scene}}$  and  $C_{\text{scene}}$  as determined in a 37-scene, statewide mosaic of FSH for Maine are shown in **Figure 4.9**. Note that values of  $S_{\text{scene}}$  and  $C_{\text{scene}}$  vary due to differing weather and soil moisture conditions that happen throughout the year and observing period of repeat-pass interferometry.

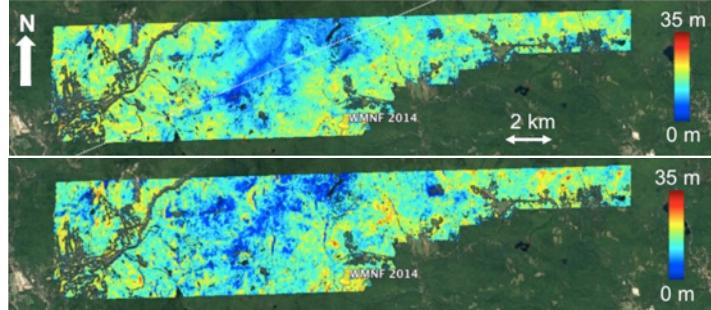
Once the coefficients for the empirical relationship between  $h_v$  and  $|\gamma_{v\theta}|$  have been established, it is a simple matter to invert the relationship (using a lookup table or otherwise) to determine FSH over an extended region.

#### 4.1.4 ACCURACY OF FINAL MEASUREMENTS

The accuracy of the estimates of FSH obtained using the methods described above is a subject of continued study. One example of the accuracy assessment is shown in **Figure 4.1**, which shows values for the Root-Mean-Square Error (RMSE) (residual) error of estimating FSH when compared to lidar data. In these cases, the error for FSH is 3.8 m when measured at a resolution of 400 x 800 m (32 ha) and using the interferometric correlation alone (i.e., not including the estimation improvement when backscatter power is used to estimate FSH for values of FSH < 10 m). When data from the coherence are combined with the backscatter power, the estimated error is better than 3.5 m when measured at a resolution of 6 ha, an improvement of more than four times.

Factors that affect the accuracy of the FSH algorithm are:

- The degree that temporal conditions affect the interferometric coherence
- The availability of SAR data at wavelengths of L-band (or P-band; C-band data from Sentinel-1a for instance, is not appropriate for FSH determination using these methods)
- Availability and quality of ground validation data



**Figure 4.11** A qualitative comparison of (a) lidar-derived vegetation height from the GRANIT sensor and (b) SAR-derived FSH for a site that is more than 300 km away from the location where the LVIS lidar was used for determining the model coefficients. The 6-ha RMSE between the two measures of FSH is 3.9 m (from Lei et al. 2019).

that can be used for determining model coefficients

- The spatial dimension (area) that the accuracy is being assessed.

With respect to this last parameter that affects accuracy, for many remote sensing applications, so long as there are no biases in the data, resolution can be traded for accuracy. In the case of the FSH algorithm, the accuracy is quoted to be 3.5 m at a 6-ha resolution. To determine the accuracy of the algorithm at a 1-ha resolution, the reporting requirement for REDD+ MRV (see **Section 4.1.7**), the extrapolated accuracy would be  $\sqrt{6 \text{ ha}/1 \text{ ha}} \times 3.5 \text{ m} = 8.6 \text{ m}$ .

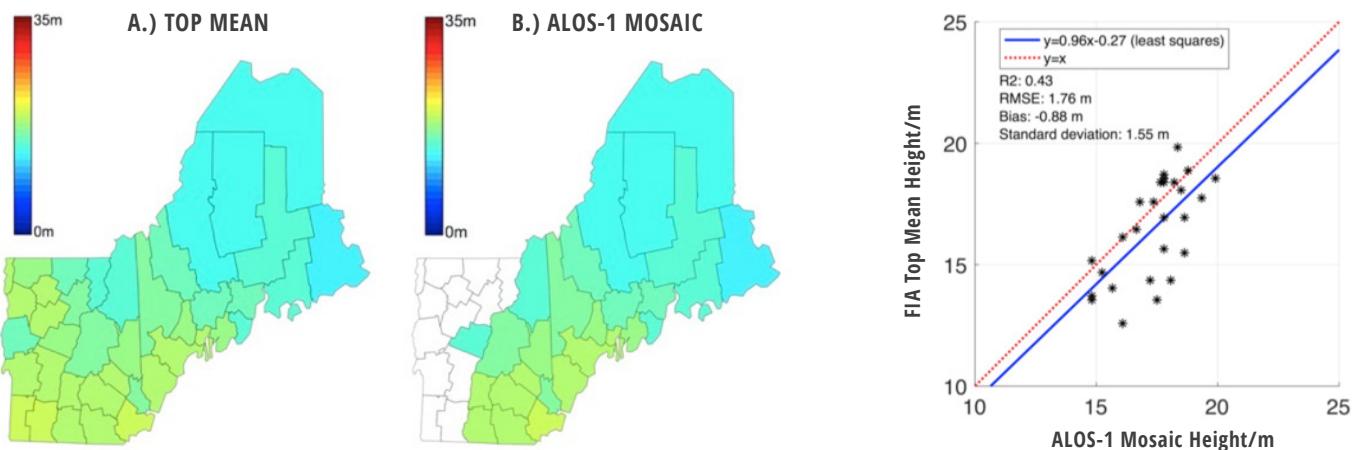
An example of a wide-area application of the FSH algorithm can be found in Lei et al. 2019, with some of the salient results shown in **Figure 4.10**.

A simple method of assessment is to show a spatial comparison between lidar-derived heights and those obtained from the FSH algorithm. For a transect extracted from the LVIS data shown in **Figure 4.4** over the Howland forest in Maine, a comparison is made in **Figure 4.9** between the lidar-derived height and the height determined from the InSAR and SAR backscat-

ter power algorithm discussed here. The plot shows excellent agreement between the two measures, but may be unsurprising in that the lidar data were used to calibrate the scene-wide coefficients used by the SAR/InSAR FSH algorithm for estimating height.

A better comparison can be assessed by finding a nearby region that is also sampled by lidar but not used in determining the model coefficients. Such a site exists in the White Mountain National Forest (WMNF) in eastern New Hampshire, U.S., more than 300 km away and distant from the originally trained SAR/InSAR scene by 5 orbits (and equivalently, at least 5 scenes). By using the overlap regions between adjacent passes of the satellite, the coefficients determined from the Howland Forest can be propagated to the WMNF scenes and then compared to the lidar data that are available there. This is shown in **Figure 4.11** in a qualitative sense. Quantitatively, the residual differences between the two datasets have a standard deviation of 3.9 m when measured at a resolution of 6 ha.

A final comparison can be made for forest heights assessed at the county level, as shown in **Figure 4.12**. In this case, data from the U.S. Forest Service’s



**Figure 4.12** Comparison of county-level assessments of vegetation height obtained (a) by the U.S. Forest Service using FIA plots and (b) those obtained using the FSH algorithm. At right is a quantitative comparison between the two datasets.

Field Inventory Analysis (FIA) program is used for creating an independent assessment of forest height for each county in the region (see Lei et al. 2019). When these county-level estimates are compared with the SAR/InSAR estimates of FSH at the same scale, the RMSE is measured to be 1.8 m, as shown in the figure.

In general, as shown in **Figure 4.11**, the differences between independently derived measures of forest height and those determined using the SAR/InSAR algorithm for FSH compare very well and have residual errors on the order of 4 m for map resolutions of 6 ha. Under the assumption that the independently derived estimates of FSH are more accurate than the SAR/InSAR approach, the dominant contributor to this residual error is due to *model error* related to the difficulty in capturing the effect of weather events on the InSAR signature. One way to overcome this type of error is for the repeat-pass observations to take place over shorter timescales than the 46-day repeat period of ALOS-1. Initial studies using ALOS-2 data (that has a 14-day repeat period) have shown that, indeed, the error is reduced. The observing plan and data distribution policies of ALOS-2, however, have not enabled a fuller assessment that could be applied over a region as large as that shown in **Figure 4.4**, and so opportune data sets where the algorithm can be further tested remain to be found.

#### 4.1.5 SOURCES OF ERROR

After presenting the SAR/InSAR algorithm for deter-

mining FSH in **Sections 4.1.1–4.1.4**, it is important to summarize the different sources of error that can confound this measurement. These sources of error are:

- **Spatially varying degree of temporal decorrelation**—The empirical models that relate SAR backscatter power and InSAR coherence to FSH described in **Section 4.1.3** rely on coefficients that are determined on a scene-wide basis (one radar scene or interferometric pair at a time). When weather affects the temporal signature on the radar imagery in a spatially varying manner within a single scene, then the scene-wide coefficients determined for the model, while correct in an average sense, will have a spatially varying error within the scene. This error can be improved by fitting the empirical models to the residual spatial variation. Such a fit would depend on the availability of ancillary data (e.g., lidar or ground validation) and would require a considerable amount of care during the fitting stage; hence, this is generally not done. A simpler approach to dealing with this type of error source would be to discard the data that suffer from this effect and substitute with data collected during a different time period.
- **Regions that are undergoing significant landcover change**—The InSAR component of the FSH algorithm relies on the temporal decorrelation signature to estimate vegetation height. When temporal decorrelation is due to causes other than the motion of vegetation proportion to
- their height, an error in the estimation of FSH will occur. An example of such error can occur in agricultural regions, where the degree of change in the landcover and field management is high. Such locations show a high degree of temporal decorrelation and hence will be evaluated by the FSH algorithm as having tall forest stands. Similarly, urban areas and regions of open water and flooded areas will also display high degrees of temporal decorrelation that will cause difficulties for the FSH algorithm. A simple approach to dealing with this type of error source is to use a landcover classification converted to an FNF map that eliminates these regions from the estimation process.
- **Regions undergoing selective logging and clearcutting**—Similar to the error sources indicated above, regions undergoing selective logging and clearcutting will display a high degree of temporal decorrelation, and the estimation process will indicate unrealistically large values of FSH (40 m and taller in regions where such tree heights are not common). In these cases, an additional post-estimation step should be exercised to identify all of those regions estimated to have a large value of FSH by the algorithm, evaluate them independently to determine the cause (using optical data or otherwise), and flag the regions as being disturbed.
- **Topographic effects**—The InSAR portion of the FSH algorithm works best when the interfer-

ometric baseline is as close to zero as possible. For this reason, the algorithm is not subject to significant topographic relief. In regions of large topographic variation, the degree of layover and shadow is spatially varying and hence should be accounted for in the assessment. One way of correcting for topographic effects is to collect the data from different aspect angles, such as can be done between ascending and descending passes of the satellite. Once an evaluation has been made as to which regions have errors associated with the viewing geometry, these errors can be minimized by combining the results from the different orbital directions of the satellite.

#### 4.1.6 COMBINATION WITH OPTICAL DATASETS

The SAR/InSAR method for estimating FSH lends itself very well to combining with optical datasets, whether active (lidar) or passive (e.g., Landsat, MODIS and Sentinel-2). In general, both serve important roles in the estimation of FSH. As explained in [Section 4.1.2](#), lidar is important in the determination of model coefficients, and optical data are often used for creating landcover classification products to derive forest cover maps. These maps are then used to determine regions where the FSH algorithm should be exercised.

After calculating FSH using the algorithms detailed here, optical data (especially lidar (as shown in [Fig. 4.11](#)) can serve the role of validation, an important component of the MRV system necessary for monitoring natural resources within a county's borders and meeting various United Nations agreements with developing countries.

#### 4.1.7 MRV SYSTEMS IN THE CONTEXT OF REDD+

The United Nations Framework Convention on Climate Change (UNFCCC) describes the need for Measurement, Report, and Verification (MRV) of forest carbon stocks, implemented through the Reducing Emissions from Deforestation and Forest Degradation in Developing Countries (REDD+) program. This program seeks methods for independently verifying the status and change of carbon stocks within developing countries, especially as they undergo varying economic, population, and climate challenges.

The methods described here, especially as demonstrated in [Figure 4.11](#), can be used for addressing this MRV need. Although in the context of this treatment, the methods have been demonstrated for an 11.6-million-ha region in the northeastern U.S., the same methods can be applied elsewhere in the world. What is required to achieve this reach are (1) the availability of repeat-pass L- or P-band SAR data, (2) an assessment of the regions where the FSH algorithms would be applied (e.g., a global/regional landcover map), and (3) the collection of ground validation or lidar data over regions near where the estimation of FSH would be applied.

#### 4.1.8 LOOKING AHEAD

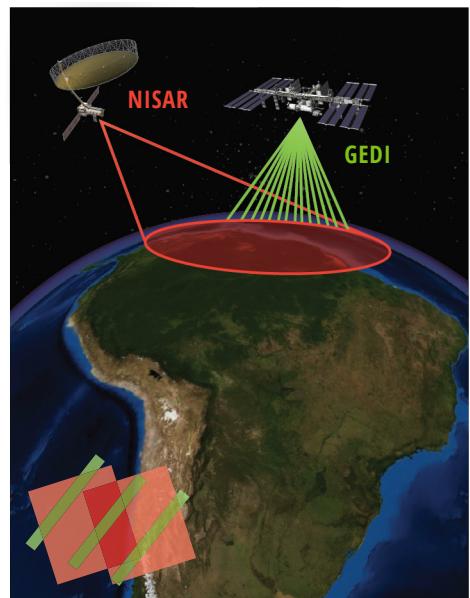
In recent years, the availability of spaceborne remote sensing data—both in terms of data distribution policy and collection of the data—has been expanding considerably. Along with this expanded availability has been an increasing need to apply these assets to better monitor natural resources on a global basis. Such monitoring is important for understanding the effects of climate, public policy, and population pressure on a changing environment.

Through the launching of NASA's [GEDI](#) and [NISAR](#) missions in 2018 and 2021, respectively, the monitoring of forest structure and FSH through the approach discussed here is well-positioned to address these needs.

[Figure 4.13](#) illustrates how this can be accomplished using the side-looking mapping capability of NISAR and the nadir-looking sampling measures of vegetation height that will come from GEDI. The figure also shows how GEDI's 14-beam lidar samples will overlap the NISAR data, which will have a 250-km swath, a 12-day repeat period, and operate at L-band.

The baseline NISAR mission will create interferograms over most of the Earth's landcover surface every 12 days at dual-polarization. In this scenario, the cross-polarized (HV) interferometric coherence and backscatter power from NISAR will be compared with forest heights measured from GEDI and used to calculate the coefficients that parameterize the empirical equations described in [Section 4.1.3](#). Even though the two missions may not be operating concurrently, the degree of change in the world forests will not be so large as to adversely affect the model parameterization.

Prior to the availability of data from these two mis-



**Figure 4.13** Illustration on how NISAR and GEDI data can be combined to create a global estimate of FSH using the algorithms described here. Viewing geometries of NISAR and GEDI are displayed, along with an inset overlap schematic of NISAR data (red) with GEDI 14-beam lidar data (green) (Lei et al., 2019).

sions, there are in principal sufficient resources from JAXA's ALOS-1 and -2 satellites as well as CONAE's SAO-COM satellite that can be combined with airborne lidar for obtaining results similar to those presented here and in published papers. The largest caveat at present is the availability of L-band SAR data, which is fairly restricted due to governmental policies, especially in the distribution of raw data. The larger scientific community, consisting of ecosystem and other Earth scientists, have been lobbying the governmental agencies of Japan and Argentina to free up some of these resources, however, and hence there is hope that some of these data will become more available, especially over the countries where the assessment and monitoring of forest resources with remote sensing data are critically important.

## 4.2 Python Scripts

A GitHub website with Python scripts written by Y. Lei, the principal developer of the FSH technique, has been set up. These scripts can be freely downloaded, along with an example-driven tutorial on the process, at <https://github.com/leiyangleon/FSH>.

## 4.3 References

- Cloude, S. R. R., & Papathanassiou, K. P. P. (2001). Single-baseline polarimetric SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 39(11), 2352–2363. <http://doi.org/10.1109/36.964971>
- Lei, Y., & Siqueira, P. (2014). Estimation of forest height using spaceborne repeat-pass L-band InSAR correlation magnitude over the US state of Maine. *Remote Sensing*, 6(11), 10252–10285. <http://doi.org/10.3390/rs61110252>
- Lei, Y., & Siqueira, P. (2015). An automatic mosaicking algorithm for the generation of a large-scale forest height map using spaceborne repeat-pass InSAR correlation magnitude. *Remote Sensing*, 7(5), 5639–5659. <http://doi.org/10.3390/rs70505639>
- Lei, Y., Siqueira, P., Torbick, N., Ducey, M., Chowdhury, D., & Salas, W. (2019). Generation of large-scale moderate-resolution forest height mosaic with spaceborne repeat-pass SAR interferometry and lidar. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2), 770–787. <http://doi.org/10.1109/TGRS.2018.2860590>
- Lei, Y., Lucas, R., Siqueira, P., Schmidt, M., & Treuhaft, R. (2018). Detection of forest disturbance with spaceborne repeat-pass SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 56(4), 2424–2439. <http://doi.org/10.1109/TGRS.2017.2780158>
- Small, D. (2011) "Flattening gamma: Radiometric terrain correction for SAR imagery" *IEEE Transactions on Geoscience and Remote Sensing* 49(8), 3081-3093
- Treuhaft, R. N., Siqueira, P. R. (2000). Vertical structure of vegetated land surfaces from interferometric and polarimetric radar. *Radio Science* 35, 141–177. <https://doi.org/10.1029/1999RS900108>



# APPENDIX C

## Estimating Forest Stand Height Using L-band SAR – Chapter 4 Training Module

Developed by Helen Baldwin and Sarva Pulla with data and scripts from Paul Siqueira and Yang Lei

### **Input datasets:**

- ALOS PALSAR or equivalent L-band dual-polarized imagery
- Forest height data (lidar or ground collection; GeoTIFF)
- A forest/non-forest mask (optional; GeoTIFF)

### **Software:**

- A Unix/Linux environment required to run the Python scripts
- QGIS/ArcGIS/GoogleEarth (suggested for visualization)
- Anaconda for Python and packages

In this tutorial, we will estimate forest stand height (FSH) using L-band SAR data. The most accurate way to estimate FSH with repeat pass interferometry is by using a combination of SAR backscatter power and InSAR coherence. For this reason, the approach described here can be referred to as a combined SAR/InSAR estimation of FSH. Since the backscatter power relationship is most appropriate to calculate values of FSH below 10 m and values above this threshold are best determined by interferometric coherence, this algorithm computes FSH from interferometric coherence first, and the backscatter power empirical relationship is used if the FSH is below that threshold.

## 1 DATA ACQUISITION

One L-band SAR scene and one ancillary dataset are necessary for this tutorial. An additional ancillary dataset is recommended. To download an example dataset, please see section 2.5 of this module.

### 1.1 ALOS PALSAR

Since the structure of vegetation is on the order of 10's of centimeters, forest vegetation is often best ob-

served with P- or L-bands (~24 cm). At this bandwidth, the Japanese Aerospace Exploration Agency (JAXA)'s JERS-1 and ALOS-1 & -2 satellites are available, but geographically limited. This tutorial utilizes ALOS-1. Please refer to Marc Simard's Training Module in Appendix E for an explanation of how to acquire ALOS PALSAR data and select the Single-Look Complex (SLC) product. This tutorial could also potentially apply to NISAR data in the future.

Processing InSAR data to estimate FSH requires either raw satellite data that have been downlinked, but not processed, or SAR data that have been processed into SLC imagery appropriate for forming interferograms. If you have access to SLCs, it is recommended that you skip section 4 and proceed to section 5. If only raw data are available, then the additional processing explained in section 4 of this tutorial is necessary. One advantage of beginning with raw data are that the output formats of the interferograms and ancillary data make it easy to follow on the processing methods with additional steps implemented to estimate FSH.

## 1.2 Ancillary Datasets

FSH ground validation data are an important component of the FSH estimation methodology. There are two types of ancillary data utilized in the algorithm. Locations where forest height has been previously determined are required to train the empirical models. A forest/non-forest mask indicating where the estimates should be calculated is an optional dataset.

### 1.2.1 Forest Height Data

Independent measurements of forest height are necessary to determine values for the empirical models that relate radar backscatter power and interferometric coherence to FSH. Lidar data are preferred, since they acquire accurate measurements of vegetation height over an extended geographic region. Freely-available satellite resources of lidar data are currently or about to become accessible, including ICESAT-1 and -2, and the upcoming NASA GEDI mission. This dataset should be in a GeoTIFF format and resampled to the same resolution as the InSAR image. The margin/NoData values must be set to NaN or some number less than zero. Within the FSH scripts, this data set is referred to as "ref\_file."

If lidar data are not available, then another form of independent forest height needs to be identified or created. A simple method is to perform a land cover classification of a region using optical data sets. Stands of different ages and species composition will have different heights, which can be estimated from the ground to the same accuracy as the FSH. This approach was used during the development and testing of the FSH algorithm with mixed results.

### 1.2.2 Map of Forest/Non-forest

The forest/non-forest map can be derived from a number of sources, or made independently by the user. Examples of sources that can be used to derive a forest/non-forest mask are i. JAXA's FNF mask, ii. the US National LandCover Dataset, and iii. The ESA's CCI Landcover (formerly GlobCover). These datasets are used to identify where forests are situated and, hence, where to estimate FSH. The forest/non-forest mask must be classified so that all regions where FSH should be estimated have a value of zero and all regions where FSH should not be estimated have a value of 1. This optional dataset should be a GeoTIFF and resampled to the same resolution as the InSAR image. This file must be in degrees; e.g., EPSG 4326. The margin/NoData values must be set to NaN or some number less than zero. Within the FSH scripts, this dataset is referred to as "mask\_file".

## 2 LINUX ENVIRONMENT AND PYTHON SETUP

While Python scripts can be run in the Windows, OSX, and Unix environments, the methods in this module require a Unix or Linux environment. Please follow the instructions in section 2.1 to setup a Linux environment on your computer using Oracle VirtualBox, section 2.2 to install Anaconda, and section 2.3 to install dependencies for the FSH scripts. If you already have a Linux environment, or have completed any of the other setup steps, please proceed to the next applicable section.

### 2.1 Download and Install VirtualBox

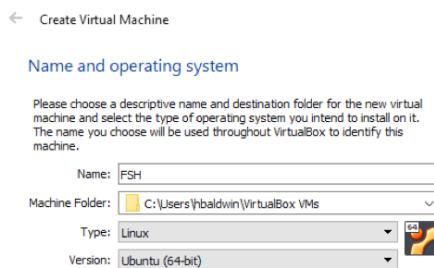
1. First, go to <https://www.virtualbox.org/> to download Oracle VM VirtualBox. Choose the host appropriate for your computer.
2. Next, go to <https://www.ubuntu.com/down>-

load/desktop and download the latest version of Ubuntu. We will use this later on while setting up our virtual machine.

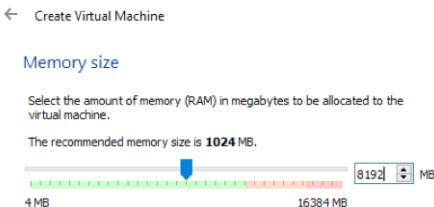
3. Follow along with the Oracle VM VirtualBox installation wizard. Once installation is finished, open the Oracle VM VirtualBox.
4. Click "New" in the menu located at the top of the Oracle VM VirtualBox Manager window to create the virtual machine you will use for this exercise. This menu bar is shown below.



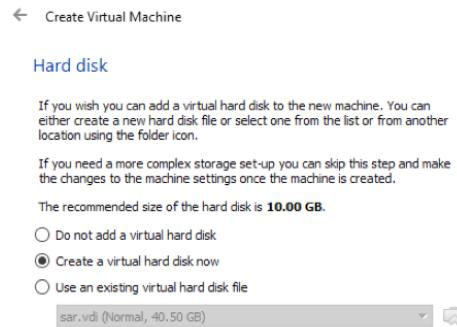
5. The "Create Virtual Machine: Name and operating system" window shown below should pop up. Enter a name for your virtual machine. For this exercise, we chose "FSH". Browse to a folder where you would like to save your machine, select "Linux" from the dropdown menu as the type of machine, and select "Ubuntu (64-bit)" as your version.



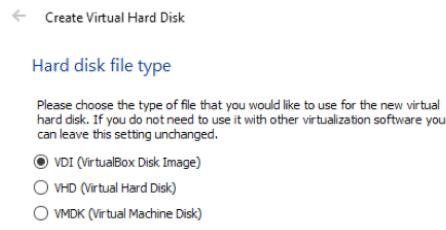
6. Once the name and operating system for your new machine are set up as shown in the image above, click next. The "Create Virtual Machine: Memory size" window should pop up.
7. Enter the amount of memory you would like to allocate to your machine. I chose 8192 MB, as shown below. Click next.



8. Leave the Hard disk selection on "Create a virtual hard disk now" and click create, as shown below.



9. Leave the Hard disk file type selection on "VDI (VirtualBox Disk Image)" and click next, as shown below.



10. Leave the Storage on physical hard disk selection on "Dynamically allocated" and select next, as shown below.



11. Set up the file location and size as shown below. Your file name should automatically populate, but you can also navigate to a new folder to create the file if necessary. I selected 40GB for the virtual hard disk size. Select create, and the "Create Virtual Hard Disk" pop up window will close.

[← Create Virtual Hard Disk](#)

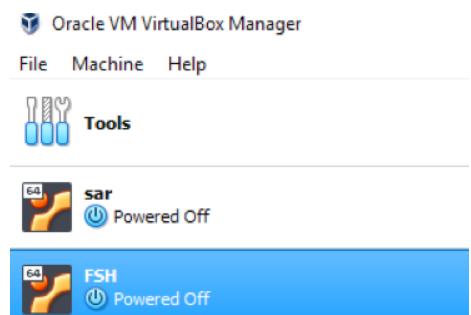
#### File location and size

Please type the name of the new virtual hard disk file into the box below or click on the folder icon to select a different folder to create the file in.

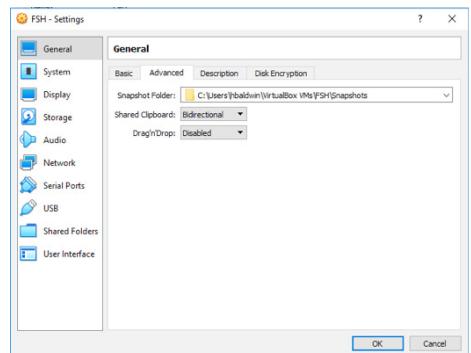
Select the size of the virtual hard disk in megabytes. This size is the limit on the amount of file data that a virtual machine will be able to store on the hard disk.

4.00 MB  2.00 TB

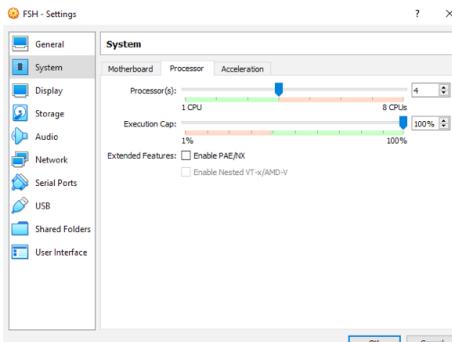
12. Notice that your new virtual machine has been added to the list of virtual machines along the left side of your Oracle VM VirtualBox Manager. As shown below, I have a virtual machine named "sar" along with the virtual machine "FSH" that I just created.



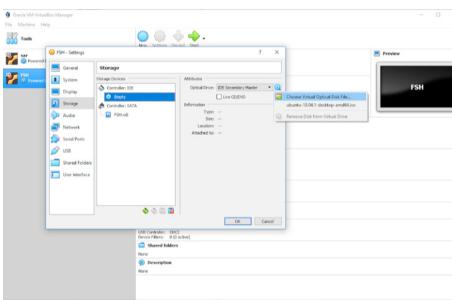
13. Select your new virtual machine from the list. It should appear highlighted, as shown above.
14. Click "Settings" in the menu located at the top of the Oracle VM VirtualBox Manager window to adjust the settings of your new virtual machine.
15. Within the settings pop up window, navigate to the advanced tab.
16. Under "Shared Clipboard," choose "Bidirectional" from the drop down menu. This will allow you to copy and paste between your host system and your virtual machine.



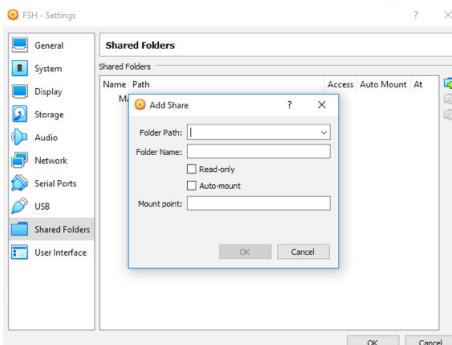
17. Navigate to "System" from the left hand menu. Choose the processor tab. Increase your number of CPUs; I chose 4, as that was the maximum within the suggested green range.



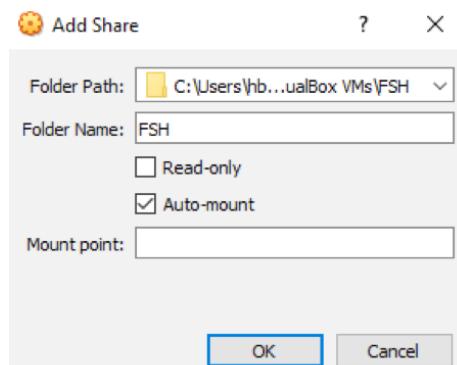
18. Navigate to "Storage" from the left hand menu.  
19. Select the "Empty" disk icon under the Controller IDE option. Under Attributes, click on the disk icon next to the optical drive selection "IDE Secondary Master." Navigate to the Ubuntu for desktop that you downloaded in step 2 using the "Choose Virtual Optical Disk File" option.



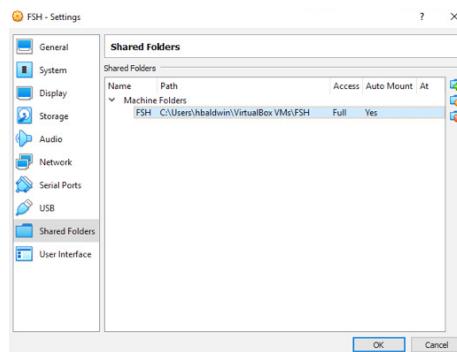
20. Navigate to "Shared Folders" from the left hand menu.  
21. Click the add folder icon along the right of the shared folders window to get to the "Add share" pop up window as shown below.



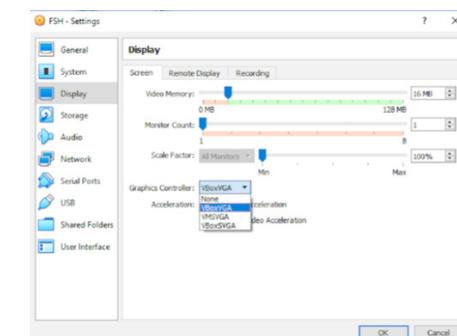
22. Navigate to the folder where your virtual machine is stored within the Folder Path option. The name of the folder will be automatically populated. Choose the "Auto-mount" option as shown below.



23. Click OK to return to the Shared Folders page. Your folder should now appear in the list of Machine Folders as shown below.

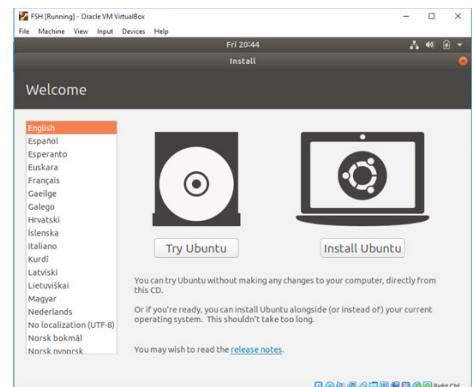


24. To avoid a blank screen after installing Guest Additions in a later step, navigate to "Display" from the left hand menu. Use the drop down menu for the Graphics Controller to select "VBoxVGA."

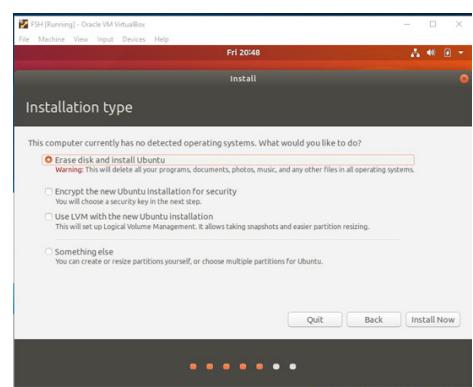


25. Click "OK" to apply these setting changes and return to the Oracle VM VirtualBox Manager. Click "Start" in the menu located at the top of the Oracle VM VirtualBox Manager window to run your new virtual machine.

26. The welcome pop up shown below should appear. Choose your preferred language from the list and click "Install Ubuntu."

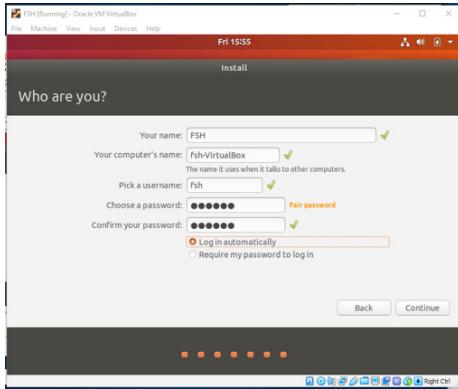


27. Click continue to utilize the default keyboard layout.  
28. Click continue to utilize the default installation and update options.  
29. Click "Install Now" with the default selections as shown below.

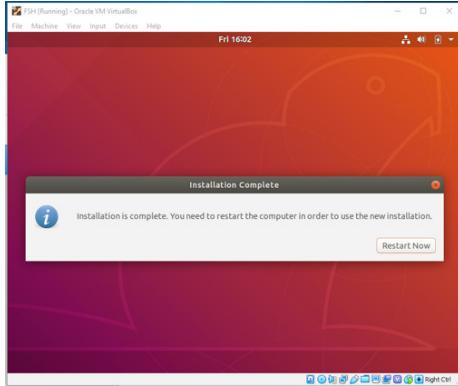


30. Click continue when the pop up window "Write the changes to disk?" appears.  
31. Click continue after selecting your time zone.

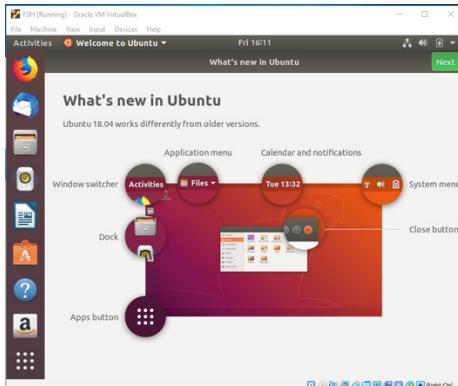
32. Fill in your preferred name and password for your virtual box as shown below.



33. Once installation is complete, the window below should appear. Choose "Restart now" to use the new installation.



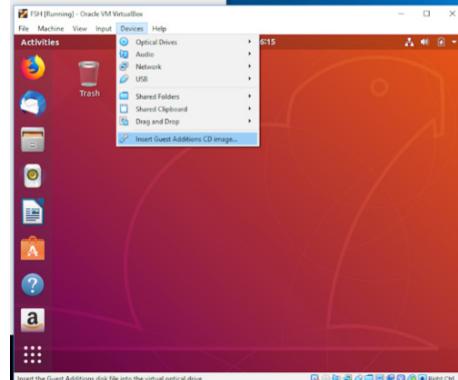
34. After a few minutes, the "What's new in Ubuntu" window shown below should appear. Click next.



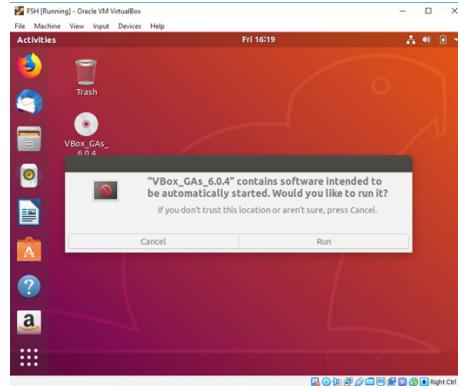
35. Click next to proceed past the Livepatch window.  
36. Click next to proceed past the "Help improve

Ubuntu" window after choosing whether or not to report information to developers for improvement.

37. Click "Done" on the "Ready to go" window.  
38. Click "Devices" in the menu on the top of your running machine and choose "Insert Guest Additions CD Image" from the drop down menu, as shown below.

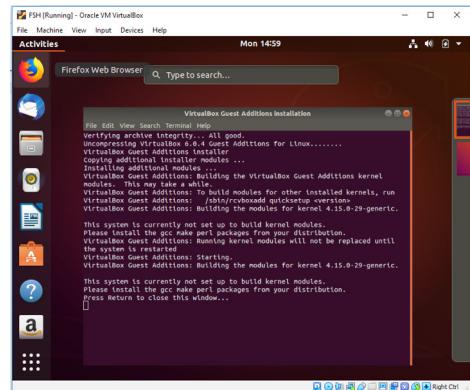


39. The VirtualBox Guest Additions CD (here: VBox\_GAs\_6.0.1) should appear on the desktop of your virtual machine and a warning window may appear as shown below. Click "run" to proceed. You may be prompted to enter in your password to run the Guest Additions disk.



40. Once the Guest Additions disk has finished running, the warning, "This system is currently not set up to build kernel modules" may appear at the end of the messages in the terminal, as shown below. If this is the case, press enter to close the window, and follow steps 40 through 48. If this warning does not appear, you may

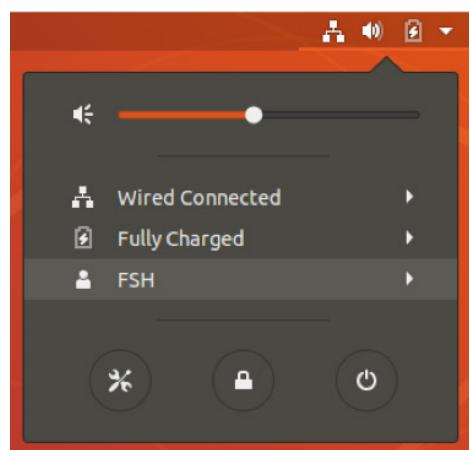
move on to installing Anaconda in section 2.2.



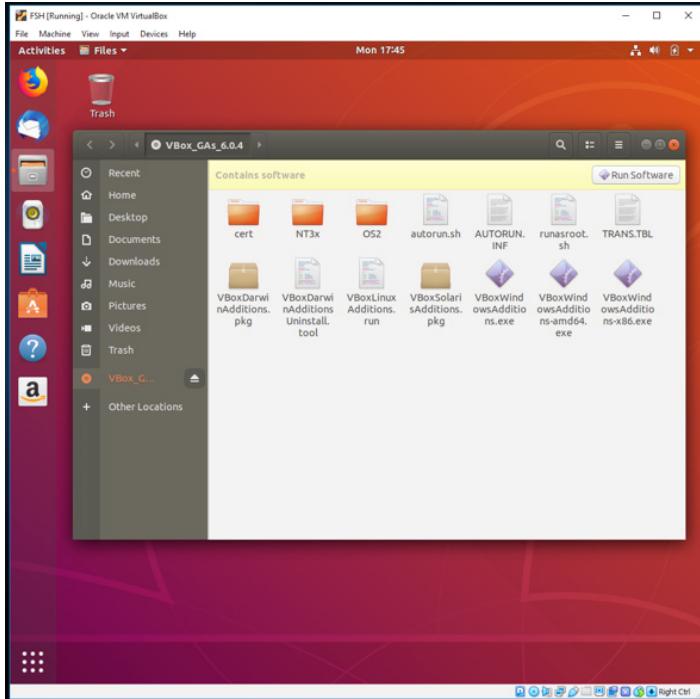
41. Open the terminal using ctr, alt, and t. Then type in the command `sudo apt-get install linux-headers-`uname -r` dkms build-essential` or `sudo apt-get install linux-headers-$(uname -r) dkms build-essential`  
42. You should be prompted to enter "y" to continue. The packages identified as missing in step 39 should now be installed. Press enter to close the window.  
43. In order to use these packages, you will have to restart the virtual machine. Select the arrow along the top right menu (shown below).



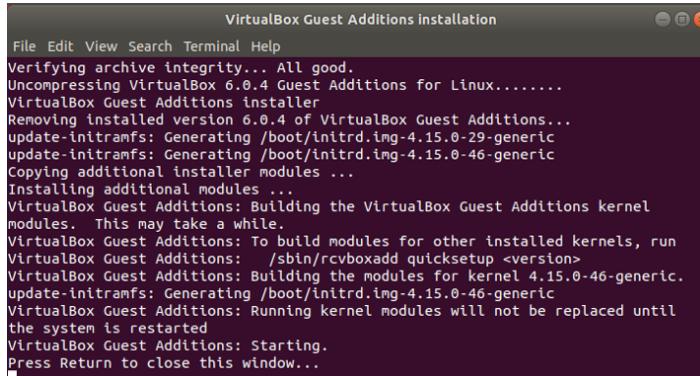
44. An additional menu, shown below, should open.



45. Click on the power icon to open the Power Off window, and choose “Restart”. When the VM restarts, rerun the VirtualBox Guest Additions CD by clicking on the file icon in the menu on the left hand side. Click the Guest Additions disk in the left hand menu on the pop up window. Then select “Run Software”, as shown below.



46. The VirtualBox Guest Additions Installation window should open as shown below. Press enter to close the window.

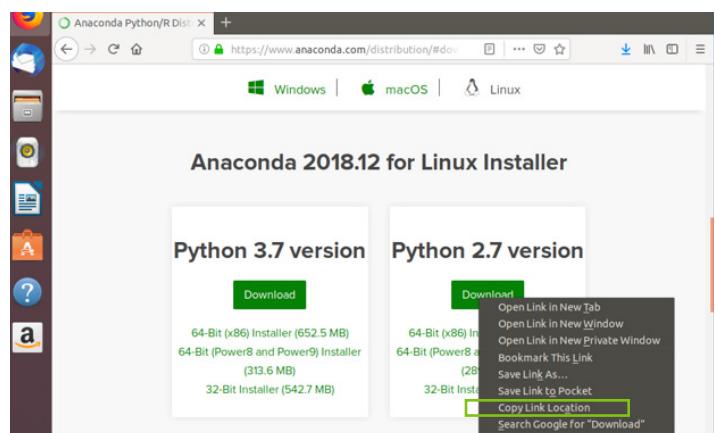


47. Restart the machine as described in steps 42 through 45.

## 2.2 Download and Install Anaconda

48. Open your web browser on your virtual machine, and navigate to <https://www.anaconda.com/distribution/#linux> or search for “install Anaconda.” Make sure to select the tab for the Linux operating system.
49. Right click on the download button for the 2.7 version of Python as shown below, as the FSH scripts were developed and tested using this version. From the menu

that appears, choose “Copy Link Location.”



50. Open the terminal using ctr, alt, and t. Type in the command “wget”, and then paste the location of the download for Python 2.7 version, as shown below.

```
fsh@fsh-VirtualBox:~$ wget https://repo.anaconda.com/archive/Anaconda2-2018.12-Linux-x86_64.sh
```

51. The “Welcome to Anaconda” text should display in your terminal as shown below. Copy the highlighted “Anaconda2-2018.12-Linux-x86\_64.sh.1” text.

```
fsh@fsh-VirtualBox:~$ wget https://repo.anaconda.com/archive/Anaconda2-2018.12-Linux-x86_64.sh
--2019-03-31 12:18:57--  https://repo.anaconda.com/archive/Anaconda2-2018.12-Linux-x86_64.sh
Resolving repo.anaconda.com (repo.anaconda.com)... 104.16.131.3, 104.16.130.3, 2006:4700::6810:8303, ...
Connecting to repo.anaconda.com (repo.anaconda.com)|104.16.131.3|:443...
HTTP request sent, awaiting response... 200 OK
Length: 658699654 (628M) [application/x-sh]
Saving to: 'Anaconda2-2018.12-Linux-x86_64.sh.1'

Anaconda2-2018.12-L 100%[=====] 628.18M 21.1MB/s   in 30s
2019-03-31 12:19:28 (20.8 MB/s) - ‘Anaconda2-2018.12-Linux-x86_64.sh.1’ saved [658699654/658699654]
fsh@fsh-VirtualBox:~$
```

52. Enter the command “bash” and paste in the “Anaconda2-2018.12-Linux-x86\_64.sh.1” text. Follow the prompts to review the Anaconda license information, and enter “yes” to confirm the installation of Anaconda when prompted.

53. Enter the location where you would like Anaconda to be saved. I chose the

```
cryptography
A Python library which exposes cryptographic recipes and primitives.

Do you accept the license terms? [yes|no]
[no] >>>
Please answer 'yes' or 'no':'
>>> yes

Anaconda2 will now be installed into this location:
/home/fsh/anaconda2
```

default as shown on the previous page.

- After the installation is finished, you will be prompted to initialize Anaconda2 in your .bashrc, as shown below. Enter "yes."

```
installing: blaze-0.11.3-py27_0 ...
installing: jupyter_console-5.2.0-py27_1 ...
installing: notebook-5.7.4-py27_0 ...
installing: qtconsole-4.4.3-py27_0 ...
installing: seaborn-0.9.0-py27_0 ...
installing: sphinx-1.8.2-py27_0 ...
installing: spyder-kernels-0.3.0-py27_0 ...
installing: anaconda-navigator-1.9.6-py27_0 ...
installing: anaconda-project-0.8.2-py27_0 ...
installing: jupyterlab_launcher-0.11.2-py27h28b3542_0 ...
installing: numpydoc-0.8.0-py27_0 ...
installing: widgetsnbextension-3.4.2-py27_0 ...
installing: ipywidgets-7.4.2-py27_0 ...
installing: jupyterlab-0.33.11-py27_0 ...
installing: spyder-3.3.2-py27_0 ...
installing: _ipyw_jlab_nb_ext_conf-0.1.0-py27_0 ...
installing: jupyter-1.0.0-py27_7 ...
installing: anaconda-2018.12-py27_0 ...
installing: conda-4.5.12-py27_0 ...
installing: conda-build-3.17.6-py27_0 ...
installation finished.
Do you wish the installer to initialize Anaconda2
in your /home/fsh/.bashrc ? [yes|no]
[no] >>> 
```

- When prompted to proceed with the installation of Microsoft VSCode, as shown below, please enter "no."

```
[no] >>> yes
Initializing Anaconda2 in /home/fsh/.bashrc
A backup will be made to: /home/fsh/.bashrc-anaconda2.bak

For this change to become active, you have to open a new terminal.

Thank you for installing Anaconda2!
=====
Anaconda is partnered with Microsoft! Microsoft VSCode is a streamlined
code editor with support for development operations like debugging, task
running and version control.

To install Visual Studio Code, you will need:
 - Administrator Privileges
 - Internet connectivity

Visual Studio Code License: https://code.visualstudio.com/license

Do you wish to proceed with the installation of Microsoft VSCode? [yes|no]
>>> 
```

- Close your terminal and open a new terminal (ctr, alt, and t) for your installation of Anaconda to become active.

## 2.3 Download and Import Dependencies

- To create a python environment named "sar" where we will store all the dependencies necessary to run the FSH scripts, enter the command "conda create -n sar python=2.7." You can choose to name your environment something other than sar.
- When prompted, enter "y" to proceed with the installation.
- To activate this python environment in the future, use the command "conda activate sar" to enter the environment and "conda deactivate" to leave it. Notice as you use these commands that you will move from "base" to "sar" environments, as shown below.

```
(base) fsh@fsh-VirtualBox:~$ conda activate sar
(sar) fsh@fsh-VirtualBox:~$ conda deactivate
(base) fsh@fsh-VirtualBox:~$ 
```

- Now, let's set up our "sar" environment with the required python packages: NumPy, SciPy, SimPy, json, pillow, OsGeo/GDAL, simplekml, mpmath. Activate the "sar" environment by entering the command "conda activate sar" into the terminal. Install gdal, numpy, pillow, simplekml and scipy by entering the command "conda install -c conda-forge gdal numpy=1.15 pillow simplekml scipy" as shown below.

```
(base) fsh@fsh-VirtualBox:~$ conda activate sar
(sar) fsh@fsh-VirtualBox:~$ conda install -c conda-forge gdal numpy=1.15 pillow
simplekml scipy
```

- When prompted, enter "y" to proceed.
- Enter the command "pip install simpy mpmath" to install additional prerequisites.
- To confirm that you have installed all of the Python packages, you can enter the command "python." Then enter "import gdal" or "import" followed by any of the other packages. If no errors pop up in your terminal and the arrows that indicate a new line appear, then the packages have been installed correctly.
- Enter the command "exit ()" to leave python.
- To view the version and other information about the packages you have installed, in the "sar" environment of the terminal, enter the command "conda list pillow" or "conda list" plus any of the packages, as shown below.

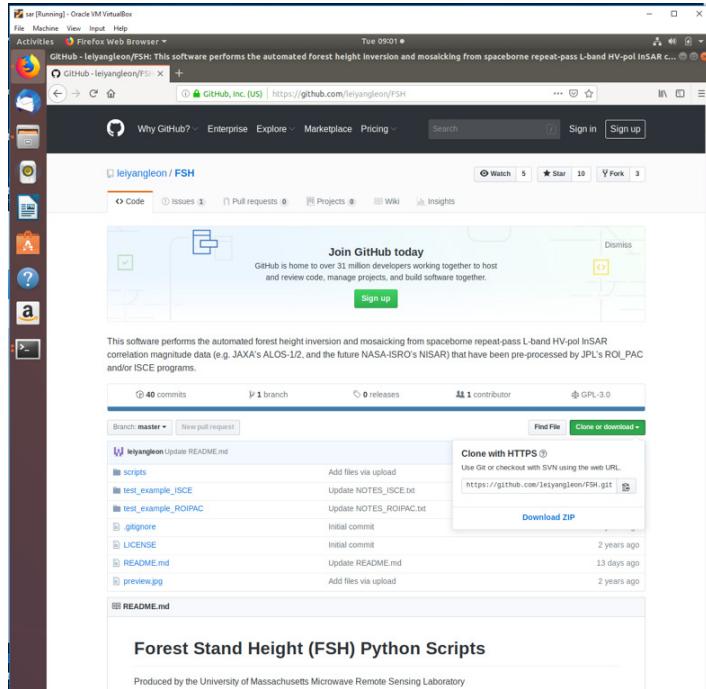
```
(sar) fsh@fsh-VirtualBox:~$ conda list pillow
# packages in environment at /home/fsh/anaconda2/envs/sar:
#
# Name           Version      Build Channel
pillow          5.4.1        py27h00a061d_1000  conda-forge
(sar) fsh@fsh-VirtualBox:~$ 
```

### 3 DOWNLOAD MATERIALS FOR THE TUTORIAL

The Python scripts needed for this tutorial, written by Y. Lei the principal developer of the FSH technique, and an example dataset can be freely downloaded from GitHub or from the SERVIR Global website. The example data are preprocessed, and using these data allow you to skip sections 4 and 5 and proceed to section 6.

#### 3.1 Obtaining the Scripts from GitHub

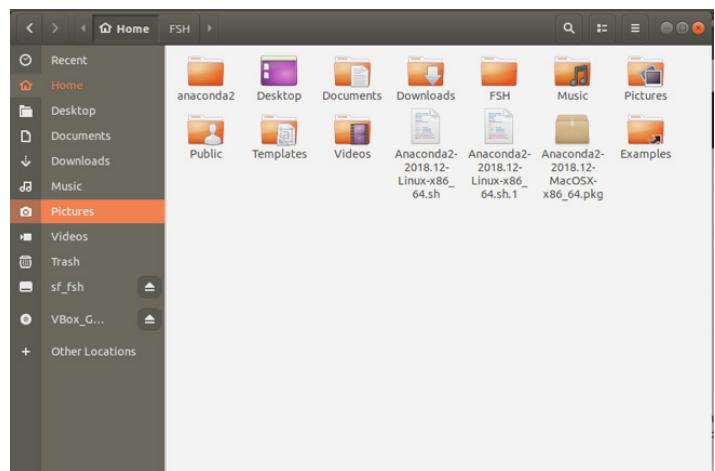
1. Navigate to the GitHub page <https://github.com/leiyangleon/FSH> using FireFox or another internet application on your virtual machine.
2. Click the green “Clone or download” button and copy the link shown under the “Clone with HTTPS” pop up window, as shown below.



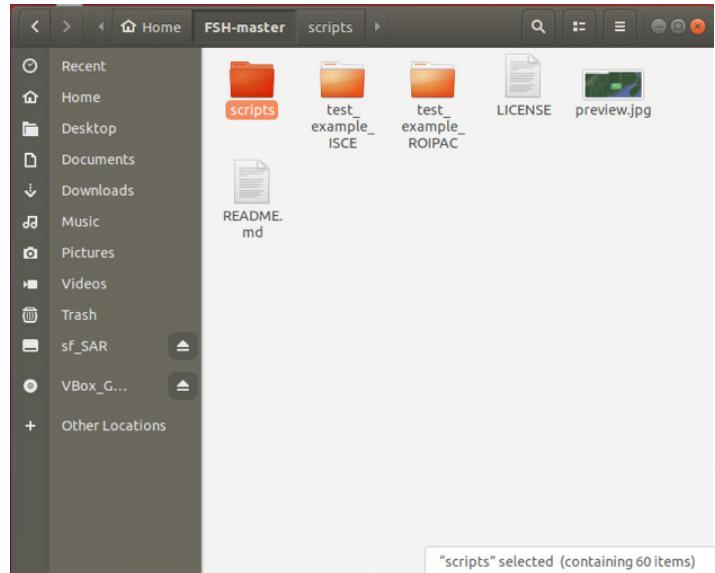
3. Open a terminal, and if you are not already in the “sar” environment created in section 2.3, navigate to the “sar” environment by entering the command “conda activate sar.”
4. Enter the command, “git clone” followed by pasting in the link you copied from the GitHub: <https://github.com/leiyangleon/FSH.git>, as shown below.

```
(sar) fsh@fsh-VirtualBox:~$ git clone https://github.com/leiyangleon/FSH.git
```

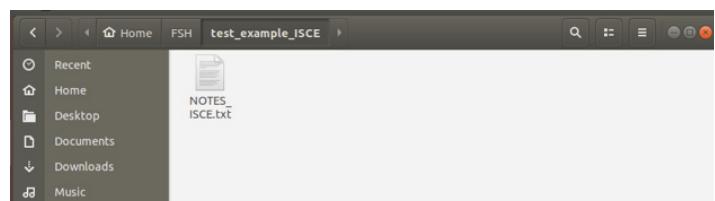
5. If git does not exist on your virtual machine, follow the prompts to install it using the command “sudo apt install git,” followed by your virtual machine’s password.
6. If you navigate to “Home” under the “Files” tab from the menu on the left hand side, you should be able to see the “FSH” folder that you downloaded with all of the scripts necessary for this tutorial.



7. Within your FSH folder, there should be three folders (scripts, test\_exampl\_ISCE, and test\_exampl\_ROIPAC) and three files (LICENSE, preview.jpg, README.md) inside, as shown below.



8. While there are two folders that seem like they should contain data (test\_exampl\_ISCE and test\_exampl\_ROIPAC), if you open these folders than you will find that they only include a text file, and no SAR data or other required files, as shown below.



9. To download the example data, please proceed to the next section (3.2).

## 3.2 Downloading Example Data

The example data consists of three scenes, including a central scene with overlapping NASA LVIS LiDAR data and two adjacent scenes.

1. You can access the link to download the example data by opening the text document within the test\_example\_ROIPAC and test\_example\_ISCE folders respectively. See below for the location of the link within the text file for the ROIPAC data.

```

NOTES_ROIPAC.txt
1. Download the directory "test_example_ROIPAC" from the link:
https://drive.google.com/file/d/0B6s-Z6YH5T12MFhxZzNqjIaUU/view?usp=sharing

2. Under test_example_ROIPAC/, you will find the flag_file ("flagfile.txt"), link_file ("linkfile.txt"), ref_file ("Howland_LVIS_NaN.tif"), mask_file ("Maine_NLCD2011_nonwildland.tif"). All of the associated files for three ALOS PALSAR HV-pol InSAR coherence scenes are grouped by their ALOS ("f$frame_o$orbit") and their acquisition dates (under the subfolder "int_$date1-$date2"). For each scene, there are seven associated files outputted by ROI_PAC: "$date1-$date2_baseline.rsc", "$date1-$date2_2rlks.amp", "geo_$date1-$date2_2rlks.cor", "geo_$date1-$date2_2rlks.int.rsc", "geo_$date1-$date2_2rlks.amp.rsc", "geo_$date1-$date2_2rlks.cor.rsc". This "NOTES_ROIPAC.txt" serves as the instructions for running the FSH software over this test example directory. Finally, a image "ROI_PAC.jpeg" shows the final output of 3-scene mosaic map (GeoTiff format) overlaid on Google Earth in a QGIS window.

***All of the ALOS images files have already been margin-cropped and geocoded in the pre-processing by ROI_PAC (Step 2 & 3 of the general workflow in Section II on the GitHub webpage, i.e. "README.md" file).***

3. Run the 1-command FSH auto-inversion (Step 4 of the general workflow in Section II on the GitHub webpage, i.e. "README.md" file)
python .../forest_stand_height.py 3 2 2 5 "linkfile.txt" "flagfile.txt" "Howland_LVIS_NaN.tif" "Maine_NLCD2011_nonwildland.tif" ".../test_example_ROIPAC/" "tif json kml mat tif" --flag_diff=1 --flag_error=1 --flag_proc=0

4. Run the 1-command FSH mosaicking (Step 5 of the general workflow in Section II on the GitHub webpage, i.e. "README.md" file)
python .../create_mosaic.py ".../test_example_ROIPAC/" "3sc_mosaic.tif" ".../test_example_ROIPAC/f890_0118/b90_118_20070808_HV_20070923_HV_fsh.tif .../test_example_ROIPAC/f890_0119/b90_119_20070710_HV_20071010_HV_fsh.tif .../test_example_ROIPAC/f890_0120/b90_120_20070727_HV_20070911_HV_fsh.tif"

5. Open the final output "3sc_mosaic.tif" in QGIS.

```

2. You can also navigate to this link through the GitHub, following the same folder tree, as shown below.

```

Branch: master | FSH / test_example_ROIPAC / NOTES_ROIPAC.txt

lelyangleon Update NOTES_ROIPAC.txt
1 contributor

15 lines (9 sloc) | 2.13 KB

1. Download the directory "test_example_ROIPAC" from the link:
https://drive.google.com/file/d/0B6s-Z6YH5T12MFhxZzNqjIaUU/view?usp=sharing

2. Under test_example_ROIPAC/, you will find the flag_file ("flagfile.txt"), link_file ("linkfile.txt"), ref_file ("Howland_LVIS_NaN.tif"), mask_file ("Maine_NLCD2011_nonwildland.tif"). All of the associated files for three ALOS PALSAR HV-pol InSAR coherence scenes are grouped by their ALOS ("f$frame_o$orbit") and their acquisition dates (under the subfolder "int_$date1-$date2"). For each scene, there are seven associated files outputted by the ROI_PAC software: "$date1-$date2_baseline.rsc", "$date1-$date2_2rlks.amp.rsc", "$date1-$date2_sim_SIM_2rlks.int.rsc", "$date1-$date2_2rlks.cor", "geo_$date1-$date2_2rlks.amp", "geo_$date1-$date2_2rlks.cor", "geo_$date1-$date2_2rlks.cor.rsc". Finally, "ROI_PAC.jpeg" shows the final output of 3-scene mosaic map (GeoTiff format) overlaid on Google Earth in a QGIS window. Please see below for the file layout.

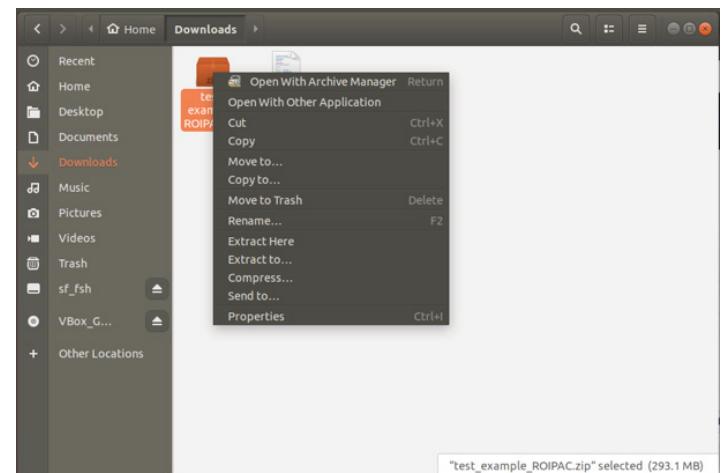
3. Run the 1-command FSH auto-inversion (Step 4 of the general workflow in Section II on the GitHub webpage, i.e. "README.md" file)
python .../forest_stand_height.py 3 2 2 5 "linkfile.txt" "flagfile.txt" "Howland_LVIS_NaN.tif" "Maine_NLCD2011_nonwildland.tif" ".../test_example_ROIPAC/" "tif json kml mat tif" --flag_diff=1 --flag_error=1 --flag_proc=0

4. Run the 1-command FSH mosaicking (Step 5 of the general workflow in Section II on the GitHub webpage, i.e. "README.md" file)
python .../create_mosaic.py ".../test_example_ROIPAC/" "3sc_mosaic.tif" ".../test_example_ROIPAC/f890_0118/b90_118_20070808_HV_20070923_HV_fsh.tif .../test_example_ROIPAC/f890_0119/b90_119_20070710_HV_20071010_HV_fsh.tif .../test_example_ROIPAC/f890_0120/b90_120_20070727_HV_20070911_HV_fsh.tif"

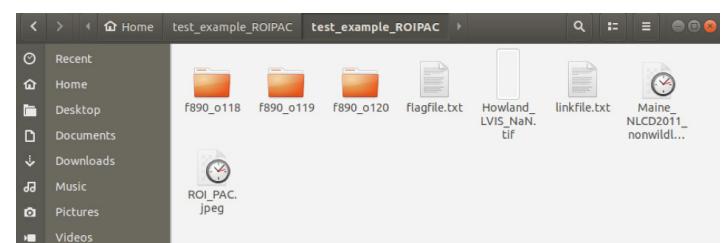
5. Open the final output "3sc_mosaic.tif" in QGIS.

```

3. Either way you choose to find the link to the data, copy and paste this link into the web browser on your virtual machine. While both datasets are compatible with the FSH scripts, we will use the ROI\_PAC as our example for this tutorial.
4. Choose the download icon to download the dataset from the Google Drive link. When prompted for confirmation due to the large size of the file, select "download anyway". When prompted to open the file, choose "save file" and press "OK."
5. Within "Files," navigate to "downloads." Right click on the example data zip, and from the pop up menu, choose "Extract to..."



6. Navigate to "Home" and press the green "Select" button to extract the example data there.
7. Within the "test\_example\_ROIPAC" folder , you will find "flagfile.txt" (referred to as the flag\_file in the scripts),"linkfile.txt" (link\_file ), "Howland\_LVIS\_NaN.tif" (ref\_file), and "Maine\_NLCD2011\_nonwildland.tif" (mask\_file). All of the associated files for the three ALOS PALSAR HV-pol InSAR coherence scenes are grouped by their ALOS ("f\$frame\_o\$orbit") and their acquisition dates (under the subfolder "int\_\$date1-\$date2"). For each scene, there are seven associated files outputted by the ROI\_PAC software: "\$date1-\$date2\_baseline.rsc", "\$date1-\$date2\_2rlks.amp.rsc", "\$date1-\$date2\_sim\_SIM\_2rlks.int.rsc", "\$date1-\$date2\_2rlks.cor", "geo\_\$date1-\$date2\_2rlks.amp", "geo\_\$date1-\$date2\_2rlks.cor", "geo\_\$date1-\$date2\_2rlks.cor.rsc". Finally, "ROI\_PAC.jpeg" shows the final output of 3-scene mosaic map (GeoTiff format) overlaid on Google Earth in a QGIS window. Please see below for the file layout.



## 4 PROCESSING RAW SAR DATA

When processing SAR data, corrections are made for the motion of the satellite and image projection effects that arise from the atmosphere, viewing geometry and topography of the Earth. The steps of processing of ALOS SAR data from raw samples for the satellite include range compression, azimuth compression resulting in an SLC, and finally projection into map coordinates. Software for processing raw data into SLCs can be obtained both commercially and through open-source licensing agreements. Of the open source licensing processors, there are two that have been used for processing raw ALOS data into SLCs and then into estimates of FSH. These are the ROI\_PAC (Repeat Orbit Interferometry PACkage) and ISCE (InSAR Scientific Computing Environment). In this tutorial, we focus on ROI\_PAC as it has completed its development lifetime and is somewhat easier to obtain than ISCE. At the time of this writing, ISCE remains under development. With this in mind, the preprocessing scripts in section 3.2 and the scripts in section 4 that estimate FSH from SLCs have been designed to work with outputs from both ROI\_PAC and ISCE.

### 4.1 Obtaining the Scripts from GitHub

1. Obtain the ROI\_PAC processing software in tgz (gzipped tar) format from: [http://www.openchannelfoundation.org/projects/ROI\\_PAC](http://www.openchannelfoundation.org/projects/ROI_PAC)
2. Download and install a fortran compiler (e.g. gfortran) and the fftw library. See <http://roipac.org/cgi-bin/moin.cgi/installation> for additional details on the installation of ROI\_PAC software.
3. Utilize the test data set that comes with the ROI\_PAC software distribution to test the software installation. You can find the details of how to test the software in the ROI\_PAC installation subdirectory: fullpath/contrib/multtest.sh where fullpath refers to the folder where you unzipped the ROI\_PAC installation archive.

### 4.2 Processing ROI\_PAC/ISCE outputs with Python scripts

1. To open the terminal within your virtual machine, press ctr, alt and t.
2. Crop the ROI\_PAC/ISCE output and eliminate the image margins by running the standalone python scripts CROP\_ROIPAC.py and CROP\_ISCE.py respectively. Please note that the amount cropped is hard coded based on the dimensions of the ALOS SAR image. The code would need to be adjusted for ALOS-2 and future NISAR images.
  - For ROI\_PAC processed results enter the command `python directory_of_scripts/CROP_ROIPAC.py dirname date1 date2`
  - For ISCE-processed results, run the following command within the execution of insarApp.py `python directory_of_scripts/CROP_ISCE.py`

You will need to replace three parameters in these commands:

- Replace directory\_of\_scripts with the location of the ROI\_PAC amp/cor files
- Replace date1 with the date for the 1st SAR acquisition
- Replace date2 with the date for the second SAR acquisition

For information on how to geocode the ROI\_PAC/ISCE output, please see the Chapter 2 training module.

## 5 FILE CREATION & ORGANIZATION

### 5.1 File Structure

The data should be organized in a file structure such that the individual folders hold results from individual interferograms between two dates (the SLCs and ancillary data for individual scene (frame) and orbit (path) numbers). For any one frame and path number, there may exist multiple interferograms, related to multiple repeat-pass combinations of data from two different dates. These interferograms should be stored in sub-directories that have the naming convention: int\_date1\_date2. Scenes from differing frame and paths can be interferometrically processed in order to create an estimate of FSH over an extended geographic region.

The interferogram subdirectories will hold all of the data and information necessary for creating and documenting interferograms made for an observation on two specific dates (date1 and date2). For ROI\_PAC-processed data, the most important file looks like geo\_date1-date2\_2rlks.cor and geo\_date1-date2\_2rlks.rsc. The resource ".rsc" file is a text file that has information the location and size of the geolocated correlation data held in geo\_date1-date2\_2rlks.cor. The format of the correlation file is known as sample-interleaved, or an rmg-format file.

Since radar data are organized in terms of orbits and scenes, in order to make a map of FSH over an extended geographic region it is necessary to mosaic the images. While the process of mosaicking can be done either before or after the estimation of FSH, it is best to do this beforehand to take advantage of the overlap region between images in adjacent paths. In these regions, while the value of the coherence magnitude may vary due to the fact that the observations (and image pairs) have occurred from different orbits (and hence, different dates), the overlap regions can be used to correct for these temporal differences and to adjust the coefficients for the empirical relationships of the SAR products to estimates of FSH.

For each ROI\_PAC-processed scene, the following files should be located in a directory with the format "f\$frame\_o\$orbit/int\_\$date1\_\$date2":

```
$date1_$date2_baseline.rsc  
$date1-$date2.amp.rsc  
$date1-$date2_2rlks.amp.rsc  
$date1-$date2-sim_SIM_2rlks.int.rsc  
geo_$date1-$date2_2rlks.amp  
geo_$date1-$date2_2rlks.cor  
geo_$date1-$date2_2rlks.cor.rsc
```

Please note that the ROI\_PAC's process\_2pass.pl should be run with 2 range looks and 10 azimuth looks in both coherence estimation and multi-looking (equivalent to a 30m-by-30m area for JAXA's ALOS), with the following lines added to the process file:

```
Rlooks_int = 2  
Rlooks_sim = 2
```

```
Rlooks_sml=2
```

```
pixel_ratio=5
```

A 5-point triangle window is hardcoded in ROI\_PAC, which is equivalent to a 2-point rectangle window. For further details on running ROI\_PAC, refer to the ROI\_PAC manual. For each ISCE-processed scene, the following files should be located in a directory with the format “f\$frame\_o\$orbit/int\_\$date1\_\$date2”:

```
isce.log  
resampOnlyImage.amp.geo  
resampOnlyImage.amp.geo.xml  
topophase.cor.geo  
topophase.cor.geo.xml
```

Please note that ISCE’s insarApp.py should be run with 2 range looks and 10 azimuth looks in both coherence estimation and multi-looking (equivalent to a 30m-by-30m area for JAXA’s ALOS), with the following lines added to the process file:

```
<property name="range looks">1</property>  
<property name="azimuth looks">5</property>
```

A 5-point triangle window is hardcoded in ISCE, which is equivalent to a 2-point rectangle window. The .amp/.cor images then need to be multilooked by a factor of two. For further details on running ISCE see the ISCE manual.

The location of the output files depends on whether they are related to the overall processing of the entire dataset, or are directly associated with a single scene. Examples of each would be the SC iteration files as a general output, and a single forest stand height image as a scene-specific output. The general outputs will be stored in a directory named “output” located within the main file directory (file\_directory). The scene specific outputs will be stored with the other scene data as described earlier.

## 5.2 Create Flag File

Once the data have been organized into directories of scenes described by their individual row and path numbers, and the interferograms have been examined to determine which SLC pairs yield the data with the highest coherence (i.e. least amount of temporal decorrelation), there remains the task of creating what is known as a “flag file” and a “link file.” In this context, the flag file is a listing of all the interferograms that will be used in creating the region-wide mosaic of FSH. In the example dataset, there are three such row/path combinations that will create a three-scene mosaic of FSH located in central Maine. The middle of the three scenes overlaps with the forest height data (ref\_file) discussed in Section 1.2, and all scenes are within the region where identifications of forest/non-forest (mask\_file) is used for determining geographic locations where the FSH algorithm will be applied. An example of the contents of a flag file in text format is:

```
001 890_120_20070727_MV_20070911_MV 070727 070911 890 120 MV  
002 890_119_20070710_MV_20071010_MV 070710 071010 890 119 MV  
003 890_118_20070808_MV_20070923_MV 070808 070923 890 118 MV
```

In this example, the first column of numbers indicates the interferogram number, the second is the root file name of the data that forms the interferogram, the third and fourth are the dates that the data were collected for the interferometric pairs, the fifth and sixth are the satellite path and orbit respectively, and the last indicates the polarization of the data.

## 5.3 Create Link File

The link file provides information on which files are expected to have some degree of geographic overlap, and will be used in propagating the coefficients of FSH. While many files may have such a geographic overlap, and that indeed this overlap can be automatically calculated, a separate link file is desired so that links can be added or broken as necessary in order to account for the varying quality of data in the overlap region used to estimate the coefficients (e.g. a scene with a particularly high degree of temporal decorrelation can be removed from the link list). A simple example of the test-formatted link file is:

```
2 1  
2 3
```

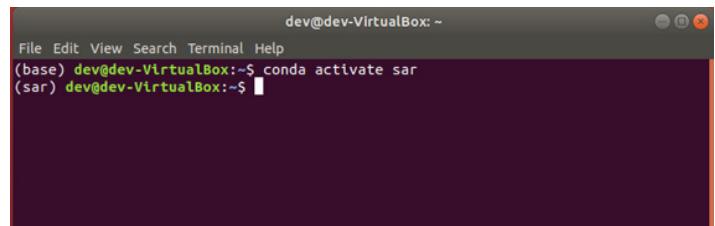
This indicates that image 2 is connected to image 1, and that image 2 is also connected to image 3 (and also that images 1 and 3 are also not connected).

## 6 ESTIMATE FOREST STAND HEIGHT

Once the SLC, forest/non-forest mask, vegetation height, link file, and flag files are created and put into place, you can run the FSH scripts by calling them in the terminal and passing the input file names and ancillary information as arguments. You can run each script one at a time, or call the main script. For this tutorial, we will run the FSH scripts from Anaconda in the virtual machine we set up. All five possible final output data types are produced. Please note that runtime does not increase linearly with each additional scene. Runtime for most of the steps are linear in the number of scenes; however, the core part of the inversion and mosaicking depends on the number of edges, which increases a bit faster as the number of scenes increases.

### 6.1 Access the Anaconda Environment

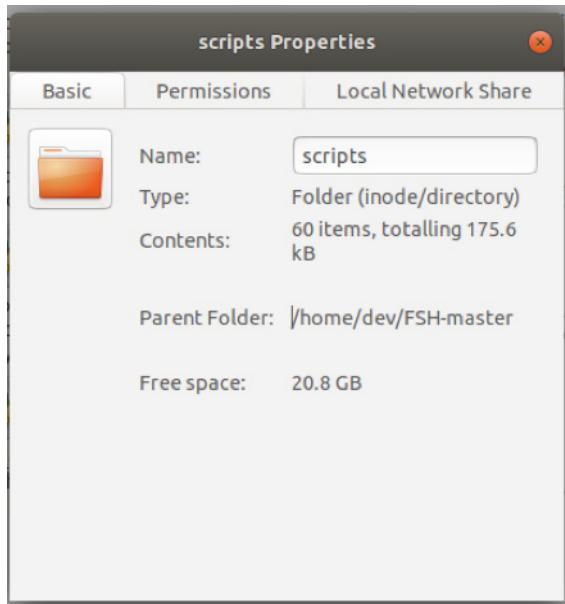
1. Press the green arrow to run your virtual machine.
2. To open the terminal press ctr, alt, and t.
3. Type in the command “conda activate sar” to access the Anaconda environment and dependencies that you installed in section 2.2. Notice that your terminal should change from “base” to “sar” environment as shown below.



The screenshot shows a terminal window titled "dev@dev-VirtualBox: ~". The window has a dark background with white text. At the top, there is a menu bar with options: File, Edit, View, Search, Terminal, Help. Below the menu, the terminal prompt "(base)" is followed by the command "dev@dev-VirtualBox:~\$ conda activate sar". The terminal then displays the response "(sar)" and the prompt "dev@dev-VirtualBox:~\$".

## 6.2 Find the Directory of Scripts and Example Data

1. In order to run the scripts from your terminal, you will need the directory to your scripts and the directory to your example data. To get to the directory of your files, right click within the folder that they are stored.
2. From the popup menu that appears, choose “Properties.”
3. You can then copy and paste the Parent Folder plus the folder name from the properties window into your script. Below is the properties window for the folder that holds my scripts.



## 6.3 Run Main FSH Script

1. Now let's call the first script! For the ROI\_PAC processed example files, enter the command into the terminal “python .../forest\_stand\_height.py 3 2 2 5 “linkfile.txt” “flagfile.txt” “Howland\_LVIS\_NaN.tif” “Maine\_NLCD2011\_non-wildland.tif” .../test\_example\_ROIPAC/ “gif json kml mat tif” --flag\_proc=0” into the terminal, where “...” is the path to your forest\_stand\_height script and your example ROI\_PAC data respectively, as shown below. For the ISCE data this would look like “python /home/dev/FSH-master/scripts/forest\_stand\_height.py 3 2 2 5 “linkfile.txt” “flagfile.txt” “Howland\_LVIS\_NaN.tif” “Maine\_NLCD2011\_nonwildland.tif” /home/dev/Downloads/test\_example\_ISCE/test\_example\_ISCE/ “gif json kml mat tif” --flag\_proc=1”

```
Resolving deltas: 100% (88/88), done.  
(sar) fsh@fsh-VirtualBox:~$ python /home/fsh/FSH/scripts/forest_stand_height.py  
3 2 2 5 "linkfile.txt" "flagfile.txt" "Howland_LVIS_NaN.tif" "Maine_NLCD2011_non  
wildland.tif" "/home/fsh/test_example_ROIPAC/test_example_ROIPAC/ "gif json kml m  
at tif" --flag_proc=0
```

2. Let's review what each of these inputs mean:
  - First, we call “python” in order to run the python scripts within the terminal. The following parameters for the FSH scripts listed in brackets are optional, while the other parameters require input.
  - file\_directory/forest\_stand\_height.py calls the main FSH script that in turn calls the rest of the scripts necessary to calculate FSH. You must pro-

vide the appropriate file directory to this script. For this example, the file directory is “/home/fsh/FSH/scripts/forest\_stand\_height.py.”

- Scenes - enter the number of scenes in the dataset. This must be an integer. If using a single radar scene, enter 1. In this example, we have 3 scenes.
- Edges - enter the number of scene to scene borders. If using a single radar scene, enter 0. In this example, we have 2 scene to scene borders.
- start\_scene (int) - flag value of the central scene that overlaps the forest stand height ground truth (e.g. LiDAR, field) data. In this example, the central scene is 2.
- iterations (int) - number of iterations to run the nonlinear least squares part of the model. In this example, we want to run the nonlinear least squares part of the model 5 times.
- link\_file - a text file that lists all the edge scene pairs. Each line consists of the two numbers that correspond to the flag numbers for those two scenes. (e.g. “2 1” would be the line for the edge of the above scenes 001 and 002). If using a single ALOS scene, this file is unneeded, and input “-” instead of the file name for the terminal arguments. For this example, the file name is “linkfile.txt.”
- flag\_file - a text file that lists all the flags and corresponding full file names and associated file information (dates, scene location (frame#, orbit#), polarization). In this example, the file name is “flagfile.txt.” Examples of what this text file would contain are:

```
001 890_120_20070727_HV_20070911_HV 070727 070911 890 120 HV  
002 890_119_20070710_HV_20071010_HV 070710 071010 890 119 HV  
003 890_118_20070708_HV_20070923_HV 070708 070923 890 118 HV
```

- ref\_file - reference tree height data (lidar or field inventory) in raster format. Currently the code is set up to use a GeoTIFF file, but other reference data in raster format could potentially be used with some code adjustments. In this example, the reference tree height data is “Howland\_LVIS\_NaN.tif.”
- mask\_file - land cover mask that excludes all water areas and areas of human disturbance (urban, agriculture). This is currently set up to be a GeoTIFF file. Other reference data in raster format could potentially be used with some code adjustments. File must be in degrees (i.e., EPSG 4326). This file is recommended, but optional. If unused, put “-” in place of the file name for the terminal arguments. For this example, the file name is “Maine\_NLCD2011\_nonwildland.tif.”
- file\_directory - the root directory to folders containing the individual SAR scenes. Each scene should have a directory named “f\$frame\_o\$orbit” (e.g. “f890\_0120” for the above scene 001). This directory contains either the input ROI\_PAC processed or ISCE processed files and is also the output location for all files that are associated with that scene. For this example, the directory is: /home/fsh/test\_example\_ROIPAC/test\_ex-

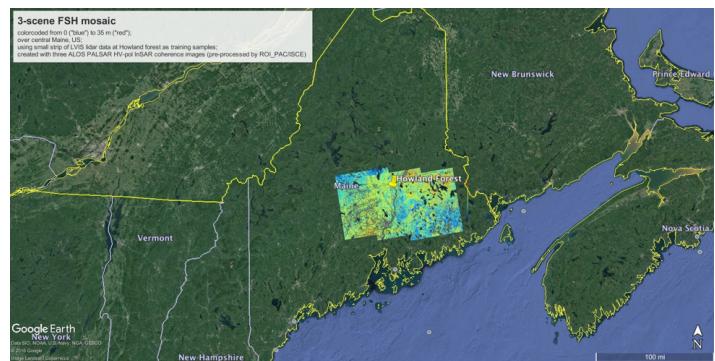
ample\_ROIPAC/. Please note that no quotes are used in the terminal for this parameter.

- Output file types - the list of output formats should be in quotes, and can contain one or all of the following: "tif kml gif mat json". In other words, output formats can be created for any of these options. For this example, all options are listed.
  - The command option --flag\_proc 0 indicates that the input data has been processed into SLCs by the ROI\_PAC algorithm. If the data was processed by ISCE, please use 1 instead. For this example, we use a 0 to indicate that the data was processed by ROI\_PAC.
3. The scripts are also able to be run with a single radar scene. To do this use “-” instead of a link\_file name, and in the input have 0 edges.
    - For example: python .../forest\_stand\_height.py 1 0 1 5 - “flagfile.txt” “Howland\_LVIS\_NaN.tif” “Maine\_NLCD2011\_nonwildland.tif” /directory\_of\_files/ “gif json kml mat tif” --flag\_proc=1
  4. In the case that you are running the FSH scripts on your own data, or would like to call each FSH script individually in the command line, please find the inputs, outputs, and terminal command lines in section 6.5. Please note that there are additional, unrequired parameters for the forest\_stand\_height.py that are explained in section 6.5 that are not included in the example. Otherwise, proceed to section 6.4 to generate a mosaic of your forest stand height estimation.

## 6.4 Generate Mosaic

1. To create a mosaic of the generated forest height maps for all the scenes in GeoTiff format, run the following command “python directory\_of\_scripts/create\_mosaic.py directory mosaic\_file list\_of\_files” in the terminal. You will need to replace three parameters.
  - Replace directory\_of\_scripts with the location of the scripts.
  - Replace mosaic\_file with the name you would like to give your final mosaic of forest stand heights.
  - Replace list\_of\_files with paths to each map that you would like to be combined within the mosaic in the format “file1 file2 file3.”
2. For example:
 

```
/home/dev/test_example_ROIPAC/test_example_ROIPAC//create_mosaic.py /home/dev/test_example_ROIPAC/test_example_ROIPAC/“3sc_mosaic.tif” .../test_example_ROIPAC/f890_0118/890_118_20070808_HV_20070923_HV_fsh.tif .../test_example_ROIPAC/f890_0119/890_119_20070710_HV_20071010_HV_fsh.tif .../test_example_ROIPAC/f890_0120/890_120_20070727_HV_20070911_HV_fsh.tif”
```
3. Following is a snapshot of the expected mosaicked forest stand height results using the example dataset.



## 6.5 Overview of Scripts

Let's review the scripts in the general order that they are called, including their main purpose, inputs, outputs, and terminal commands.

1. forest\_stand\_height.py is the main script, which in turn calls nine other scripts with a total runtime of around 23 minutes 22 secs for the example data. Some of the other scripts call additional scripts.

The command line call is:

- python file\_directory/forest\_stand\_height.py scenes edges start\_scene iterations link\_file flag\_file ref\_file mask\_file file\_directory “output\_file\_types” [--Nd\_pairwise] [-Nd\_self] [-N\_pairwise] [-N\_self] [-bin\_size] [-flag\_sparse] [-flag\_diff] [-flag\_error] [—numLooks] [—noiselevel] [-flag\_proc] [-flag\_grad].”

The inputs for this script in the order entered into the terminal are:

- scenes (int) - number of scenes in the data set
- edges (int) - number of edges (aka scene-scene borders)
- start\_scene (int) - flag value of the central scene that overlaps the ground truth (e.g. LiDAR, field) data
- iterations (int) - number of iterations to run the nonlinear least squares part of the model
- Link\_file (string) - file name of the file that lists all the edge scene pairs or ‘-’ if processing a single scene
- flag\_file (string) - file name of the file that lists all the scene flags and corresponding full file names and associated file date (dates, scene location (frame#,orbit#), polarization)
- ref\_file (string) - filename of reference data raster file (ground truth data, e.g. LiDAR, field)
- maskfile (string) - filename of the mask file that excludes all non-forest areas (mask excluding water and human disturbed areas such as urban and agriculture is also acceptable; if no mask is available input ‘-’ as the filename)
- file\_directory (string) - directory path of where the input and output files are located

- a. filetypes (string) - list of the desired output file types formatted as a single string with quotation marks (e.g. "kml json tif")
- b. [-Nd\_pairwise] (int) - optional pixel-averaging parameter for edge fitting (default=20)
- c. [-Nd\_self] (int) - optional pixel-averaging parameter for central scene fitting (default=10)
- d. [-N\_pairwise] (int) - optional pixel-averaging parameter for edge error metrics (default=20)
- e. [-N\_self] (int) - optional pixel-averaging parameter for central scene error metrics (default=10)
- f. [---bin\_size] (int) - optional bin size for density calculation in sparse data cloud fitting (default=100)
- g. [-flag\_sparse] (int) - optional flag for sparse data cloud filtering (choose 0 or 1, default=0)
- h. [-flag\_diff] (int) - optional flag for exporting differential height maps (choose 0 or 1, default=0)
- i. [-flag\_error] (int) - optional flag for exporting .json error metric files (choose 0 or 1, default=0)
- j. [-numLooks] (int) - number of looks in the correlation estimation (default=20)
- k. [-noisefile] (float) - sensor thermal noise level (ALOS's value hardcoded as default if no value provided)
- l. [-flag\_proc] (int) - flag for InSAR processor selection (choose 0 for ROI\_PAC or 1 for ISCE, default=0)
- m. [-flag\_grad] (int) - flag for correction of large-scale temporal change gradient (choose 0 or 1, default=0)

There are no direct outputs from this script, as all the file outputs are created within the scripts that are called by this main script.

The scripts called by forest\_stand\_height.py are: auto\_tree\_height.py, read\_linkfile.py, intermediate.py, intermediate\_self.py, auto\_mosaicking\_new.py, write\_deltaSC.py, write\_mapfile\_new.py, write\_diff\_height\_map.py, and cal\_error\_metric.py

2. auto\_tree\_height\_many.py is called by the forest\_stand\_height. This script extracts data from ROI\_PAC/ISCE output files and formats them for use in the rest of the scripts. For each scene, this script runs auto\_tree\_height\_single.py, and then saves the output correlation magnitudes, kz, and coordinates in a .mat file, and geo data (lines, samples, corner latitude and longitude, and latitude and longitude step size) in a text file.

The command line call for this script is python directory\_of\_scripts/auto\_tree\_height\_many.py scenes flagfile directory

The inputs for auto\_tree\_height\_many.py are:

- scenes (int) - number of scenes in the data set
- flagfile (string) - file name of the file that lists all the scene flags and corresponding full file names and associated file date (dates, scene location (frame#,orbit#), polarization)
- directory (string) - directory path of where the input and output files are located
- [-numLooks] (int) - number of looks in the correlation estimation (default=20)
- [-flag\_proc] (int) - flag for InSAR processor selection (input 0 for ROI\_PAC or 1 for ISCE, default=0)
- [-flag\_grad] (int) - flag for correction of large-scale temporal change gradient (input 0 or 1, default=0)

The outputs for this script are:

- scenename\_orig.mat - .mat file that stores correlation map, kz value, and corner coordinates
- scenename\_geo.txt - text file that stores the geodata (width, lines, corner lat and lon, and lat and lon step values)

Auto\_tree\_height\_many.py calls auto\_tree\_height\_single\_ROIPAC and auto\_tree\_height\_single\_ISCE.

3. auto\_tree\_height\_single\_ROIPAC.py calls the script read\_rsc\_data.py in order to read the value of the given parameter from the rsc file produced by ROI\_PAC processing of SAR data. This script also calls remove\_corr\_bias.py to remove correlation bias associated with ROI\_PAC. This script is called by auto\_tree\_height\_many.py and cannot be run in the terminal on its own as it needs to be iterated for each scene in the analysis.

The inputs for this script are:

- directory (string) - directory path of where the input and output files are located
- date1 (string) - date of the first image of the interferogram (format however they are listed in the scene data text file, such as 070911 for September 11, 2007)
- date2 (string) - date of the second image of the interferogram (same format as date1)
- numLooks (int) - number of looks in the correlation estimation
- noisefile (float) - sensor thermal noise level (ALOS's value hardcoded as default)
- flag\_grad (int) - flag for correction of large-scale temporal change gradient (input 0 or 1)

The outputs for this script are:

- corr\_vs (numpy array) - array of the correlation magnitudes
  - kz (float) - kz parameter
  - coords (numpy array) - array of max lat and lon values in the format [north, south, west, east]
  - geo\_width (int) - number of columns of image data
  - geo\_nlines (int) - number of rows of image data
  - corner\_lat (float) - max latitude value (north)
  - corner\_lon (float) - min latitude value (west)
  - step\_lat (float) - latitude pixel size in decimal degrees
  - step\_lon (float) - longitude pixel size in decimal degrees
4. read\_rsc\_data.py reads a parameter from the ROI\_PAC.rsc text output file. This script is called by auto\_tree\_height\_single\_ROIPAC.py and is not meant to be run in the terminal.

Inputs for this script are:

- filename (string) - file name of the ROI\_PAC text output file containing the desired parameter (may include subdirectories containing the ROI\_PAC output files - everything lower than the main file directory)
- directory (string) - directory path of where the input and output files are located
- param (string) - name of the desired parameter

Outputs for this script are the parameter values as floats (result)

5. auto\_tree\_height\_single\_ISCE.py calls remove\_corr\_bias.py to remove correlation bias associated with ISCE. This script is called by auto\_tree\_height\_many.py and cannot be run in the terminal on its own since it needs to be iterated for each scene in the analysis.

The inputs for this script are:

- directory (string) - directory path of where the input and output files are located
- date1 (string) - date of the first image of the interferogram (format however they are listed in the scene data text file, such as 070911 for September 11, 2007)
- date2 (string) - date of the second image of the interferogram (same format as date1)
- numLooks (int) - number of looks in the correlation estimation
- noiselevel (float) - sensor thermal noise level (ALOS's value hardcoded as

default if no value provided)

- flag\_grad (int) - flag for correction of large-scale temporal change gradient (input 0 or 1)

The outputs for this script are:

- corr\_vs (numpy array) - array of the correlation magnitudes
- kz (float) - kz parameter
- coords (numpy array) - array of max lat and lon values in the format [north, south, west, east]
- geo\_width (int) - number of columns of image data
- geo\_nlines (int) - number of rows of image data
- corner\_lat (float) - max latitude value (north)
- corner\_lon (float) - min latitude value (west)
- step\_lat (float) - latitude pixel size in decimal degrees
- step\_lon (float) - longitude pixel size in decimal degrees

6. intermediate.py calculates the overlap between each pair of images. This script is called by forest\_stand\_height.py.

To run in the terminal, enter the command: python directory\_of\_scripts/intermediate.py edges start\_scene linkfile maskfile flagfile ref\_file directory

The inputs for this script are:

- edges (int) - number of edges (aka scene-scene borders)
- start\_scene (int) - flag value of the central scene that overlaps the ground truth forest height data
- linkarray (numpy array) - array of the scene pairs that correspond to each edge in the format array([[scene1, scene2], [scene1, scene3], etc])
- maskfile (string) - filename of the mask file that excludes all non-forest areas (mask excluding water and human disturbed areas such as urban and agriculture is also acceptable)
- flagfile (string) - file name of the file that lists all the scene flags and corresponding full file names and associated file date (dates, scene location (frame#,orbit#), polarization)
- ref\_file (string) - filename of the reference data raster file
- directory (string) - directory path of where the input and output files are located

There's no direct output for this script since all file outputs are created in subprocesses.

Intermediate.py calls intermediate\_self.py and intermediate\_pairwise.py.

7. `remove_corr_bias.py` removes the correlation bias associated with processing by ROI\_PAC or ISCE.

The inputs for this script are:

- `C` (numpy array) - correlation magnitude array
- `numLooks` (int) - number of looks in the correlation estimation

The output for this script is `YC` (numpy array) - correlation magnitude array (with bias removed)

8. `intermediate_pairwise.py` calculates the overlap between each pair of scenes, reading the data directly from `auto_tree_height_single` rather than from an intermediary file. This script in turn calls `flag_scene_file.py` and `remove_nonforest.py`. This script is called by `auto_tree_height_single.py` and is not meant to be run from the terminal.

The inputs for this script are:

- `flag1` (int) - flag value of one scene in the pair
- `flag2` (int) - flag value of the other scene in the pair
- `flagfile` (string) - file name of the file that lists all the scene flags and corresponding full file names and associated file date (dates, scene location (frame#,orbit#), polarization)
- `maskfile` (string) - filename of the mask file that excludes all non-forest areas (mask excluding water and human disturbed areas such as urban and agriculture is also acceptable)
- `directory` (string) - directory path of where the input and output files are located
- `filename1_orig.mat`: correlation map and associated parameters for the first scene (generated in previous steps)
- `filename2_orig.mat`: correlation map and associated parameters for the second scene (generated in previous steps)

The outputs for this script are link files: one for each overlapping edge region, with the filename format `flag1_flag2.mat`

9. `intermediate_self.py` calculates the overlap between the forest height validation data and central scene. This script in turn calls `flag_scene_file.py` and `remove_nonforest.py`. This script is called by `intermediate.py` and is not meant to be run from the terminal.

The inputs for this script are:

- `start_scene` (int) - flag value of the central scene that overlaps the ground truth data
- `flagfile` (string) - file name of the file that lists all the scene flags and corresponding full file names and associated file date (dates, scene location (frame#,orbit#), polarization)

- `directory` (string) - directory path of where the input and output files are located
- `filename_orig.mat`: correlation map and associated parameters for the central scene (generated in previous steps)
- `reference` data raster file (already exists; main input)

The output for this script is `self.mat`, a link file for the central scene-ground truth overlap region

10. `flag_scene_file.py` associates flag numbers with the name, dates, ALOS location (frame and orbit), and polarization of each scene. This script is called by `intermediate_pairwise.py`, `write_deltaSC.py`, and `write_mapfile_new.py` and is not meant to be run from the terminal.

The inputs for this script are:

- `flagfilename` (string) - file name of the file that lists all the flags and corresponding full file names and associated file date (dates, scene location (frame#,orbit#), polarization)
- `flag` (int) - flag of the desired scene
- `directory` (string) - directory path of where the input and output files are located

The output for this script is a `data_array` (list) - list of the data associated with the given flag number.

11. `remove_nonforest.py` removes all non-forest areas from the image based on the non-forest `mask_file`. This script is called by `intermediate_pairwise.py` and `write_mapfile_new.py` and is not meant to be run in the terminal.

The inputs for this script are:

- `I` (numpy array) - the image data
- `func_coords` (numpy array) - array of corner coordinates
- `maskfile` (string) - filename of the mask file that excludes all non-forest areas (mask excluding water and human disturbed areas such as urban and agriculture is also acceptable)
- `directory` (string) - directory path of where the input and output files are located

The output for this script is `O` (numpy array) - image without the non-forest sections.

12. `auto_mosaicking_new.py` calculates the S and C parameters automatically by iterating through all the scenes in preparation for forest height estimation. This script is called by `forest_stand_height.py`. `auto_mosaicking_new.py` calls `ls_deltASC.py` and `read_linkfile.py`

To run in the terminal, enter the command: `python directory_of_scripts/auto_mosaicking_new.py scenes start_scene N linkfile directory`

The inputs for this script are:

- scenes (int) - number of scenes in the data set
- edges (int) - number of edges (aka scene-scene borders)
- start\_scene (int) - flag value of the central scene that overlaps the ground truth data
- N (int) - number of iterations to run the nonlinear least squares part of the model
- Linkfile - the filename of the file that lists all the edge scene pairs.
- linkarray (numpy array) - array of the scene pairs that correspond to each edge in the format array([[scene1, scene2], [scene1, scene3], etc])
- directory (string) - directory path of where the input and output files are located
- [-Nd\_pairwise] (int) - pixel-averaging number for image fitting between two overlapped radar scenes (default=20)
- [-Nd\_self] (int) - pixel-averaging number for image fitting between single radar scene and the overlapped ground truth data (default=10)
- [-bin\_size] (int) - bin size for density calculation in scatter plot fitting when ground truth data are sparse (default=100)
- [-flag\_sparse] (int) - flag for sparse data cloud fitting (input 0 or 1, default=0)

The outputs produced by this script are iteration files (.json format; e.g. "SC\_#\_iter.json" for "#th iteration) that store the increment steps of S and C parameters and the residual; no values are returned by the function.

13. ls\_deltaSC.py runs least squares on the change in S and C parameters. This script in turn calls cal\_KB.py. This script is called by auto\_mosaicking\_new.py and is not meant to be run from the terminal.

The inputs for this script are:

- dp (numpy array) - array of increment steps of S and C parameter values
- edges (int) - number of edges (aka scene-scene borders)
- scenes (int) - number of scenes in the data set
- start\_scene (int) - flag value of the central scene that overlaps the ground truth data
- linkarray (numpy array) - array of the scene pairs that correspond to each edge in the format array([[scene1, scene2], [scene1, scene3], etc])
- directory (string) - directory path of where the input and output files are located
- Nd\_pairwise (int) - pixel-averaging number for image fitting between two overlapped radar scenes
- Nd\_self (int) - pixel-averaging number for image fitting between single

radar scene and the overlapped ground truth data

- bin\_size (int) - bin size for density calculation in scatter plot fitting
- flag\_sparse (int) - flag for sparse data cloud filtering (input 0 or 1)

The outputs for this script are:

- changeSC (numpy array) - updated S and C parameters as referenced to the average S (=0.6) and C (=13)
  - res (float) - residual k and b error compared to k = 1 and b = 0
14. cal\_KB.py calculates the K and B parameters. This script in turn calls cal\_KB\_pairwise\_new.py and cal\_KB\_self\_new.py. This script is called by ls\_deltaSC.py and is not meant to be run from the terminal.

The inputs for this script are:

- R (float) - R parameter for this edge
- RSME (float) - RSME parameter for this edge
- R\_RSME\_files: one for each edge with the filename format scene1\_scene2\_l1andl2.json

The output for this script is YY (numpy array), an array of k and b values.

15. cal\_KB\_pairwise\_new.py calculates K and B between image pairs. In turn, this script calls arc\_sinc.py, mean\_wo\_nan.py, extract\_scatterplot\_density.py, and remove\_outlier.py. This script is called by cal\_KB and is not meant to be run in the terminal.

The inputs for this script are:

- scene1 (int) - flag value of one scene in the pair
- scene2 (int) - flag value of the other scene in the pair
- deltaS1 (float) - change in S value for one scene in the pair
- deltaC1 (float) - change in C value for one scene in the pair
- deltaS2 (float) - change in S value for the other scene in the pair
- deltaC2 (float) - change in C value for the other scene in the pair
- directory (string) - directory path of where the input and output files are located
- Nd\_pairwise (int) - pixel-averaging number for image fitting between two overlapped radar scenes
- bin\_size (int) - bin size for density calculation in scatter plot fitting
- link files: one for each overlapping edge region, with the filename format scene1\_scene2.mat (generated in previous steps)

The outputs for this script are:

- k (float) - k parameter for this edge
- b (float) - b parameter for this edge

16. `cal_KB_self_new.py` calculates K and B between the central image and the forest height validation data. In turn, this script calls `arc_sinc.py`, `mean_wo_nan.py`, `extract_scatterplot_density.py`, and `remove_outlier.py`. This script is called by `cal_KB` and is not meant to be run in the terminal.

The inputs for this script are:

- `deltaS2` (float) - change in S value for the central scene
- `deltaC2` (float) - change in C value for the central scene
- `directory` (string) - directory path of where the input and output files are located
- `Nd_self` (int) - pixel-averaging number for image fitting between single radar scene and the overlapped ground truth data
- `bin_size` (int) - bin size for density calculation in scatter plot fitting
- `flag_sparse` (int) - flag for sparse data cloud filtering (input 0 or 1)
- `self.mat`: link file for the central scene-ground truth overlap region (generated in previous steps)

The outputs for this script are:

- `k` (float) - k parameter for this edge
- `b` (float) - b parameter for this edge

17. `arc_sinc.py` calculates the inverse sinc function as part of calculating K and B values. This script is called by `cal_KB_pairwise` and `write_mapfile_new.py` and is not meant to be run in the terminal.

The inputs for this script are:

- `X` - A numpy array of x values for the inverse sinc function
- `c_params` - C parameter (float) from the Forest Stand Height model

The outputs for this script are:

- `y` - a numpy array of y values of inverse sinc function satisfying  $x = \text{sinc}(y/C)$

18. `mean_wo_nan.py` calculates and returns the mean of all number values in an array as part of calculating K and B values. This script is called by `cal_KB_pairwise_new.py` and is not meant to be run in the terminal.

Inputs for this script are:

- `A` (numpy array) - input array of values

Outputs for this script are:

- mean of B (A excluding NaN values) (float)

19. `extract_scatterplot_density.py` calculates the 2D histogram of the scatterplot between pairs of forest height and returns the forest height pairs with relatively large density. This script is intended to replace `remove_outlier.py` in order to distinguish between forest disturbance and forest height estimation. This script is called by `cal_KB_pairwise` and is not intended to be run in the terminal.

The inputs for this script are:

- `x` (numpy array) - array of x values of points
- `y` (numpy array) - array of y values of points
- `bin_size` (int) - bin size for density calculation in scatter plot fitting (default = 100)
- `threshold` (float) - density threshold (default = 0.5)

The outputs for this script are:

- `Hm_den` (numpy array) - array of x values of the points with densities above the inputted threshold
- `Pm_den` (numpy array) - array of y values of the points with densities above the inputted threshold

20. `remove_outlier.py` this script is called by `cal_KB_self_new.py`, `cal_KB_pairwise.py`, `cal_KB_pairwise_new.py`, `cal_error_metric_pairwise.py`, and `cal_error_metric_self.py` to remove outliers, and is supplemented by the function of `extract_scatterplot_density.py`.

The inputs for this script are:

- `x` (numpy array) - array of x values of points
- `y` (numpy array) - array of y values of points
- `window_size` (float) - window size to search for neighboring points (defaults to 0.5)
- `threshold` (int) - number of neighboring points needed within the window to not count as an outlier (defaults to 5)

The outputs for this script are:

- `XX` (numpy array) - array of x values of the points excluding those counted as outliers
- `YY` (numpy array) - array of y values of the points excluding those counted as outliers

21. `read_linkfile.py` reads in a text file containing a list of all the scene pairs and returns a 2D array of the pairs. This script is called by `auto_mosaicking_new.py`

To run this script in the terminal, use the following command: `python directory_of_scripts/read_linkfile.py edges filename directory`

The inputs for this script are:

- edges (int) - number of edges (aka scene-scene borders)
- filename (string) - file name of the file that lists all the edge scene pairs
- directory (string) - directory path of where the input and output files are located

The outputs for this script is linkarray (numpy array) - array of the scene pairs that correspond to each edge in the format array([[scene1, scene2], [scene1, scene3], etc])

22. write\_deltaSC.py calculates the temporal change parameters (S and C) as referenced to the average values: S=0.6, C=13 based on the final iteration. This script is called by forest\_stand\_height.py. write\_deltaSC.py in turn calls flag\_scene\_file.py.

To run in the terminal, enter the command: python directory\_of\_scripts/write\_deltaSC.py scenes N flagfile directory

The inputs for this script are:

- scenes (int) - number of scenes in the data set
- N (int) - number of iterations to run the nonlinear least squares part of the model
- flagfile (string) - file name of the file that lists all the scene flags and corresponding full file names and associated file date (dates, scene location (frame#,orbit#), polarization)
- directory (string) - directory path of where the input and output files are located
- SC\_#\_iter.json: final iteration file (generated in previous steps)

The output for this script is one file per scene that contains delta S and C. The file name format is "scenename\_tempD.json"

23. write\_mapfile\_new.py calculates and writes the tree height map to a file. This script is called by forest\_stand\_height.py. This script calls flag\_scene\_file.py, arc\_sinc.py, remove\_nonforest.py and write\_file\_type.py.

The inputs for this script are:

- scenes (int) - number of scenes in the data set
- flagfile (string) - file name of the file that lists all the scene flags and corresponding full file names and associated file date (dates, scene location (frame#,orbit#), polarization)
- maskfile (string) - filename of the mask file that excludes all non-forest areas (mask excluding water and human disturbed areas such as urban and agriculture is also acceptable) (optional - if no mask available use '-' as an input to forest\_stand\_height.py)

- directory (string) - directory path of where the input and output files are located
- output\_files (string) - list of the desired output file types formatted as a single string (e.g. "kml json tif")
- scenename\_orig.mat: correlation map and associated parameters for the central scene (generated in previous steps)
- scenename\_tempD.json: delta S and C files produced (generated in previous steps)

There's no direct output (all file output created in write\_file\_type.py).

24. write\_file\_type.py writes the input array from the tree height map or the diff\_height map to a file, with the file type depending on input parameters: gif, json, kml, mat, or tif. In turn this script calls read\_geo\_data.py. This script is called by write\_mapfile\_new.py and is not meant to be run in the terminal.

The inputs for this script are:

- data (numpy array) - array to be written to the file
- outtype (string) - string to signify which input (tree height "stand\_height" or differential height "diff\_height") is being output
- filename (string) - scene file name
- directory (string) - directory path of where the input and output files are located
- filetype (string) - file extension for the desired output file type (.gif, .json, .kml, .mat, and .tif accepted -> input without the "." (e.g. "kml" instead of ".kml")
- coords (numpy array) - array of max lat and lon values in the format [north, south, west, east]
- reffile (string) - reference filename containing ground truth data (optional; only needed for differential height map)

The outputs for this script are output files(s) of the array image saved in the file type specified in the input.

25. read\_geo\_data.py reads in latitude, longitude, pixel size, and image size from a GeoTIFF or text file based on ROI\_PAC output. This script is called by write\_file\_type.py and is not meant to be run in the terminal.

Inputs for this script are:

- coord\_file (string) - file name of the input data file with the location information (lat/long, step size, image size)
- directory (string) - directory path of where the input and output files are located

Outputs for this script are:

- width (int) - width/number of columns of the image
  - nlines (int) - lines/number of rows of the image
  - corner\_lat (float) - latitude of the upper left corner
  - corner\_long (float) - longitude of the upper left corner
  - post\_lat (float) - latitude step size
  - post\_long (float) - longitude step size
26. write\_diff\_height\_map.py writes the forest differential height map between SAR and overlapping forest height ground truth images. This script is called from forest\_stand\_height if the parameter --flag\_diff is entered.

Inputs for this script are:

- start\_scene (int) - flag value of the central scene that overlaps the ground truth data
- reffile (string) - reference filename containing ground truth data
- flagfile (string) - file name of the file that lists all the scene flags and corresponding full file names and associated file date (dates, scene location (frame#,orbit#), polarization)
- maskfile (string) - filename of the mask file that excludes all non-forest areas (mask excluding water and human disturbed areas such as urban and agriculture is also acceptable) (optional; if no masks are available, use '-' as an input to forest\_stand\_height.py)
- directory (string) - directory path of where the input and output files are located
- output\_files (string) - list of the desired output file types formatted as a single string (e.g. "kml json tif")

There is no direct output for this script, as all file output is created in write\_file\_type.py.

27. cal\_error\_metric.py calculates the R and RMSE error metrics for the model. This script is called from forest\_stand\_height.py if the parameter --flag\_error is entered. This script calls cal\_error\_metric\_pairwise.py and cal\_error\_metric\_self.py.

The inputs for this script are:

- dp (numpy array) - array of increment steps of S and C parameter values
- edges (int) - number of edges (aka scene-scene borders)
- start\_scene (int) - flag value of the central scene that overlaps the ground truth data
- link (numpy array) - array of the scene pairs that correspond to each edge in the format array([[scene1, scene2], [scene1, scene3], etc])
- directory (string) - directory path of where the input and output files are located

located

- N\_pairwise (int) - pixel-averaging number for scatter plot
- N\_self (int) - pixel-averaging number for scatter plot

The output for this script is YY, a numpy array of R and RMSE values.

28. cal\_error\_metric\_pairwise.py calculates the R and RMSE error metrics. This script calls arc\_sinc.py, mean\_wo\_nan.py and remove\_outlier.py. It is called by cal\_error\_metric.py and is not meant to be run in the terminal.

The inputs for this script are:

- scene1 (int) - flag value of one scene in the pair
- scene2 (int) - flag value of the other scene in the pair
- deltaS1 (float) - change in S value for one scene in the pair
- deltaC1 (float) - change in C value for one scene in the pair
- deltaS2 (float) - change in S value for the other scene in the pair
- deltaC2 (float) - change in C value for the other scene in the pair
- directory (string) - directory path to where the input and output files are located
- N\_pairwise (int) - pixel-averaging number for the scatter plot
- link files: one for each overlapping edge region, with the filename format scene1\_scene2.mat (generated in previous steps)

The outputs for this script are:

- R (float) - R parameter for this edge
- RSME (float) - RSME parameter for this edge
- R\_RSME\_files: one for each edge, with the filename format scene1\_scene2\_l1andl2.json

29. cal\_error\_metric\_self.py calculates R and RMSE between the central image and the forest height ground validation data. This script calls arc\_sinc.py, mean\_wo\_nan.py, and remove\_outlier.py. This script is called by cal\_error\_metric.py and is not meant to be run in the terminal.

The inputs for this script are:

- deltaS2 (float) - change in S value for the central scene
- deltaC2 (float) - change in C value for the central scene
- directory (string) - directory path of where the input and output files are located
- N\_self (int) - pixel-averaging number for scatter plot
- self.mat: link file for the central scene-ground truth overlap region (generated in previous steps)

The output for this script is YY (numpy array) - array of R and RMSE values.





**DR. SASSAN SAATCHI** is a senior scientist with the carbon cycle and ecosystems section at the Jet Propulsion Laboratory (JPL), California Institute of Technology and an adjunct professor at the UCLA Institute of the Environment and Sustainability and the Center for Tropical Research. He received his PhD in Electrophysics and Applied Mathematics from Georgetown University in 1988, and held a postdoctoral fellowship at the NASA/GSFC hydropheric branch before joining JPL in 1991. During the past 20 years, Dr. Saatchi has participated in several NASA-supported interdisciplinary and international projects as principal investigator and has developed remote sensing techniques and algorithms for regional-scale ecological and hydrological variables for NASA airborne and spaceborne missions.

Dr. Saatchi's present research interests include the global carbon cycle, in particular forest biomass/carbon dynamics, land use and land cover change, forest structure and regeneration, and the impact of climate change and variability such as droughts on global forest function and resilience. He is currently directing several interdisciplinary research projects studying the carbon cycling of tropical forests, particularly in the Amazon and Congo Basin using satellite remote sensing data. As an adjunct professor at UCLA, he has also been collaborating with biologists to integrate satellite observations for mapping species functional diversity and habitat characteristics, and with social scientists to model and understand how human activities have impacted vegetation ecosystems. Dr. Saatchi has been developing new active remote sensing techniques from airborne and spaceborne platforms to quantify vegetation structure and water status and creating a data assimilation and modeling framework to improve understanding of processes that impact forest carbon and water cycling. He is actively involved in planned NASA's NISAR and ESA's BIOMASS satellite missions by developing algorithms for retrieving changes of vegetation above ground biomass across different landscapes and ecoregions.

*I would like to thank the SERVIR program for providing the opportunity, support, and guidance for the material presented in this chapter. I particularly like to thank Daniel Irwin for inviting me to participate in this effort and for his leadership role in making SERVIR a successful program and extending the data and technical resources and capacity available at NASA to a larger international community. The content of this chapter was improved significantly after the reviews and recommendations provided by the SERVIR program. I also wish to acknowledge the contributions of those cited below in providing data, figures, and examples that helped with the tutorial material and the content of this chapter: Alan Xu, Yan Yang, Antonio Ferraz, Yifan Yu, Victoria Meyer. The methods and products presented in this chapter were developed from my past and present research in the field of ground and remote sensing of forest structure and carbon stocks and dynamics. I also wish to thank the Africa Flores for her guidance and persistence in developing the material for the chapter and Dr. Rajesh Bahadur Thapa, as the SERVIR capacity building lead for organizing the SAR biomass tutorial in Kathmandu, Nepal and his hospitality and technical discussions during my visit.*

# CHAPTER 5

## SAR Methods for Mapping and Monitoring Forest Biomass

Sassan Saatchi, Senior Research Scientist, Carbon Cycle and Ecosystems Section, Jet Propulsion Laboratory, California Institute of Technology

### ABSTRACT

Forests play a major role in the global carbon cycle, sequestering more than 25% of the carbon emitted to the atmosphere from fossil fuel consumption and land-use changes. The accumulation of carbon in forests has therefore become an effective strategy for mitigating climate change and an important mechanism for countries to meet their emission requirements under many international protocols and agreements. Remote sensing techniques are considered the most promising approach for providing up-to-date information on the status of forest cover and carbon stocks at different scales. Among remote sensing techniques, Synthetic Aperture Radar (SAR) sensors at long wavelengths have the advantage of strong sensitivity to the forest Above Ground Biomass (AGB) and the ability to quantify and monitor carbon stocks at the scale in which human activities occur. This chapter provides a summary of the methodologies and techniques for estimating forest AGB and monitoring changes from existing and future SAR satellite systems. The material in this chapter is designed to help both practitioners and remote sensing students and experts use SAR imagery for mapping and monitoring forest biomass. The examples and the bibliography capture the state of the art in SAR remote sensing of vegetation structure and biomass, and provide resources for enthusiasts to follow future developments in the technology and the methodology.

## 5.1 Background

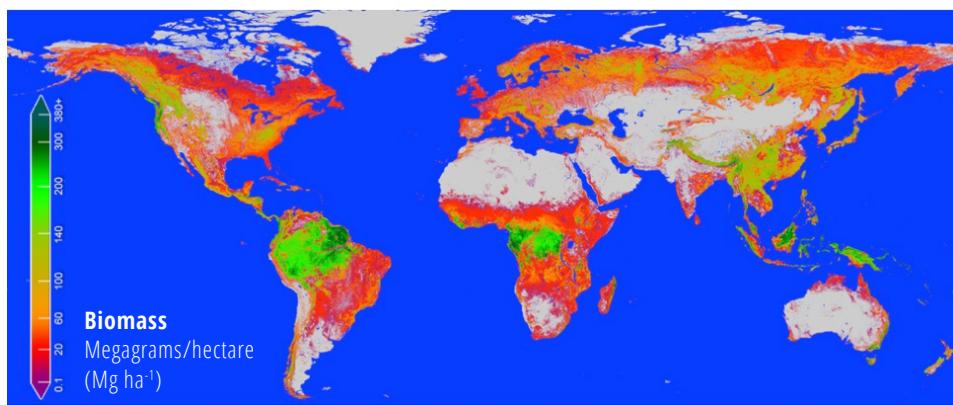
### 5.1.1 GLOBAL DISTRIBUTION OF FOREST BIOMASS

Vegetation in terrestrial ecosystems takes up a significant fraction (~30%, or  $3 \text{ PgC year}^{-1}$ ) of carbon released to the atmosphere from fossil fuel and deforestation (LeQuere et al. 2018, Schimel et al. 2015) and creates the land residual sink with a destiny dependent on future climate conditions and human activities (Ciais et al. 2013, Bonan 2008). Almost all of this sink is in forests, covering about 3.8 billion ha (FAO 2015) of the land surface (~30%) and storing large reservoirs of carbon, approximately double the amount in the atmosphere (Canadell & Raupach 2008, Sabine et al. 2004). Together, the carbon stored and sequestered in these ecosystems are major contributors to mitigating climate change and the economic benefits of emission Reductions from Deforestation and Degradation (REDD) (IPCC 2007, Gibbs et al. 2014). There are, however, large uncertainties surrounding the magnitude of the carbon stored in forests, particularly at landscape scales (1–100 ha) where mitigation benefits and ecosystem services are evaluated (Gibbs et al. 2007). A recent attempt to

put together the information from different types of measurements on a global scale captures the overall distribution of forest Above Ground Biomass (AGB) and carbon stored in global ecosystems (**Fig. 5.1**).

The structure of forests (i.e., the three-dimensional arrangement of individual trees) is a direct indicator of how much carbon is stored in the ecosystem. Carbon stored in an ecosystem has a profound effect on how the ecosystem functions (i.e., how it cycles

carbon, water, and nutrients). Additionally, there is an increased need to understand local to global storage and dynamics of carbon in ecosystems, as carbon storage is a prerequisite to understanding the coupling of the biosphere to other components of Earth systems. For example, the amount of carbon in a system determines how much is eventually emitted to the atmosphere (as  $\text{CO}_2$ ,  $\text{CO}$ , and  $\text{CH}_4$  through burning and decay) when ecosystems are disturbed due to



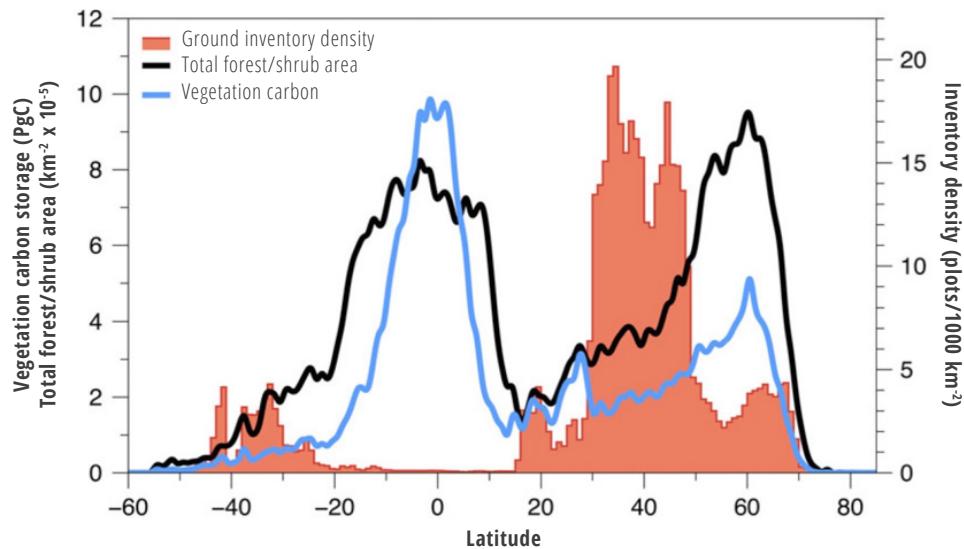
**Figure 5.1** Distribution of forest AGB density in global ecosystems showing the high biomass in tropical rainforest regions and relatively lower biomass in extratropics extending to temperate and boreal regions with vast areas of forest cover. Map is produced at 1-km spatial resolution using a combination of ground, lidar, and radar measurements by Saatchi's team at the Jet Propulsion Laboratory, California Institute of Technology.

deforestation and degradation or from climate-driven stress and fire. The amount of carbon stored in the system can be estimated from AGB, which is estimated from measurements of structure (e.g., the size and density of trees) and the mass of trees. As such, AGB is considered a crucial variable for a range of applications, including forest fire assessment, management of the timber industry, monitoring land-use change, and other ecosystem services such as biodiversity and production of food and fiber, as well as greenhouse gas accounting.

Although many of these applications may be accounted for by using operational satellite observations of forest cover change, the understanding of changes in terrestrial AGB remains rudimentary (Saatchi et al. 2011). For example, it is known that changes in land use, largely from tropical deforestation and fire, are estimated to have reduced biomass globally, while the global carbon balance suggests that terrestrial carbon storage has increased; albeit the exact magnitude, location, and causes of this residual terrestrial sink are still not well quantified (Schimel et al. 2015a, Sellers et al. 2018). There is strong evidence that the residual sinks are spread in different forest ecosystems with locations that may change due to climate change and anomalies. Yet the magnitude and fate of these terrestrial sinks are crucial to future climate projections, and any uncertainties in the spatial locations or the temporal behavior of them directly influences the current status of global carbon cycle and climate (Houghton et al. 2018, Schimel et al. 2015a).

### 5.1.2 GROUND INVENTORY OF FOREST BIOMASS

Knowledge of the distribution and amount of AGB is based almost entirely on ground inventory measurements over an extremely small (and possibly biased) set of samples, with many regions left unmeasured (**Fig. 5.2**). Conventional forest inventory data known as the National Forest Inventory (NFI) are based on systematic sampling of forests and are mainly designed for monodominant, evenly aged forests in managed temperate and boreal regions. Although the basic statistical techniques can be used for tropical forests, there are differences in terms of



**Figure 5.2** The distribution of woody (forest and shrubland) area and biomass derived from a variety of sources from field and remote sensing data. The red histogram shows forest inventory plot density in 1,000 km<sup>2</sup> grid cells (Schimel et al. 2015b), suggesting an uneven distribution of inventory plots in the Northern Hemisphere and a lack of data in tropical regions.

plot size, number of plots, and plot locations that have not been worked out for tropical forests.

- Conventional NFI can provide accurate estimates of forest carbon density at the national and potentially subnational levels depending on the density of the plots. However, they cannot provide spatial maps unless combined with remote sensing data.
- In tropical and unmanaged forests, implementation of NFI is extremely difficult, because of limited access to the site and the cost of establishing and monitoring plots over time. Using the protocols of the U.S. or northern Scandinavian NFI to the tropics requires a large number of plots.
- Conventional NFI data include 5–10 years of repeated measurements, and the timing of the measurements is not coordinated among the countries, making it difficult to conduct a global assessment for any period. For Greenhouse Gas (GHG) emissions, the use of a national inventory along with remote sensing estimation of forest cover change can provide national-level emissions estimates, but those estimates may involve uncertainty due to the lack of forest estimates in areas where deforestation occurs.

At large scales, robust AGB estimates are acquired

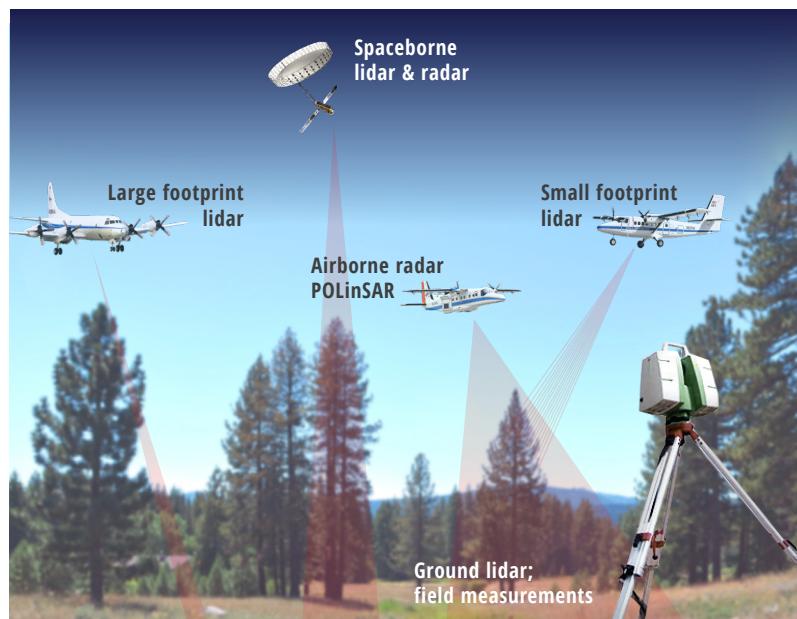
from ground-based forest censuses that are based on labor-intensive fieldwork (plot inventories) conducted by trained operators. As such, these plot inventories cannot be repeated frequently or at a low cost everywhere. Thus, plot inventories are limited to managed forests in a number of developed countries in the Northern Hemisphere where systematic sampling of forest inventories are performed on a regular basis (5- to 10-year cycles). Information on most carbon-rich global forests is missing, particularly in developing and tropical countries, even though this is where most living biomass is located (63% of carbon in intact tropical forests versus 15% in boreal forests and 13% in temperate forests, according to a recent and comprehensive estimate (FAO 2015)). Furthermore, land-use activities, along with increasing disturbances from climate and human stresses, are rapidly changing plot inventory requirements to include more frequent observations of forest ecosystems.

### 5.1.3 REMOTE SENSING OF FOREST BIOMASS

There is a strong synergy between ground and remote sensing measurements for quantifying AGB (**Fig. 5.3**). Ground data (generally consisting of all tree diameters above a threshold, a sampling of tree heights, and species identification that permits

inference of wood densities) are more comprehensive locally than remote sensing data that generally measure aggregate canopy height (in the case of lidar sensors) or some indicators of forest height and volume (in the case of radar sensors). In contrast, airborne or satellite remote sensing-based data are far more extensive, with millions of measurements over regional or continental scales compared to plots and providing a more spatially comprehensive measure of forest biomass variations. However, both ground inventory and remote sensing observations focus on measuring some physical properties (e.g., height or diameter, volume, etc.) that are not forest biomass (Clark & Kellner 2012). Both efforts rely on statistical techniques to estimate biomass, using single-tree allometry in the case of field plots and plot-aggregate allometry in the case of satellite data. Furthermore, both approaches are subject to several measurement and algorithmic errors.

A variety of remote sensing sensors provide measurements of biophysical and structural characteristics of forests based on the interaction of light or microwave energy with forest canopy and woody components. These sensors are typically categorized into passive sensors, such as spectrometers or radiometers that measure reflected or emitted radiation from the Earth's surfaces, and active sensors, which internally generate and emit energy and then measure different attributes of the returned energy bouncing back from the surface. Passive remote sensors measure different ranges of wavelengths of reflected solar radiation (optical and microwave), providing two-dimensional information that can be indirectly linked to biophysical properties of vegetation (Rosette et al. 2012, Shugart et al. 2010). Examples of passive systems include Landsat (measuring the visible spectrum), QuickBird (visible to near-infrared), AVIRIS, and MODIS, with the latter two measuring from visible to infrared (Hyde et al. 2006). On the other hand, active sensors are designed to work at limited wavelengths, such as lidar in visible or near-infrared wavelengths (Drake et al. 2002) or radar in microwave long wavelengths (Shugart et al. 2010). For more details on remote sensing techniques for forestry applications, see Zhang & Ni-meister (2013), Wulder & Franklin (2012), Zolkos



**Figure 5.3**  
Ground and remote sensing measurement techniques to quantify forest structure and AGB.

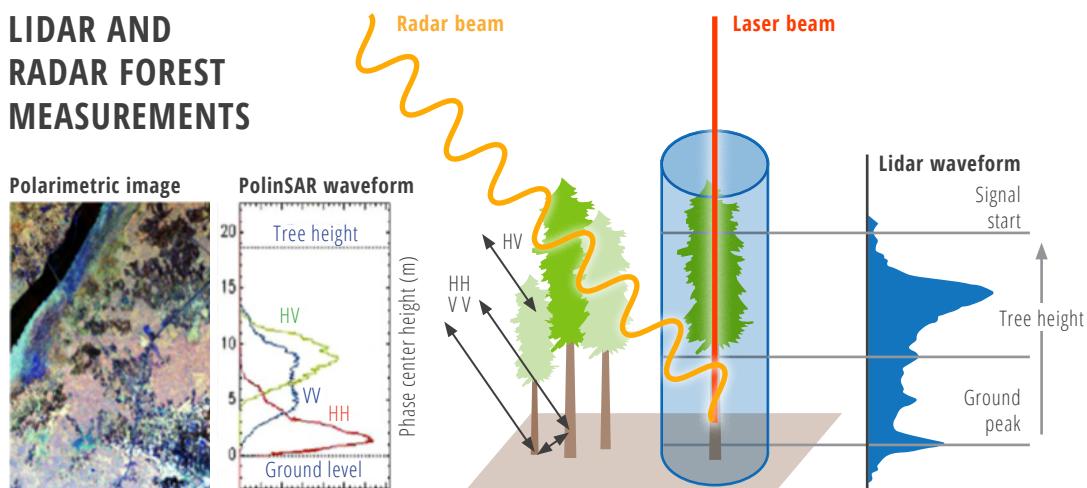
et al. (2013), Saatchi et al. (2011b), and LeToan et al. (2011). Here, for the sake of brevity, the remote sensing techniques for forest structure and biomass are divided into two categories:

- (1) The first category refers to remote sensing observations that provide the most direct measurements of forest structure, such as canopy height from lidar sensors on either airborne or spaceborne platforms. Lidar sensor measurements must be treated similarly to ground measurements such as tree height measurements using a laser ranger or clinometers in the field. In both cases, the measurements are relatively direct. Height is measured from laser altimetry from air or space, and from distance and angle measurements in the ground. There is strong evidence that tree height can be measured as accurately if not better than ground measurements using small-footprint (<1 m) lidar systems (Asner et al. 2010). Here, the measurement errors can be treated the same as measurement errors in the field (Dubayah et al. 2000, Lefsky et al. 2002, Lefsky 2010).
- (2) The second category refers to active remote sensing observations that provide indirect measurements of forest structure, such as active radar sensors for forest volume or biomass and height. In this case, radar backscatter

measurements provide strong sensitivity to forest structure and biomass. This sensitivity may be asymptotically reduced when biomass increases to a range of more than 100 to 150 Mg/ha at L-band wavelengths (~25 cm) (Saatchi et al. 2011b, Mitchard et al. 2011, Mermoz et al. 2015), and more than 200 to 300 Mg/ha at P-band wavelengths (~70) (Saatchi et al. 2011b, LeToan et al. 2011, Sandberg et al. 2011). By adding interferometric radar techniques as in PolInSAR and TomoSAR measurements, the sensitivity of radar sensors may increase over the entire biomass range in tropical forests (Hajnsek et al. 2009, Minh et al. 2015, Neumann et al. 2012). The high-resolution, two-dimensional radar measurements (backscatter power) have provided separation of tropical forest biomass based on their canopy gaps, structure, and spatial heterogeneity (Hoekman et al. 2000), and have been used as an important deforestation and degradation monitoring tool.

Lidar and radar remote sensing techniques are currently recognized as the best approaches for quantifying and monitoring forest AGB changes globally. Therefore, numerous space agencies are attempting to improve the presence of these techniques for spaceborne observation of forest bio-

## LIDAR AND RADAR FOREST MEASUREMENTS



**Figure 5.4** Lidar and radar forest inventory from air or space platforms capturing vertical and horizontal structure of forest ecosystems.

mass, with NASA and the European Space Agency (ESA) having already approved plans to develop and launch lidar and radar sensors in the near future (**Fig. 5.4**).

NASA's GEDI (launch 2018) and NISAR (launch 2021) missions, and ESA's Biomass (launch 2022), share similar objectives for developing regional or global estimates of forest structure and AGB. These missions will address one of NASA's key strategic goals for understanding changes in the Earth's climate by focusing on the most uncertain component of the global carbon cycle related to terrestrial carbon sources and sinks. All missions providing active remote sensing measurements of forest structure must be converted to AGB using algorithmic models and validated by ground-estimated AGB distributed globally in different forest types. These missions have significant overlaps in terms of science objectives and products, but each focuses on different observations, employs different algorithms, and retrieves different AGB ranges at different spatial and temporal scales. The success of these missions strongly depends on how their science products can advance scientific and societal benefits.

Biomass observations at P-band will be particularly useful for high biomass density forests in tropical regions where there is a large uncertainty in quantifying forest biomass due to the lack of national inventory data and low efficacy of existing radar and optical remote sensing techniques. ESA's Biomass mission's unique contribution to the global carbon cycle is to provide annual carbon stocks and changes

for old growth, secondary, and degraded tropical forests. It is expected that the Biomass mission's measurement sensitivity will allow for the estimation of high-biomass forests ( $>100 \text{ Mg/ha}$ ). However, for areas of low biomass density ( $<100 \text{ Mg/ha}$ ), NASA's NISAR mission at L-band frequency will perform better in terms of accuracy and spatial resolution ( $<100 \text{ m}$ ). GEDI lidar sampling measurements of forest height will be acquired approximately 12 to 18 months prior to Biomass and NISAR data acquisitions, allowing GEDI-derived forest structure to integrate with Biomass and NISAR algorithms for improving the radar estimations of forest structure and biomass.

## 5.2 Forest Biomass – Ground Inventory

In this section, forest inventory is discussed as the most reliable approach for quantifying AGB at the local scale, as well as using airborne small-footprint lidar measurements as the state-of-the-art remote sensing technique for most accurately estimating AGB at landscape scales. Currently, both techniques are used extensively in quantifying forest carbon stocks at the local, regional, and national scales and are considered the most reliable for integrating with radar observations to estimate AGB. Particularly, airborne lidar data will allow upscaling inventory measurements from small plots to a scale that can be useful in calibrating radar measurements and developing radar-based models and algorithms for AGB. This section will also be considered as the first step

toward understanding how AGB is quantified and to what extent knowledge gained from ground and lidar AGB estimates could improve the radar techniques for AGB estimation. This section provides general information about ground and lidar quantification of AGB and also provides an example discussed during the SAR tutorial for demonstration.

### 5.2.1. FOREST INVENTORY SAMPLING

Forest inventory measurements include both the direct measurement of biomass of individual trees from destructive harvesting, or indirect estimation through measurements of tree size and inference using allometric relationships (Gibbs et al. 2007, Brown 1997, Chave et al. 2005, Keller et al. 2001). However, before an allometric equation can be used, ground-based forest inventory data must be collected using standard techniques at local, regional, or national scales. Systematic or random sampling designs (either of the entire forest area or stratified segments) are two broad techniques used to collect data that allow mean biomass to be estimated with low uncertainty.

Stratification of sampling with broad forest types can greatly increase the efficiency of surveys by ensuring that major variations are captured. These approaches are well established within the forestry community in most developed countries and can be readily adopted in tropical regions if access, cost, and institutional infrastructure issues are resolved. However, despite the availability of numerous methodologies for quantifying forest biomass in tropical regions

from ground sampling, there are still fundamental problems associated with sampling, measurement, and allometric uncertainty that must be addressed by the research community (Chave et al. 2014, Saatchi et al. 2015, Ngomanda et al. 2014, Lima et al. 2012, Chen et al. 2015, Katerrings et al. 2001).

### 5.2.1.1 Statistical Sampling

The conventional methodology for estimating the forest AGB in any location relies on statistical sampling approaches and is recommended by various protocols and guidelines for GHG inventory in forestlands (IPCC Chapter 4 2006). These sampling techniques have been used in most NFI systems in developed countries and include systematic random sampling approaches, as in examples of U.S. forest inventory data (Fig. 2.1) (Heath et al. 2011), Swedish NFI (Reese et al. 2003), Finland NFI (Tomppo et al. 2011), Canada NFI (Stinson et al. 2011), and China NFI (Zeng et al. 2015). A concise summary of the sampling designs in European countries can be found in the literature (Tomppo et al. 2011, Lawrence et al. 2010). Most of these countries use either detached field sample plots or clusters of plots, and there are variations in sampling density and the associated uncertainty. The forest area represented by one plot varies from 50 ha in the Walloon region in Belgium to about 2,500 ha in the U.S. and 267,700 ha in Canada. There is also quite a high level of diversity in estimation methods and the use of tree allometry based on the measurements. The reports in Tomppo et al. (2010) present more detailed descriptions of these countries' inventory methods and changes in the designs (Zeng et al. 2015).

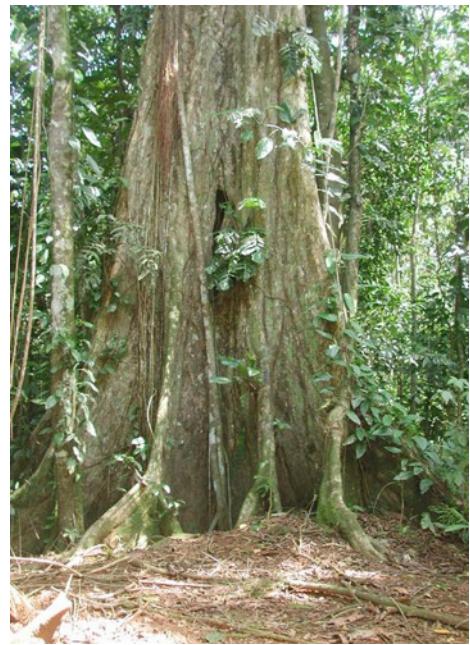
An approach similar to the NFI systems for boreal and temperate forests can be applied to tropical countries with the additional consideration of diversity of species, structure, and requirements for precision of estimates. The sample size and the shape and the configuration of the samples will be an important element in creating a probabilistic sampling design at the national or regional scale. Large plots and a higher number of samples provide more precise AGB estimates at the national or subnational scales. However, other factors such as the degree of difficulty in establishing large plots in complex terrains, costs,

and the time associated with field surveys significantly contribute to the choice of sampling size and configurations (McRoberts et al. 2013).

### 5.2.1.2 Inventory Measurements and Biomass Allometry

Inventory has a long history from tree-based size and density measurements for harvesting and timber extractions. In general, trees are constrained in their geometry and display striking regularities in their structures. These regularities allow tree diameter measurements to be transformed into other variables of interest. There are two prevalent explanations for these regularities: One involves the mechanical strength required to support standing wood structures, and the other involves the constraints of transporting water up through a tall structure composed of hollow tubes. Trees essentially respond to both of these constraints by developing a complex but regular architecture that can be characterized in either case by the use of statistically calibrated equations known as "allometric equations." Also, tree diameter can be related to other attributes such as total tree mass, the area of a tree's foliage, etc., by allometric equations (West & Brown 2005, Chave et al. 2005).

Most trees do not grow symmetrically over their lifespans. Small trees have a disproportionately larger amount of leaves and less woody tissue than large trees (Hallé & Oldemann 1975, Hallé et al. 1978). Structural models based on tree size and mechanical strength were derived for engineering problems for constructing ships where diameter, height, and type of wood were used to calculate the mass. In forestry, similar type measurements have been used to quantify the size of trees and the density of the wood for logging and commercial use of wood. An allometric relationship can be found between tree height and sapwood area that scales isometrically, on average, with the tree trunk cross section. This relationship varies as a consequence of morphological and ecophysiological species-specific responses to different habitats and hydraulic constraints. However, it will ultimately converge on an approximately two-thirds scaling rule as the size of the tree increases (McMahon 1973).



**Figure 5.5** Trees with complex structure associated with tree buttress. Photo by Sasan Saatchi, Costa Rica, 2007.

The allometric models are developed for each forest type and are based on empirical relationships between mass and tree diameter and height. However, these empirical relationships are difficult to obtain logically, particularly for remote locations and tropical forests. Most calibrations are sparse with respect to data on larger diameter trees. Since the equations are fitted to the data using a log-transformed model, the errors associated with the larger diameter trees are very large (Chave et al. 2005, Chambers et al. 2001). In mature natural forests, a large percentage of the total mass is associated with the largest trees, so this is potentially a significant source of error and bias (Shugart et al. 2010).

### 5.2.2 PRACTICAL GUIDE FOR PLOT DESIGN AND SAMPLING

Several guidelines exist for designing plots for forest or general vegetation inventory and for structure and biomass characterization. It is recommended that interested readers consult with documents such as the [RAINFOR protocols](#) for plot design and measurements and [Winrock International](#). The documents can be downloaded from the following links:

- <http://www.rainfor.org/upload/ManualsEn>

[glish/RAINFOR\\_field\\_manual\\_version\\_June\\_2009\\_ENG.pdf](http://www.rainfor.org/ENGLISH/RAINFOR_field_manual_version_June_2009_ENG.pdf)

- [https://www.winrock.org/wp-content/uploads/2016/03/Winrock\\_Terrestrial\\_Carbon\\_Field\\_SOP\\_Manual\\_2012\\_Version.pdf](https://www.winrock.org/wp-content/uploads/2016/03/Winrock_Terrestrial_Carbon_Field_SOP_Manual_2012_Version.pdf)

The following guidelines are designed to help in establishing plots for remote sensing, specially SAR biomass estimation applications:

(1) Location. Select the general area of the plot locations for the study area. Depending on the remote sensing applications, the general location may be selected from an area with the following criteria:

- Reasonably homogenous soil parent material and soil type
- Adequate access
- Reasonably sloped or flat terrain to avoid complex plot establishment and difficult of relating it to radar or lidar data
- Sufficient long-term security from human disturbance
- Sufficient long-term institutional support in case of permanent and monitoring plots
- Avoid areas that have not had frequent disturbance, particularly if the plots are used for developing models for remote sensing mapping, or calibration and validation of remote sensing products

(2) Sample design also depends on the application.

For most inventory applications, the landscape is divided based on some stratification of vegetation type, soil, or topography; and the samples are designed to represent the structure of each strata. Within strata, plots should be randomly located, to avoid ‘majestic forest’ bias and provide statistically unbiased estimate of the structure and biomass for each strata. If maps are available, plot location should be randomly assigned prior to going to the field. If not, in the field, the position of the plot starting point can be randomized by locating it in a random direction at a random distance of the original location.

(3) Plot Size, Shape, and Orientation. Sample plots can be designed in a variety of size, shape, and orientations depending on some trade-off between accuracy, time, and cost of measurements. In addition, the vegetation type and the terrain

may also influence the choice of plot characteristics. Different requirements for plots were discussed and presented earlier. The guidelines here will cover the plot size and shape for both the ground-estimation of biomass and for remote sensing data analysis.

• Plots can be circles, squares, or rectangles. Experience has shown that small circular plots are more efficient because the actual boundary around the plot does not need to be marked. But these plots are often used for national inventory and may not be used to represent remote sensing pixels. Circular plots are easy to establish when they are small. Large circular plots are difficult to establish on the ground because of uncertainty in delineating the boundary. Rectangular plots are also easy to establish and depending on the size of the rectangle and its orientation, the plot can be easily matched with pixels. If the rectangular plots are elongated in shape when laid out on the ground, they may significantly longer edges than circular plots that may introduce errors in number of trees.

• The choice of plot size also depends on the application or remote sensing data, the accuracy of biomass estimation, and the type of forests. For SAR studies, large plot size >0.25 ha or >1-ha depending on the SAR pixel size and speckle noise is recommended. It is possible to calculate the appropriate plot size specifically for each project; however, this adds an additional complication and an additional effort to the process. The size of trees and the diameter threshold of trees may also influence the plot size. It is possible to calculate the size of the plot based on precision and effort and the application. Prior to initiation of plot measurement, it is recommended that limited sampling take place to determine the size of the largest trees. In a land cover stratum with few trees greater than 50 cm dbh, the minimum stem diameter measured within the largest nest may need to be adapted. For non-forest, savanna, and woodland strata, nest plot sizes, and stem diameter sizes will

need to be delineated.

- There are also nested plots that may help to have large plots and a cost efficient approach in collecting tree measurements. Nested plots are composed of several plots (typically 2 to 4, depending upon forest structure) plots and each plot in the nest should be viewed as being a separate plot. According to Winrock guidelines, in ecosystems with low structural variation, such as single species, even-aged plantations, or in areas where trees do not exist, a single plot can be effectively used.
- For orientation, N/S and E/W directions for the principal axes of the plot are the most convenient and also most compatible with the remote sensing pixel comparison. Note that when establishing plots using GPS, record the true or magnetic north to be able to accurately delineate the boundaries of the plot in the remote sensing imagery.

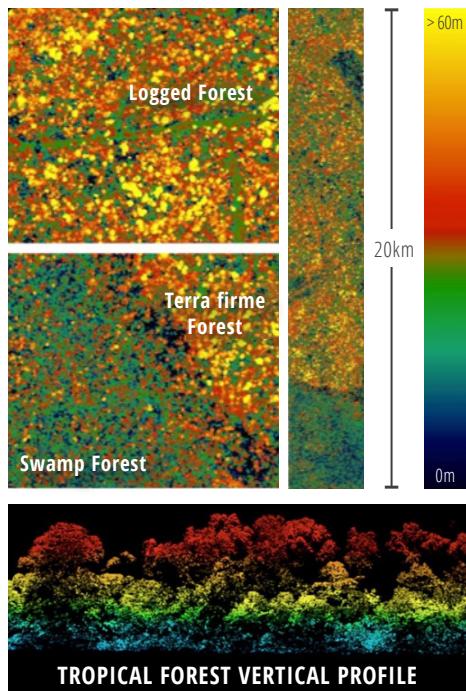
(4) Topography may impact the plot size and orientation in the field. It is important to record the planar distance if used to set up the plot and the angle of the slope. These values will allow calculating the area of the plots established on sloped terrains.

(5) Measurements in the plots also depends on the size of trees and the type of vegetation. However, in general the measurements should include:

- The size of trees (diameter, height, crown size, etc.), identification of tree species for quantifying their wood density or specific gravity from existing data or measurements of wood density (see for example measurement protocols by Jerome Chave: [http://www.rainfor.org/upload/ManualsEnglish/wood\\_density\\_english\[1\].pdf](http://www.rainfor.org/upload/ManualsEnglish/wood_density_english[1].pdf)).
- Plot dimensions and location by using GPS units. Latitude/longitude, among other measurements for the plot geometry and location will be elevation, bearings of plot boundaries, and local landmarks to assist plot relocation. It is recommended that GPS measurements include several plots along different axis of the plots (e.g., GPS for every 20 m within the plot for a 1-ha plot (100 m x 100 m) to increase the accuracy of plot location, size, and orientation.

## 5.3 Forest Biomass – Lidar Remote Sensing Inventory

AIRBORNE LIDAR CANOPY HEIGHT MODEL

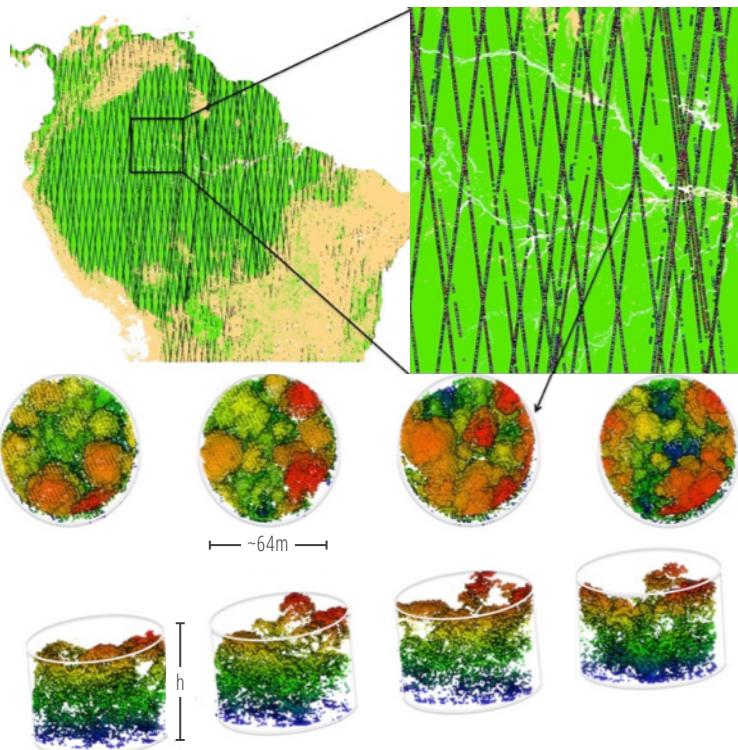


**Figure 5.6** Example of forest canopy height measured by airborne lidar over old growth, degraded, and swamp forests of the Congo Basin in Democratic Republic of Congo (data from WWF/UCLA Carbon Map and Model Project).

### 5.3.1 LIDAR FROM AIR AND SPACE

Airborne lidar measurements can be used for both mapping and sampling inventory of forest structure, as in most national inventory techniques (**Figure 5.6**). This is mainly due to the accuracy of high-resolution airborne lidar measurements for measuring tree height, vertical structure, and horizontal distribution of tree crowns and gaps (Ferraz et al. 2016). For airborne sensors, a significant area over the landscape (100–10,000 ha) can be readily mapped at about 1-m spatial resolution (Asner et al. 2010).

Capable of acquiring elevations with centimeter-level accuracy, small-footprint airborne lidar has had a revolutionary impact on 3D imaging of the Earth's surface and forest structure. More commonly, small-footprint airborne lidar sensors have been employed to detect vegetation and describe the canopy structure



**Figure 5.7** GLAS lidar measurements across tropical forests showing systematic sampling of forest vertical structure at large footprints suitable for estimating AGB for each sample location.

for applications such as habitat modeling, forest inventory, and biomass studies. Airborne small-footprint (<1 m) lidar measurements are mainly discrete-return or waveform sensors working in near-infrared (1,064 nm) wavelengths and flying at low altitudes, depending on the presence of cloud and lidar measurement requirements. Other new lidar technologies working in different optical wavelengths and photon counting capabilities are available for a combination of applications (Moussavi et al. 2014). Small-footprint lidar records multiple of points for each unit area ( $1 \text{ m}^2$ ) with high precision of the altitude of each point within the canopy, allowing a detailed measurement of the forest vertical profile. The airborne sensors are widely available in tropical regions and can be used to acquire data over significant areas either for wall-to-wall coverage (Mascaro et al. 2011b, Meyer et al. 2013) or as inventory samples for regional and national carbon assessments (e.g., BioREDD in Colombia, the World Wildlife Fund (WWF) program in the Democratic Republic of the Congo (DRC), lidar inventory in Brazil, and the NASA Carbon Monitoring System (CMS) program

in Kalimantan). These airborne lidar inventory samples are all based on a Verified Carbon Standard (VCS) VT0005 methodology tool developed by Sassan Saatchi in Colombia and certified by Terra Global Capital (Tittmann & Saatchi 2015).

Existing spaceborne lidar technology works at only large-footprint (25- to 80-m radius) elliptical or circular plots over the landscapes along orbital tracks or sensor beams, providing a systematic sampling of forest structure (Lefsky 2010). In this case, the density of samples will increase as the satellite's orbit drifts along the Earth's surface. Large-footprint lidar measurements have the advantage of being treated as a plot including a large number of trees and being matched with ground measurements for relating the sensor forest height measurements to AGB.

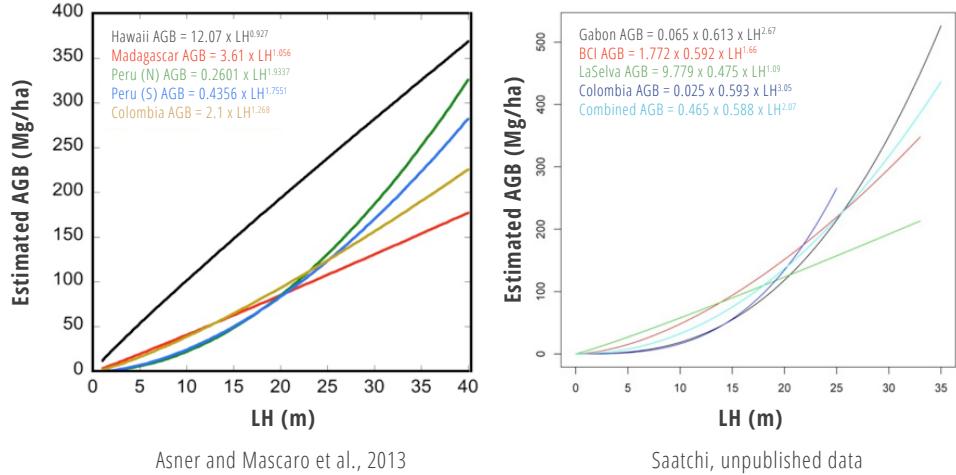
Data acquired over global forests in 2003–2008 from the Geoscience Laser Altimeter System (GLAS) on board the Ice, Cloud, and land Elevation Satellite (ICE-Sat) provided millions of footprints that can be treated as inventory samples (**Fig. 5.7**).

These footprints have an average size of approxi-

mately 0.25 ha (0.16–0.5 ha) spaced at about 172-m intervals along the orbits over the landscape (see **Fig. 5.7**). The GLAS lidar samples do not follow any *a priori* design, as they randomly capture different forest types and provide a reasonable set of data to be treated as forest inventory. A series of studies using GLAS data have successfully demonstrated GLAS data capabilities for estimating forest canopy heights (Lefsky et al. 2007, Rosette et al. 2008) and forest biomass (Lefsky et al. 2005, Nelson et al. 2009, Neigh et al. 2013). The studies consider the statistical nature of GLAS shots and the potential spatial correlations of samples for estimating regional mean and variance of forest structure or biomass (Neigh et al. 2013, Næsset et al. 2011, Saatchi et al. 2011a, Baccini et al. 2012).

### 5.3.2 LIDAR BIOMASS MODELS

Allometric models for converting lidar measurements of forest height or vertical structure into AGB have been developed for different forest types globally (Næsset et al. 2010, Nelson et al. 2010, Asner & Mascaro et al. 2014). These models are often in the form of power law and based on one or several lidar height metrics (Drake et al. 2002). The most common models use the mean top canopy height from small-footprint lidar or a height metric such as the height of the median energy (HOME) or percentile height from large-footprint lidar from air and spaceborne sensors (Asner & Mascaro 2014, Drake et al. 2002). Similar to ground estimation of AGB, the allometry models may vary from location to location, capturing differences in the tree growth and diameter height allometry of forests. Some examples of allometric model variations show significant variations in height to biomass models (**Fig. 5.8**). The use of multiple height metrics derived from the pseudo-waveforms from either small-footprint lidar or large-footprint lidar sensors can contribute to improving biomass estimation uncertainty over larger regions (Meyer et al. 2013, Saatchi et al. 2011, Neigh et al. 2013, Andersen et al. 2014). However, so far there is no universal model to convert the lidar height measurements into AGB on a continental scale, and by acquiring data in different forest types and calibrating the lidar data with ground forest inventory plots, new models are being developed.



**Figure 5.8** Examples of lidar biomass allometric models used in converting airborne lidar data to AGB. Variation across models suggests that the lidar models focused on one parameter only may vary significantly for different forest types, similar to ground allometric models

### 5.3.3 PRACTICAL GUIDE FOR PRODUCING LIDAR AGB MAPS

Lidar-biomass models are developed from ground plot level estimates of biomass and lidar height metrics. The following six steps must be considered in the model development:

- (1) Relation between ground estimation of biomass and lidar height metrics depends strongly on the plot size. For developing models, the plots sizes have to be large enough to include a large number of trees (50–100) such that the mean biomass density estimate of the plot from the allometric model has low uncertainty.
- (2) Depending on the forest types and size of trees, the plot size may vary. For boreal forests dominated by conifers, plots of >0.1 ha may contain enough trees and have accurate ground estimates of biomass. For tropical forests, plot sizes must be larger than 0.25 ha to guarantee the presence of enough trees for ground estimates of biomass with low uncertainty and lidar metrics that represent forest structure at a scale much larger than the crown of a large tree.
- (3) The shape of the plots may also influence the accuracy of the lidar-biomass models. Square plots are recommended as the best options for most forest types, because square plots of any size are easy to establish and have smaller edge lengths compared to rectangular plots. Circular plots are difficult to establish unless they are small, particularly in tropical forests.
- (4) Models developed from small plots may introduce large bias in biomass estimation (see **Fig. 5.9**) due to edge effects and large variations of biomass at small scales that cannot be explained by forest height only. This is particularly the case in unmanaged forests in temperate and tropical regions (Chave et al. 2004, Meyer et al. 2013).
- (5) The height metrics used in developing a lidar-biomass model are important in large-scale applicability of the model. It is recommended that models are developed with height metrics that remain strongly related to AGB across the landscape when the forest structure varies due to variations of soil type and moisture, topography, and various levels of successional stages. For example, the mean top canopy height (MCH) is shown to be a robust metric for capturing the biomass variations across the landscape (Asner & Mascaro 2014, Meyer et al. 2013, Lefsky 2010). MCH from small-footprint lidar has not only information about the height of trees within the plot but also carries information about the vertical structure of the forest.

mation about gaps and spatial extent of tree cover. Theoretically, MCH includes the average of tree heights or crown areas within an area and therefore shows strong correlation to basal area, and hence AGB. The equivalent of MCH in ground measurements is not the mean height of trees but the basal area weighted height of the trees within the plot, the so-called "Lore's Height of forest plot" (Lefsky 2010, Saatchi et al. 2011a).

- (6) The form of the model may also become important in biomass estimation and error assessment. In most applications, the use of a power-law between the AGB and the height metrics provides the most reliable model for converting forest structure to biomass. A power-law or model also appears to be used extensively in allometric models developed from tree harvesting (Chave et al. 2005,

Brown et al. 2001). The use of a power law or logarithmic model between AGB and forest height metrics derived from airborne lidar data is recommended.

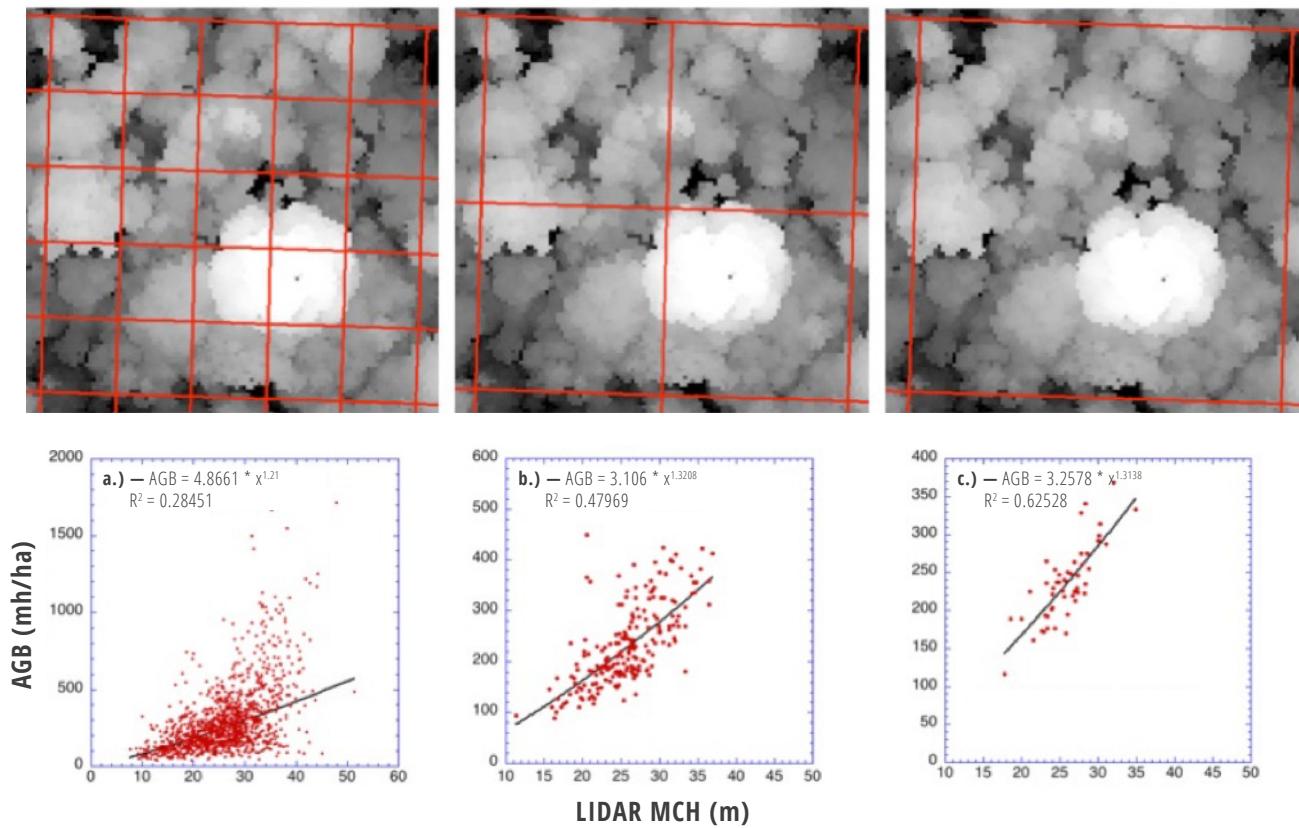
## 5.4 SAR Remote Sensing of Forest Biomass

SAR backscatter measurements are sensitive to vegetation AGB. Observations from a spaceborne SAR can thus be used for mapping AGB globally. However, radar sensitivity to AGB values changes depending on the wavelength and geometry of the radar measurements and is influenced by surface topography, structure of vegetation, and environmental conditions such as soil moisture and vegetation phenology or moisture. All algorithms or models used to estimate AGB from SAR measurements must account for all variables that impact SAR measurements. This

section provides a discussion of the overall sensitivity of radar backscatter to AGB to assist users in choosing the best combination of frequency, polarization, and incidence angles to develop AGB estimation models or algorithms. The impacts of forest structure spatial variation and errors associated with the geolocation of the plots used to relate the backscatter to biomass, the radar measurement geometry, and speckle noise all are important factors that influence radar sensitivity to forest structure and AGB.

### 5.4.1 RADAR SENSITIVITY TO FOREST STRUCTURE AND BIOMASS

Radar observations of vegetation have been studied for more than four decades, both theoretically and experimentally (Ulaby et al. 1982, Tsang et al. 1985, Ulaby & Dobson 1989, Cloude 2014). These studies have shown that the radar measurements depend strongly on the structure, dielectric proper-



**Figure 5.9** Ground plots of different size and lidar-derived models with MCH in tropical forests of Barro Colorado Island in Panama. The plots under the 1-m resolution lidar data suggest that at scale of 20 m x 20 m, (a) there is large bias in the model but gradually at areas of 50 m x 50 m, (b) 100 m x 100 m, and (c) the model improves, and the estimate of biomass can be done without significant bias.

ties of vegetation components, and underlying soil surface depending on the frequency of the operation (Saatchi et al. 1994, Saatchi & McDonald 1997, Ulaby et al. 1990). The soil is most commonly described as a homogeneous medium having a complex dielectric constant that is a function of the volumetric soil moisture, as well as the soil texture, temperature, and bulk density; several empirical models exist for this relationship (Dobson & Ulaby 1986, Hallikainen et al. 1985, Mironov et al. 2004, Peplinski et al. 1995). Studies of soil surface scattering and soil moisture remote sensing at L-band have shown that surface scattering can be expressed in terms of soil dielectric constant at the top 5 cm and the surface roughness characteristics in terms of Root Mean Square (RMS) roughness height and spatial correlation length (Fung et al. 1992). In most SAR-related models for the remote sensing of soil surfaces, it is assumed that the effect of the spatial correlation is reduced significantly during the SAR azimuthal processing and multi-looking, and that the radar signature sensitivity to soil surface RMS height variation remains as the dominant surface structure influencing the surface scattering (Oh et al. 1992, Shi et al. 1997, Dubois et al. 1995, Baghdadi et al. 2002, Bryant et al. 2007). Other landscape features such as directional row or tillage may impact radar cross sections at 100-m spatial resolution but are assumed irrelevant in natural vegetation such as forests and shrublands.

In general, the radar-transmitted energy, in the form of an electromagnetic pulse, penetrates into the forest canopy and reflects back from forest components such as leaves, branches, stems, and underlying soil. Knowing the magnitude of transmitted and received energy, a physical relationship based on electromagnetic theory has been developed to relate the ratio of these energies to properties of the forest. The radar measurements are performed in different frequencies or wavelengths, each providing a different penetration into the vegetation and soil and sensitivity to vegetation biomass.

The measurements are performed in a combination of transmit and receive polarizations (Horizontal (H) and Vertical (V)) at an off-nadir incidence angle and at a spatial resolution projected on the radar range direction. Therefore, radar backscatter sensi-

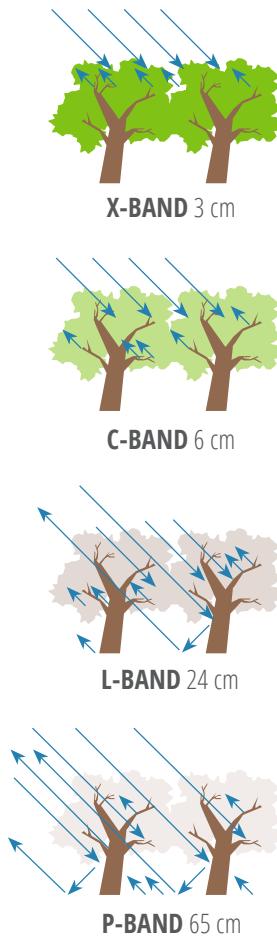
tivity to AGB at any frequency and polarization combination (e.g., HH, HV, VV) depends on two sets of parameters: (1) measurement geometry (such as incidence angle and location and size of the image pixels with respect to the size and the orientation of ground plots) and (2) forest structural parameters (such as the size (volume) and density of trees (number per resolution cell), orientation of forest components (leaves, branches, stems), underlying surface conditions (moisture, roughness, and slope)); and (3) the dielectric constant that in turn depends on the vegetation water content or specific gravity (i.e., the wood density) (Dobson et al. 1995, Saatchi & Moghaddam 2000). In the following subsections, the sensitivity of SAR measurements to these parameters are briefly examined, and examples and references for further reading are provided.

This section focuses on radar frequencies that are either operational or will be operational in future, and have strong sensitivity to vegetation AGB. Examples of SAR imagery are provided at C-band, L-band, and P-band frequencies. Among these frequencies, C-band (Sentinel, RadarSAT) and L-band (ALOS, PALSAR) are operational satellites and will be continued in the future for forest biomass monitoring in the L-band NISAR system (launch 2021). In 2022, ESA will launch a P-band SAR mission dedicated directly to monitoring forest structure and biomass globally.

#### 5.4.2 RADAR WAVELENGTHS AND FOREST STRUCTURE

Usually SAR data are acquired at X-, C-, and L-band frequencies for remote sensing of the environment from airborne and spaceborne platforms. Other frequencies such as P-band and S-band have also been used for remote sensing applications but only on airborne platforms, with plans to be implemented for space observations in near future. A P-band sensor has been designed for ESA's future Earth Explorer Biomass mission, and an S-band sensor is ISRO's contribution to the NISAR mission. A summary for typical radar frequencies and wavelengths is shown in [Chapter 2, Table 2.3](#).

Excellent studies have been previously conducted on examining radar backscatter properties from forest areas (e.g., Freeman & Durden 1998, Dobson



**Figure 5.10** Sensitivity of SAR measurements to forest structure and penetration into the canopy at different wavelengths used for airborne or spaceborne remote sensing observations of the land surface.

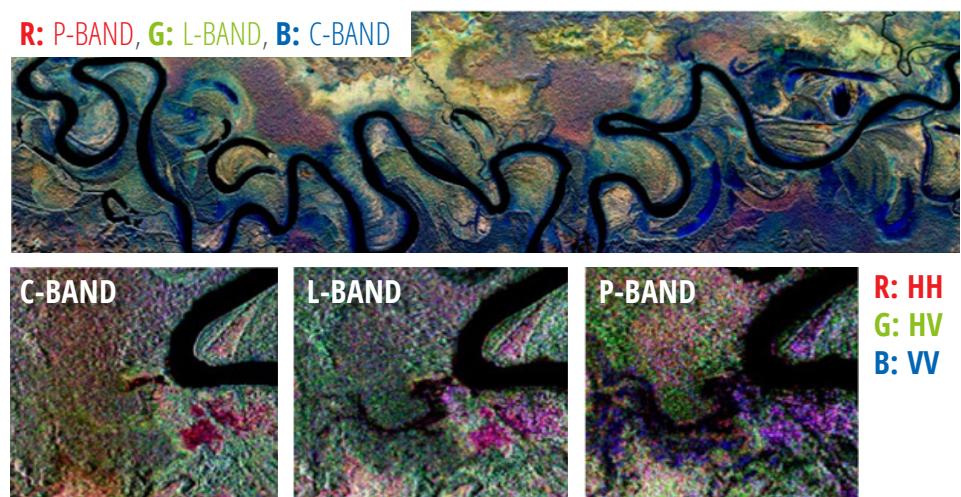
et al. 1992, Ranson et al. 1997). Most scattering occurs when the particles are on the scale of the radar wavelength. Thus, in the case of forests, L-band backscatter arises more from the trunk and the branches of trees, whereas X-band backscatter arises more from their leaves and needles. Also, microwave penetration depth in forests varies depending on the frequency. While L- and P-band can penetrate deep into forests, X-band can get reflection from the canopy level. The backscatter sensitivity to forest components as seen by SAR systems operating at different frequencies is shown in [Figure 5.10](#). For biomass estimation, L-band and P-band sensors are therefore preferred over higher frequencies and smaller wavelengths for two reasons: (1) at these bands, the radar

waves or energy can penetrate the tree canopy and scatter from larger woody components of the forest, and (2) the scattering from larger tree components, unlike leaves, are more stable temporally and remain highly coherent over the acquisition period in the case of repeated measurements for change detection or interferometric applications (Le Toan et al. 1992).

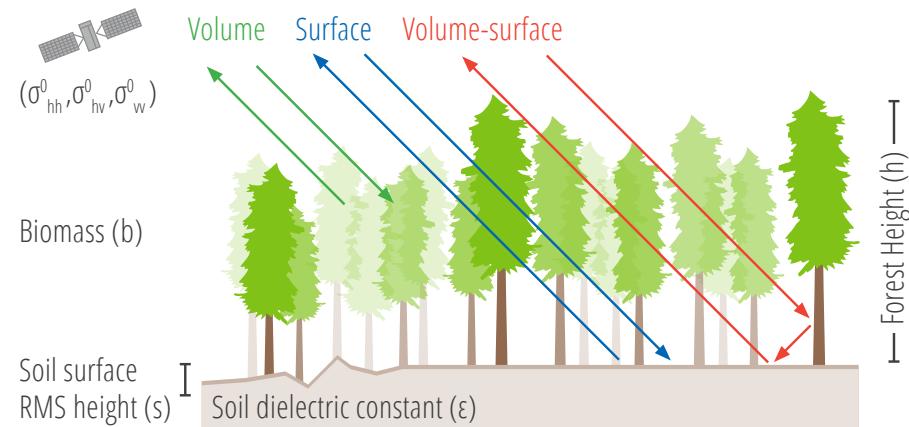
At higher X- and C-band frequencies, SAR penetration will be limited to the upper forest canopy dominated by leaves and smaller branches unless used in very sparse forest covers such as woodlands and savannas. High-frequency radar systems such as Sentinel and RadarSAT operating at C-band and Terra-X SAR operating at X-band provide measurements that are more sensitive to the biomass in low-density forests (e.g., sparse savannas), shrublands, grasslands, or agricultural crops (Wigneron et al. 1999, Saatchi et al. 1994).

Recent studies have focused on the relationship between AGB and radar typically use spaceborne SAR data from ALOS PALSAR (L-band,  $\lambda = 23.62$  cm), and airborne SAR data from both P-band and L-band frequencies (LeToan et al. 2011, Saatchi et al. 2011b).

The radar scattering forest stem and large branches at low frequencies or large wavelengths is considered the main reason radar sensors are used for estimating forest volume and biomass. to trunk and crown biomass and moisture content [16,25]. Past studies have found that the radar backscatter increases with increasing forest AGB from low to medium levels of AGB, but gradually loses its sensitivity to higher AGB levels and asymptotes to a saturation level, resulting in a logarithmic or sigmoidal relationship between AGB and backscatter (Dobson et al. 1992, LeToan et al. 1992, Saatchi et al. 2011). The asymptotic or the saturation level varies based on the radar wavelength and forest type and structure. Results from the airborne AIRSAR (Fig. 5.11) and E-SAR data suggest that saturation may vary between 80 and 150 Mg·ha<sup>-1</sup> for L-band radar (15–30 cm wavelength) and 200–350 Mg·ha<sup>-1</sup> for P-bands, with a wavelength of ~70 cm (Saatchi et al. 2011, LeToan et al. 2011, Mitchard et al. 2009, Bouvet et al. 2018, Villard et al. 2015).



**Figure 5.11** Examples of SAR imagery at C-, L-, and P-band frequencies from the AIRSAR system over tropical forests along the Ja River in Papua New Guinea showing differences of penetration and impacts of forest structure and underlying moisture on SAR false color composite (HH, HV, VV) imagery.



**Figure 5.12** Dominant scattering mechanisms of L-band SAR measurements of forest ecosystems contributing to polarimetric backscatter observations.

#### 5.4.3 RADAR SCATTERING AND FOREST STRUCTURE

The impact of vegetation structure and biomass on SAR data can be investigated by modeling the dominant scattering mechanisms controlling the SAR measurements. A variety of approaches exist for modeling vegetation media, including the characterization of forest vegetation structure, known as scatterers or scattering components such as stems, branches, and leaves in terms of canonical dielectric cylindrical or disk shapes with specified size and orientation distributions. The dielectric constants are assigned to each scattering component to reflect the live wood of trees and leaf material as well as their water content (Saatchi

et al. 1994, Saatchi & McDonald 1997, Saatchi & Moghadam 2000, Yueh et al. 1992, Lang et al. 1983, Karam et al. 1992, Ulaby et al. 1990). The total SAR backscatter from vegetation arises from a combination of scattering and attenuation of individual scattering components that can be represented as a sparse scattering medium (Lang 1981, Saatchi et al. 1994, Chauhan et al. 1994). This approach requires knowledge of tree structure (size, orientation, and density; or equivalently species and biome), dielectric constant, and ground characteristics (RMS height, correlation length, and dielectric constant of soil surface). Figure 5.12 provides a general schematic of the three dominant SAR scattering mechanisms in the forest ecosystems.

The backscattering coefficient measurement by SAR systems can be expressed as the combination of three scattering components (Fig. 5.12): (1) volume (*vol*) scattering, (2) volume and surface interaction (*vol-surf*), and (3) surface scattering (*surf*):

$$\sigma_{pq}^0 = \sigma_{pq-vol}^0 + \sigma_{pq-vol-surf}^0 + \sigma_{pq-surf}^0 \quad (5.1)$$

where *p* and *q* denote polarization of transmitted and received radar signals, respectively, that can be assigned either vertical (V) or horizontal (H) for a linear polarization radar system. The three dominant scattering terms are derived from basic electromagnetic theory by solving Maxwell's equations in a discrete random media (Saatchi & Lang 1989, Lang 1981, Tsang & Kong 1988, Saatchi & McDonald 1997, Chauhan et al. 1991).

There are simpler approaches that only use the Vegetation Water Content (VWC) to provide analytical forms for attenuation and scattering effects. The most common model used in microwave frequencies is the Water Cloud Model, which includes two scattering components from vegetation volume and its underlying ground but ignores the volume-ground interaction (Attema & Ulaby 1978) that becomes dominant in forest ecosystems and for longer wavelength radar observations. Therefore, the Water Cloud Model is mainly applicable at shorter wavelengths (C-band and above) (Matzler 1994, Ulaby & El-rayes 1987) fails to represent the SAR vegetation interaction at longer wavelengths.

#### 5.4.4 SAR POLARIZATIONS AND FOREST STRUCTURE

Transmitted and received radar signals propagate in a certain plane of polarization. Most radars are designed to transmit microwave radiation either horizontally polarized (H) or vertically polarized (V). Similarly, the radar antenna can receive either the horizontally or vertically polarized back-scattered energy, and some radars can receive both. Different combination options for radar polarization (listed below) will provide different image characteristics:

- **Single-polarization**—the radar system

operates with the same polarization for transmitting and receiving the signal

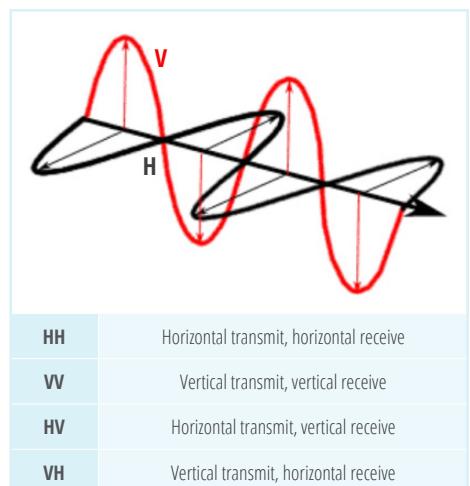
- **Cross-polarization**—a different polarization is used to transmit and receive the signal
- **Dual-polarization**—the radar system operates with one polarization to transmit the signal and both polarizations simultaneously to receive the signal
- **Quad-polarization**—H and V polarizations are used for alternate pulses to transmit the signal and with both simultaneously to receive the signal (Fig. 5.13).

Among the quad-polarization configurations, there are also several variations as in the fully polarimetric measurements that include all components of amplitude and phase of the scattering matrix, and quasi-quad-polarization that includes only the amplitudes and not the phase due to switching the polarizations on different SAR transmit and receive pulses separating the HH/HV measurements from VV/VH (Raney 2007, Hensley et al. 2014).

Polarization is therefore the key characteristic of radar signals propagating into tree canopies or vegetation volume and scatter from individual vegetation components that collectively contribute to the backscatter energy measured by the radar receiver system. Polarization as the orientation of radar wave vectors (at H, V, or any other polarization) interact with vegetation components and backscatter according to the size and orientation of scatterers. For example, a standing live tree with near-vertical orientation depolarizes the incoming waves with different strengths than branches or leaves. Using radars that provide measurements in different polarizations allows separate vegetation with different structures to be reflected in the average size and orientation of different components. The best way to demonstrate this effect is by examining the radar imagery over agricultural landscapes with distinct crop types with uniform shapes and orientations (Fig. 5.14).

#### 5.4.5. CONFOUNDING FACTORS IN RADAR SENSITIVITY TO BIOMASS

The confounding variables that impact SAR measurements and make interpreting those measure-



**Figure 5.13** Electromagnetic waves radiated to the landscape in horizontal and vertical orientations providing different linearly polarized measurements.

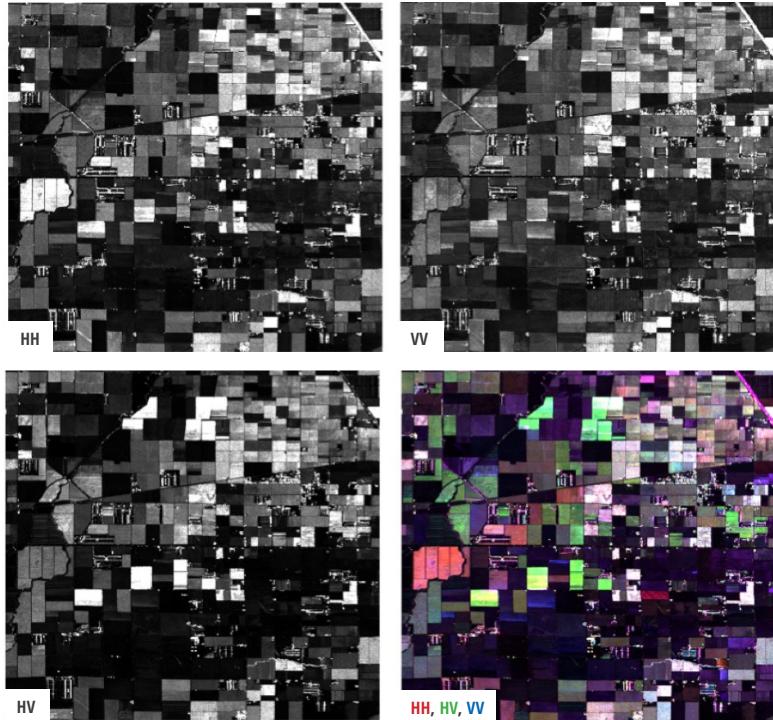
ments ambiguous can be divided into two categories: (1) environmental and (2) geometrical.

##### 5.4.5.1 Environmental Factors

- Two radar backscatter measurements of vegetated surfaces taken from the same instrument using exactly the same characteristics and observational geometry may be significantly different without any changes of the vegetation structure or biomass. The differences may be attributed to surface conditions or environmental changes (Fig. 5.15) between the two radar measurements and must be considered when analyzing the data (Table 5.1).

##### 5.4.5.2 Geometrical Factors

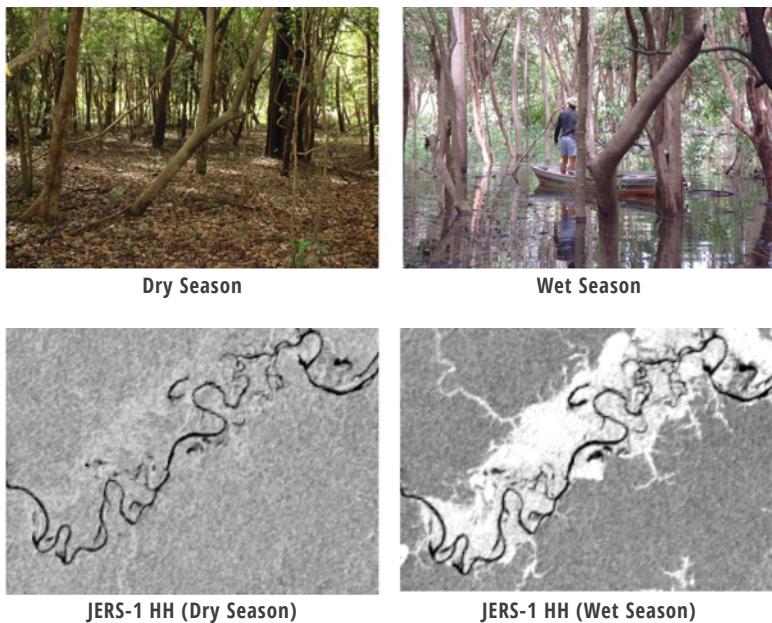
Unlike optical passive and lidar sensors, SAR measurements are performed at an off-nadir look direction, and being an active sensor, both the geometry of the observations and the geometry of the targets (including both vegetated and non-vegetated surfaces) impact these measurements. The surface topography and the orientation of slopes and aspects of the observed surface are perhaps the most dominant effects on the radar measurements. However, other factors such as the orientation of trees, branches, leaves, and



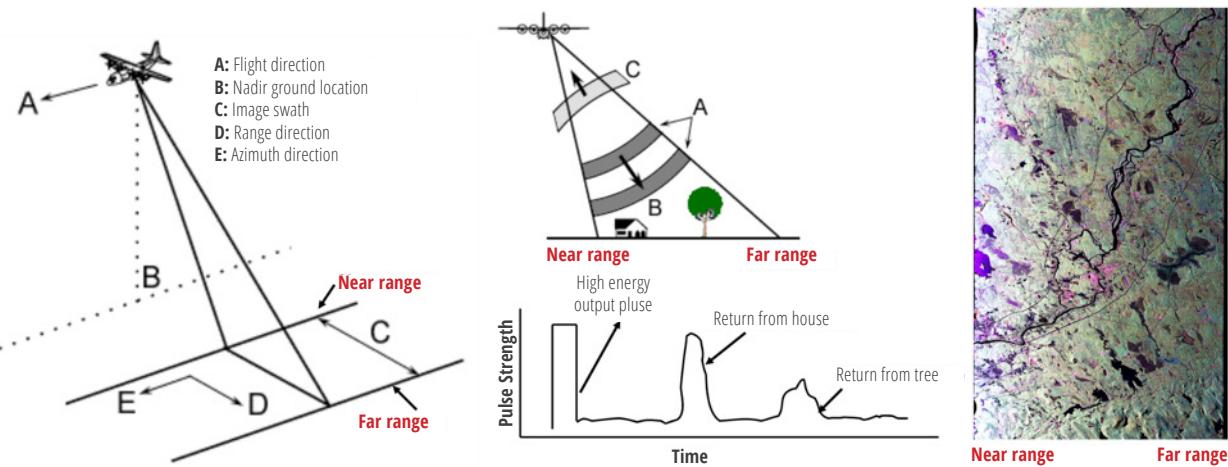
**Figure 5.14** JPL UAVSAR image acquired by L-band radar showing three backscatter polarizations and the false-colored composites over an area in California's Central Valley covered by orchards and different crops. The strength of each polarized backscatter is shown, relatively suggesting how certain crops are relatively higher in one of the HH, HV, and VV polarizations.

VARIABLE	IMPACTS	BACKSCATTER CHANGES
<b>Soil Moisture</b>	SAR backscattered measurement of forests is sensitive to underlying moisture condition and any changes of soil moisture due to precipitation events or irrigation can influence backscatter values.	HH and VV backscatter, significantly and HV to a smaller degree, change with soil moisture depending on the density of vegetation cover. The volume-surface scattering mechanism and direct surface scattering are responsible for changes in backscatter. Similarly, SAR coherence between the data takes is impacted by changes of moisture.
<b>Surface Inundation</b>	Vegetated surfaces, particularly near rivers or in low elevation areas in wetlands, may be inundated seasonally or permanently due to the rise of the water level creating a smooth water body submerging the vegetation at different levels into the water.	Forest inundation increase the backscatter power by a large factor. The increase in power is significant in HH and VV due to volume-surface interactions. HV backscatter may also change due to inundation due to geometry and forest canopy density and the SAR wavelength.
<b>Wind</b>	Presence of wind may change the orientation of the leaves, twigs and small branches with respect to radar observational geometry.	The effects of wind often show up as random differences in the SAR backscatter between observations, introducing noise in backscatter, and reduction of coherence between two SAR observations.
<b>Intercepted Water</b>	After any rain events or early in the morning due to development of dews, there are water droplets on the leaves, causing both scattering and attenuation of the SAR signal.	Depending on the amount of intercepted water or the size of water droplets, and the wavelength, the radar backscatter may increase (at X-band and C-band) or reduce (at L-band and P-band) causing enhanced scattering or attenuation respectively.
<b>Water Content</b>	Changes of water content in trees and leaves from either stress, or diurnally and seasonally due to water loss and recharge of soil moisture impact radar backscatter.	Radar backscatter responds to dielectric constant of vegetation components and therefore the water content. Changes in water content can create significant changes (1-2 dB) in backscatter in all polarizations. Observations of the same time of the day and season can reduce this effect in SAR observations.

**Table 5.1** Summary of environmental impacts on SAR measurements.



**Figure 5.15** Changes of SAR backscatter in wetland forests acquired during the dry and wet seasons showing large backscatter difference due to inundation and an increase in the surface-volume scattering interaction in HH polarization.



**Figure 5.16** Fundamental arrangement and geometry of SAR measurements over the landscape showing (a) the radar look direction, imaging swath and near- and far-range locations, (b) radar pulses and returns across the slant range and the location of targets in the radar image, and (c) a UAVSAR image over mixed boreal forests of northern Maine at L-band polarizations showing the impact incidence angles on backscatter image.

other structures with the respect to the SAR observational geometry may also impact SAR measurements (Outlined in Fig. 5.16, Table 5.2).

## 5.5 SAR Processing Steps for Biomass Estimation

Before biomass estimation from SAR measurements, SAR data must be processed such that the pixel size and geometric attributes and environmental effects are all normalized and radiometrically calibrated. Although it may be possible to include all the SAR processing steps within the biomass estimation algorithm, preparing SAR imagery before algorithm implementation allows for separating the biomass estimation process from the data quality and calibration process.

### 5.5.1 SPECKLE AND IMAGE MULTI-LOOKING

One of the significant differences between active or coherent sensor imagery such as SAR (or laser) to passive sensors (such as that used in Landsat) is the effect of speckle in the spatial resolution of the sensor. Images obtained from coherent sensors are characterized by speckle. This is a spatially random, multiplicative noise due to coherent superposition of multiple backscatter sources within a SAR resolution element. In other words, speckle is a statistical fluctuation

VARIABLE	IMPACTS	BACKSCATTER CHANGES
Incidence angle	SAR measurements are acquired at off-nadir geometry. For each look direction, the radar beam scans the surface over a range of incidence angles. The range of incidence angles is larger for airborne systems (~ 20–70 degrees) but remains confined to only 6 to 10 degrees for spaceborne sensors.	The backscatter of vegetation surfaces vary by a factor of 2 or more from near range (e.g. 20 degrees) to far range (e.g. 70 degrees). If the terrain is topographically complex, the impacts of incidence angle variations will be larger. Often at near range angles the radar backscatter return is larger than at far range, due to the larger path length of radar waves into the vegetation and stronger attenuation.
Surface Topography	SAR's side-looking geometry introduces displacements for tall objects and relief structures. The impacts of surface topography in radar imagery are of three kinds: shadows, foreshortening, and layover (Elachi et al. 1988). Radiometric Terrain Correction (RTC) techniques will help removing/reducing the effects of topography.	The changes of backscatter from surface topography can be significant depending on the slope and aspect of the surface and the incidence angle. Shadows appear dark in the image with very low backscatter. As the incidence angle of an image increases from near-range to far-range, shadowing becomes more prominent toward far-range. Foreshortening can cause compression of features in radar imagery. In the case of layover, the reflected signal from the upper portion of a surface feature is received before the return from the lower portion causing backscatter distortion.

**Table 5.2** Summary of geometrical impacts on SAR measurements.

associated with the radar reflectivity (brightness) of each pixel in the image of a scene. The spatial resolution of a SAR sensor defines the minimum separation between measurements the sensor is able to discriminate and determines the amount of speckle introduced into the system. The higher the spatial resolution of the sensor, the more objects on the ground can be discriminated. The term "spatial resolution" is often confused with the pixel size, which is the spacing of the pixels in the azimuth and ground range direction after processing the data. A first step to reduce speckle—at the expense of spatial resolution—is usually performed during the multi-looking, where range

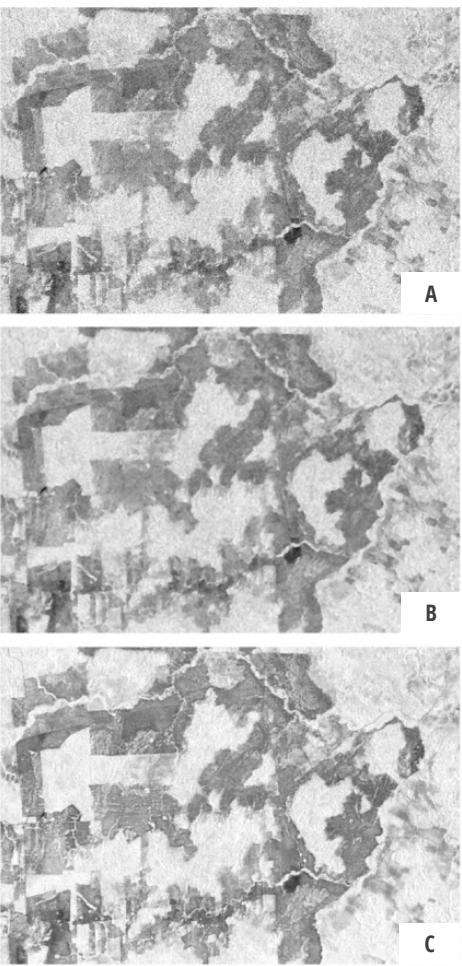
and/or azimuth resolution cells are averaged. The more looks used to process an image, the less speckle there is.

The SAR signal processor can use the full synthetic aperture and the complete signal data history in order to produce the highest possible resolution, albeit very speckled. The data often received from SAR data are in different formats: Single-Look Complex (SLC) or Multi-Look Complex (MLC). SLC image data are calibrated single-look complex files for each polarization (HH, HV, VH, and VV) that are often in floating point format, whereas MLC files are calibrated multi-looked cross products that may be in either amplitude or

power for each polarization and may be provided either as an integer (scaled amplitude) or floating point (backscatter power).

### 5.5.2 SAR PIXEL SIZE CHARACTERISTICS

For this application, the focus is on the multi-looked SAR imagery at pixel sizes that are square and can be readily projected on the ground using the local incidence angle. The user may improve the SAR image quality by further removing the speckle with spatial or temporal averaging at the expense of spatial and temporal resolution of the data. Speckle reduction is particularly important when using SAR data for estimating forest biomass or performing other operations such as classification and image segmentation.



**Figure 5.17** Speckle reduction of SAR imagery from (a) 25-m (5-look) resolution ALOS PALSAR image to (b) 45-look (effective 75-m) spatial filtering to (c) 25-look (5 ALOS images) temporal filtering.

When developing models with SAR backscatter and ground-estimated biomass from plots, the relation is significantly improved when speckle is reduced in SAR imagery. Examples of speckle reduction in imagery and SAR backscatter are shown in **Figure 5.17**.

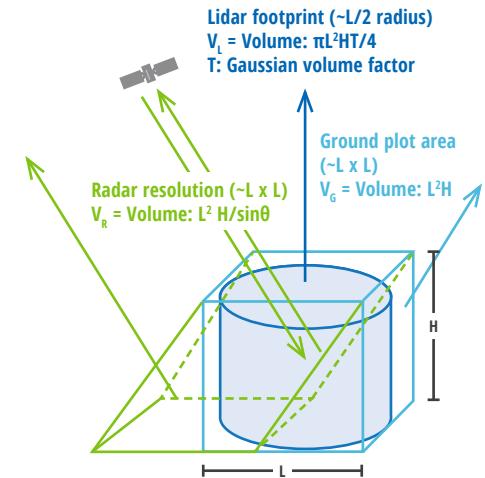
The speckle reduction from spatial averaging impacts the radar backscatter measurements and improves the relationship between the SAR pixel and the ground or lidar measurements. The differences between side-looking SAR pixels and ground plot and lidar pixel are shown in **Figure 5.18**. SAR collects data along a slant range that samples only a slice of the forest medium under the pixel. For bare surfaces without a volume of vegetation, the projection of the pixel on the ground can readily relate the SAR measurements to the surface characteristics. However, in forest ecosystems, the sampling across the volumes always covers a sliced region into the canopy different from the footprint of lidar and the location of the ground plots.

### 5.5.3 SAR RADIOMETRIC CORRECTIONS

For a correct interpretation of backscatter signatures, correcting for the effects of local incidence angle due to topography and normalization for the true pixel area are necessary steps before biomass retrieval. Many studies have shown that uncompensated topographic effects induce a 2- to 7-dB dispersion of the L-band backscatter, which is about the same order of backscatter range used to distinguish forest and non-forest contrast in SAR imagery. The RTC, including the incidence angle normalization, will modify the backscatter values from  $\sigma^0$  (sigma-nought) to  $\gamma^0$  (gamma-nought). As the process of performing terrain correction is covered in other chapters, this section covers the basic information on how to convert  $\sigma^0$  to  $\gamma^0$  according to:

$$\gamma^0 = \sigma^0 \frac{A_{\text{flat}}}{A_{\text{slope}}} \left( \frac{\cos \theta_{\text{ref}}}{\cos \theta_{\text{loc}}} \right)^n, \quad (5.2)$$

where  $\theta_{\text{ref}}$  and  $\theta_{\text{loc}}$  respectively represent the reference angle for the normalization of the backscatter (e.g., the incidence angle at the midswath of the SAR image) and local incidence angle derived from the geometry of radar with respect to the surface topog-



**Figure 5.18** Schematic showing the SAR volume sampling of a forest ecosystem within a pixel in comparison with the ground plot and lidar samples. Differences between the volumes of each sensor are also shown. The difference in sampled area is much larger between SAR and ground or lidar when the pixel or plot size is small or over the topographically complex terrain due to edge effects and sampled areas. At larger pixels (~1 ha), the difference becomes small, and the relation between SAR measurements and ground- or lidar-estimated forest structure and biomass improves.

raphy (slope and aspect).  $A_{\text{flat}}$  and  $A_{\text{slope}}$  represent the local pixel area for a theoretically flat terrain and the true pixel area due to the sloped terrain, respectively. The power  $n$  represents the power of the fit of the angle correction due to radar backscatter variations across incidence angles. For a bare surface, the exponent is equal to 1, but for vegetated surfaces, it can be less than 1 due to variations in scattering mechanisms (volume over slope) originating from canopy gaps and different radar penetration into the canopy. The value of  $n$  may also vary with polarization. However, for simplicity,  $n$  may be considered to be 1 for all polarizations and for most practical cases.

All existing RTC algorithms are based quantifying the local incidence angle and  $A_{\text{slope}}$  over terrain with significant topography. These approaches are based on estimating the local illuminated area  $A_{\text{slope}}$  through either (1) the estimation of the local incidence angle or the projection angle (Ulander 1996) or (2) the integration of the Digital Elevation Model (DEM) (Small 2011, Small et al. 1998). While methods based on

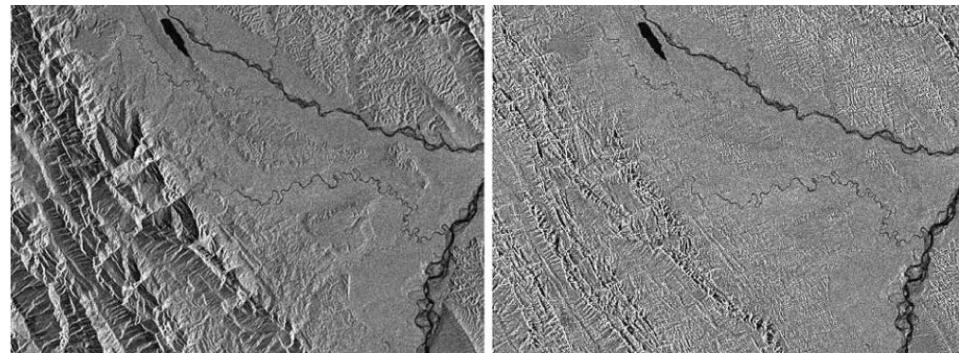
local incidence angle have the advantage of being simpler, methods that include DEM integration have been shown to be more accurate, particularly in steep terrain (e.g. Fig. 5.19). The DEM integration approach involves determining the number of DEM pixels belonging to each radar range and azimuth pixel through knowledge of the geocoding process. It is recommended that users of SAR imagery consult with existing tutorials on terrain correction available on [NASA](#) and [ESA](#) websites.

## 5.5.4 SAR Polarimetric Indices

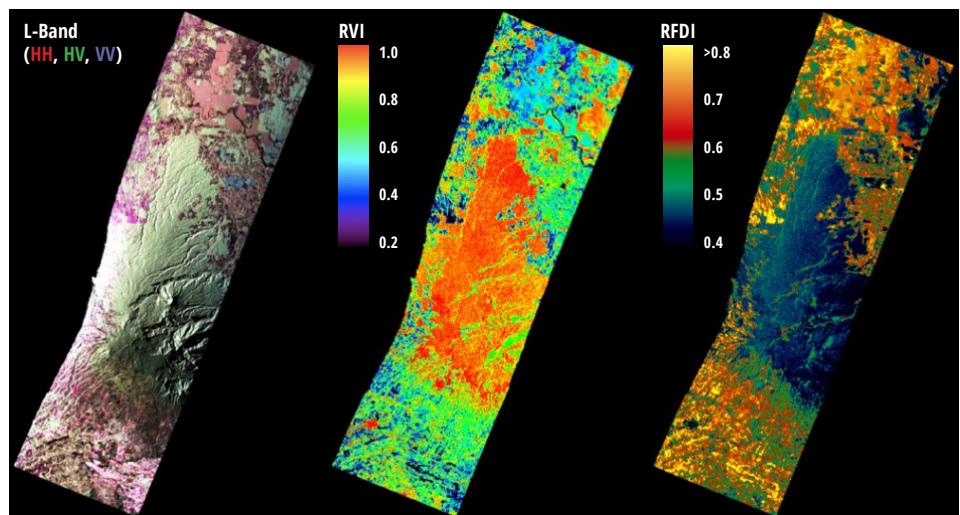
The following section contains a brief discussion on how polarimetric signatures or indices can be used for monitoring forest cover or biomass in different landscapes. Use of signatures or indices are important because they are developed from a combination of radar measurements, which can improve the sensitivity for estimating or monitoring a surface characteristic and can reduce other impacts. For monitoring forest biomass, radar backscatter measurements can be impacted by variations in forest type and structural form (type and orientation), environmental conditions (e.g., moisture and phenology), or radar imaging geometry (e.g., incidence angle and topography). Choosing a combination of polarimetric or radar measurements that can reduce these effects and increase a radar image's sensitivity to forest cover or biomass can be regarded as a reliable monitoring index or parameter. Though there are more complex types that can be developed from either airborne polarimetric systems or from polarimetric interferometric measurements, two simple polarization indices—the Radar Vegetation Index (RVI) and the Radar Forest Degradation Index (RFDI)—are proposed below for monitoring forest types and which can be readily produced from existing satellite SAR systems:

$$RVI = \frac{8\gamma_{HV}^0}{(\gamma_{HH}^0 + \gamma_W^0 + 2\gamma_{HV}^0)},$$

where  $\gamma^0$  represents the radiometrically and geometrically corrected SAR backscattering coefficient for each polarization combination in linear units ( $m^2/m^2$ ). RVI is a ratio of cross-polarization to approximate the total power from all polarization channels;



**Figure 5.19** Examples of SAR imagery (a) before and (b) after RTC over a test site in mountains of Bolivia. A Sentinel SAR image before RTC (left) shows areas that are stretched and compressed due to the topography and geometry of image acquisition. These areas are shown corrected (right) as unstretched and adjusted for backscatter values after applying the RTC from the Gamma algorithm.



**Figure 5.20** UAVSAR L-band polarimetric images and polarization indices over the La Selva Biological Station in tropical forests of Costa Rica showing: (a) three polarized channel color composite showing areas of relatively intact rainforest across a mountain range and low-biomass areas in the northern and southern parts of the image, (b) RVI image showing higher forest biomass areas in red and crops and agroforestry and secondary forests in green and blue, and (c) RFDI image showing more intact forests in dark blue and degraded, secondary, and low-biomass values in lighter blue, green, and red.

it generally ranges between 0 and 1 and is a measure of the randomness of scattering. The RVI is near 0 for a smooth bare surface, increases with vegetation growth, and has an enhanced sensitivity to vegetation cover and biomass. By being a ratio, the RVI has less sensitivity to radar measurement geometry and topography and remains insensitive to absolute calibration errors in radar data.

The RFDI is calculated as

$$RFDI = \frac{\gamma_{HH}^0 - \gamma_{HV}^0}{\gamma_{HH}^0 + \gamma_{HV}^0},$$

where the terms are all radiometrically corrected imagery. However, the ratio can also be used before any radiometric or geometric correction of the SAR imagery. The value of RFDI varies between 0 and 1 because in almost in all conditions, even in most topographically complex terrain, HH remains larger than HV. However, the values of RFDI remain mainly at >0.3 for dense forests, to values of about 0.4 or more for degraded forests, and >0.6 for deforested landscapes (e.g. Fig. 5.20). RFDI can be used with dual-polarization imagery such as the ALOS PALSAR Fine Beam Dual (FBD) datasets.

Using data from the same satellite orbits, the geometry and incidence angle do not vary over SAR pixels, allowing temporal analysis of RFDI without concerns for changes in geometry and incidence angle. In fact, RFDI from satellite imagery such as ALOS PALSAR or Sentinel can be computed without any correction for incidence angle and topography. The main application of RFDI is defined as an index to monitor changes in forest cover due to deforestation and degradation. The low values refer to forests where the effect of volume-surface interaction is either small (e.g., forests with shorter stems and dense canopies) or relatively equal in both channels (e.g., forests over slopes). The high values refer to forests with large differences between HH and HV, suggesting they are open or recently degraded forests, or inundated forests. Theoretically, RFDI can be used at any radar resolution; however, the best spatial resolution for developing RFDI depends strongly on the speckle noise in radar backscatter and the natural heterogeneity of forest structure and gap size variations over the landscape where the contribution of volume-surface interaction is larger in HH compared to HV backscatter. In general, RFDI can be used to detect both the loss of forest cover and its recovery from disturbances resulting from logging or other types of natural or anthropogenic events.

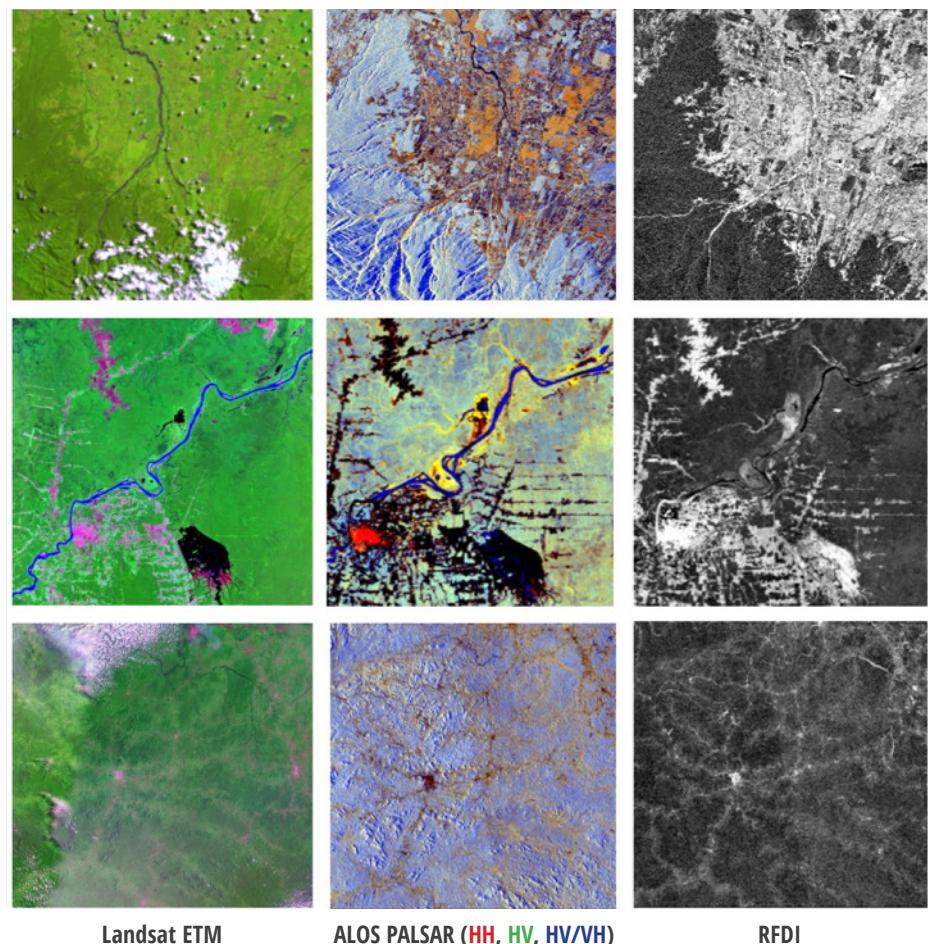
### 5.5.5 PRACTICAL SAR IMAGE PROCESSING FOR BIOMASS ESTIMATION

Five practical approaches for SAR processing before the data analysis for biomass estimation are summarized as follows:

- (1) Download the SAR intensity imagery in any format and create imagery in linear power (not in dB).

Linear power data, which are often provided in floating point, are considered the calibrated radar imagery that can be used to relate to any surface parameter or integrated in the models. Note that working with backscatter values in dB may introduce large statistical errors in the analysis because all mathematical equations and algebraic relations must consider the logarithmic quantities.

- (2) SAR images can be multi-looked (simple averaging) to create images at coarser spatial resolution with reduced speckle. Speckle in SAR imagery can also be reduced by using various SAR filters (Lee



**Figure 5.21** Satellite imagery over three tropical study areas, with Braulio Carrillo National Park in the top row; Rondônia, Brazil in the middle row; and Ebolowa in Cameroon in the bottom row. Images include false color Landsat ETM (RGB:543 bands) (left column), false color ALOS PALSAR images (RGB: HH, HV, HV/HV) (middle column), and RFDI (right column)

INDEX	APPLICATION	INTERPRETATION
<b>Radar Vegetation Index (RVI)</b>	Monitoring vegetation cover, water content, and aboveground biomass with quad-pol or quasi-quad-pol data.	Range (0-1): low values refer to low vegetation cover and water content. The low threshold can be used to separate forest and nonforest.
<b>Radar Forest Degradation Index (RFDI)</b>	Detecting forest degradation and deforestation, biomass loss and gain with dual-pol or quad-pol data.	Range (0-1): Low values refer to high biomass and intact forests. Values change gradually to higher values for degraded and nonforest areas. The values remain independent of topography.

**Table 5.3** SAR vegetation indices used with dual- and quad-polarized SAR measurements.

et al. 1999).

- (3) All SAR images acquired from satellite or airborne datasets must be georeferenced such that each ground-projected pixel has geographic coordinate. Note that multi-looked SAR images with reduced speckle improve the relationship between ground measurements and SAR backscatter. De-

pending on SAR data, the multi-looked imagery can have different resolutions. For example, for ALOS PALSAR data (originally at 20 m with 3 looks), a 100-m image can be regarded as an image of about 75 looks with significantly reduced speckle.

- (4) If the data downloaded do not include RTC, use any commercially or freely available software to

perform terrain correction. NASA's Alaska SAR Facility provides several software applications for SAR processing, including [MapReady](#) for terrain correction. RTC can be applied on multi-looked imagery given the geometry of the SAR observations and the terrain model (such as the Shuttle Radar Technology Mission (SRTM) data). This process includes finding the local incidence angle and the area normalization factors in map (in the final radar ground projection) coordinates for projection angle corrections.

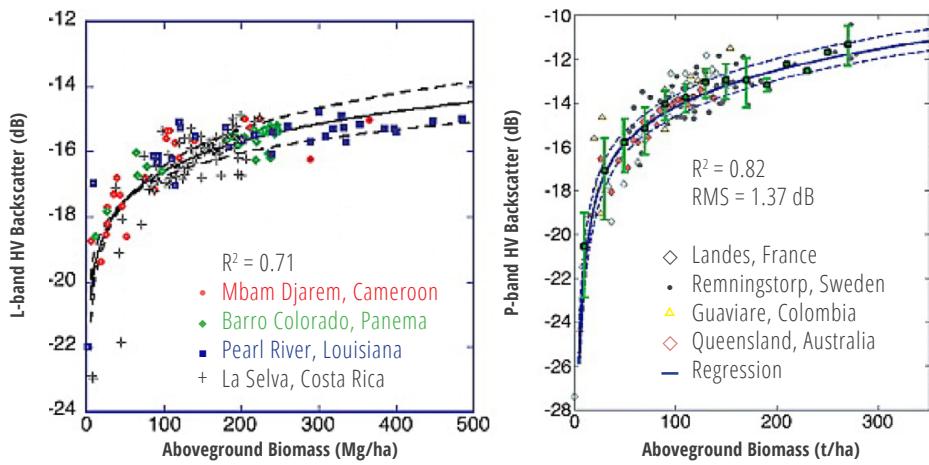
- (5) Develop color composite images from SAR polarizations and SAR polarimetric indices to identify different features of landscape and vegetation covers in color without the impacts of topography and or SAR geometry.

## 5.6 SAR Biomass Estimation Algorithm

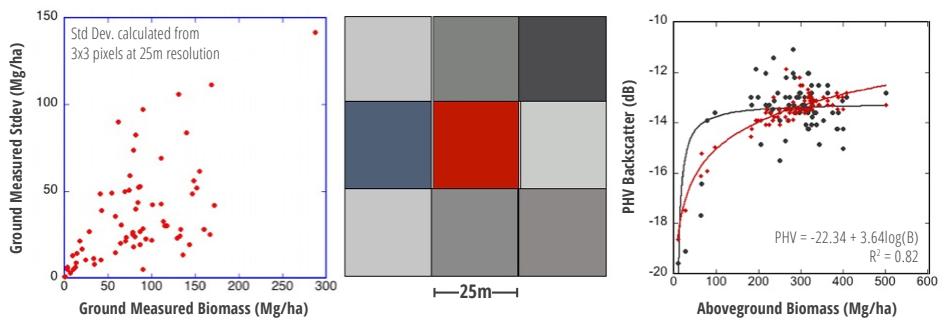
Performing all necessary SAR processing results in SAR imagery that has been corrected for terrain effects through RTC techniques and projected on the ground and multi-looked to a certain pixel spacing (e.g., 25 m–100 m) depending on the original image resolution. Basic techniques are presented here to develop a radar-biomass model or algorithm for estimating vegetation/forest AGB. This section is designed to show: (1) the general relationship between vegetation biomass and radar measurements, (2) development of site-specific statistical models from either airborne or satellite radar data, (3) development of a physically based model that includes different scattering matrix components, and (4) use of machine-learning algorithms for large-scale biomass mapping.

### 5.6.1 GENERAL RADAR-BIOMASS RELATIONSHIP

To demonstrate the relationship between radar measurements and AGB, this section concentrates on low-frequency (large-wavelength) radar systems at the L-band frequency due to its availability from space through ALOS PALSAR and due to its strong backscatter sensitivity to biomass at low- to mid-ranges. However, some results are shown and examples given from C- and P-band SAR imagery, as they are available from Sentinel series and, in the future, from ESA's Biomass mission.



**Figure 5.22** Sensitivity of radar backscatter measurement at L-band and P-band frequencies and HV polarization to forest AGB over sites distributed in boreal, temperate, and tropical ecosystems (Shugart et al. 2010).



**Figure 5.23** Impacts of geolocation error and small plot size on radar backscatter-biomass relationship shown by large variations of backscatter in small plots (25 m × 25 m) (left), the geolocation error causing the correct pixel (red) or plot location being misplaced by one of the surrounding pixels (center), and the impact of the geolocation error on radar backscatter relation and saturation (right). The use of correction location or larger plots will improve the relationship and the asymptote.

As discussed previously, the dominant scattering mechanisms as well as the size (volume) and dielectric constant (moisture or wood density) of forests determine the magnitude and behavior of the backscatter at each polarization to AGB. As a result, the backscatter radar energy at linear polarizations is related to the volume and biomass of forest components (**Fig. 5.22**).

The radar backscatter biomass relationships shown in **Figure 5.22** for both L-band data from ALOS PALSAR (Mitchard et al. 2009, Saatchi et al. 2011b) and airborne P-band data (Le Toan et al. 2011) from several study sites are based on a direct comparison of radar measurements and AGB from ground plots. The form of the relationship suggests that there is a rapid rise of backscatter with biomass for low-biomass plots, and then a slower increase to an asymptote value at higher biomass values. For L-band, the asymptote may arrive

at values of about 100 Mg/ha or more depending on forest types (Yu & Saatchi 2016), and for P-band, the asymptote may reach 300 Mg/ha or higher. Note that both the form of the relationship and the asymptote or saturation values may change significantly depending on the data quality and analysis:

- (1) The plot data used for comparison of radar imagery are small; therefore, the backscatter power may be noisy due to the presence of speckle noise in radar measurements. The noisy data may introduce a false saturation at lower biomass values. Large plots >100 m in size will readily improve the relationship.
- (2) Geolocation errors in both SAR and ground plot locations will introduce noise in the data when comparing the ground plots and the radar backscatter (**Fig. 5.23**). Similar to the plot size,

the geolocation error can introduce false saturation. The use of larger plots and averaging SAR pixels (multi-looking) reduces geolocation errors.

(3) Changes in SAR geometry across the plots may also introduce noise in the data, impacting the relationship and the saturation level of the radar-biomass relationship. If the image is not corrected radiometrically (RTC) and for incidence angle variations, plots with similar biomass may have significantly different backscatter (**Fig. 5.24**).

(4) Environmental differences and changes in soil moisture on radar backscatter can introduce noise in the relationship. Using multitemporal SAR imagery will allow averaging out the moisture effects and will improve the backscatter values, allowing them to become more stable spatially and temporally for biomass estimation (**Fig. 5.24**).

(5) Differences in the time of image acquisition and plot data can also introduce noise in the relationship. If the inventory plots are established a long time before or after the SAR acquisition, changes in biomass and forest structure from both disturbance and recovery during this period will influence the SAR backscatter. It is recommended that the dates between ground and radar acquisitions are minimized.

## 5.6.2 RADAR-BIOMASS STATISTICAL MODELS

Depending on the wavelength of the measurement, the radar backscatter from a forest can be related to scattering from live stems, branches, and foliage based upon their abundance and moisture content within a resolution cell as:

$$\gamma_{pq}^0 \propto f_{pq}(n_i, V_i, \varepsilon_i), \quad (5.3)$$

where  $f_{pq}$  is a function averaged all possible orientation and size distributions,  $p$  and  $q$  represent the transmit and received polarizations (H, V),  $n_i$  represents the density of trees,  $V_i$  represents the volume of trees, and  $\varepsilon_i$  represents the moisture or dielectric constant of forest components (stems, branches, and leaves). This equation symbolically represents the radar backscatter relationship to forest structure and wood density that, along with orientation and tree size distributions, can be used to generate a model for estimating forest volume or biomass. In addition to forest structure and biomass, other parameters such as soil moisture and surface structure (slope and roughness) impact the function. The most important task in radar biomass estimation is the development of the model  $f_{pq}$ .

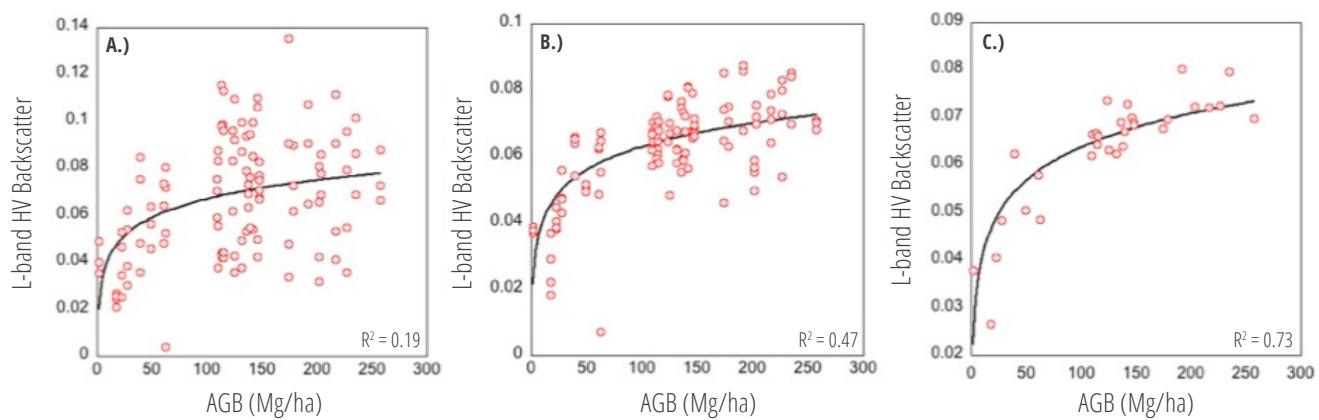
The analogy of this model in forestry applications is a parametric or regression-type relationship designed to directly estimate AGB with respect to forest structure. Here, some options of statistical models are pro-

vided that are developed by fitting a regression-type equation to backscatter relation to AGB. The following is the starting point in the analysis of data:

- (1) The radar backscatter is radiometrically corrected and normalized for incidence angle ( $y_{pq}$ ), and is converted into the linear unit and not in logarithmic scale (dB). To convert radar backscatter to a linear unit,  $10^{(0.1 \text{ dB})}$  is used, where dB refers to the backscatter value in dB, which is often a negative number.
- (2) The plot size is large enough and has good geolocation and shape to match the radar pixels.
- (3) Radar backscatter data are from multi-looked imagery for large enough pixels with reduced speckle noise.
- (4) Plot-level radar backscatter at each polarization or for radar vegetation indices are derived from averaging pixel-based backscatter at the linear unit.

Here, a statistical model is introduced based on the nonlinear combination of radar backscatter for the estimation of biomass. Based on previous studies, the regression model is between an unknown power of AGB and a linear combination of backscatter measurements at three polarizations (Ranson & Sun 1994, Saatchi et al. 2007, Saatchi et al. 2011b):

$$AGB^\lambda = a_0 + a_1 \gamma_{HH}^0 + a_2 \gamma_{HV}^0 + a_3 \gamma_W^0, \quad (5.4)$$



**Figure 5.24** Radar backscatter and biomass relationship influenced by geometry and environmental conditions: (a) radar backscatter and AGB from ground plots acquired over different local incidence angles and different times, (b) correction of radar backscatter for the local incidence angle reducing the noise in the data, and (c) multitemporal averaging of backscatter (over five dates) reduce the noise due to environmental factors and moisture and improve the radar-biomass relationship (data from UAVSAR and ground plots in region of Howland, Maine, U.S.)

where the unknown coefficients ( $\lambda$ ,  $a_0$ ,  $a_1$ ,  $a_2$ , and  $a_3$ ) will be determined statistically by using radar measurements and field data. The above model has the advantage of being flexible and can be used either for a single polarization, such as  $\gamma_{HV}^0$  (Le Toan et al. 2011), or multiple polarization. It can also use other information such as the interferometric height estimation from radar or any other spectral information to improve the prediction of the model. The power  $\lambda$  for AGB is designed to adjust for the asymptote in the model (Saatchi et al. 2011b).

There are other options for AGB estimation from backscattering coefficients that have been used both in temperate and tropical forests (Saatchi et al. 2007):

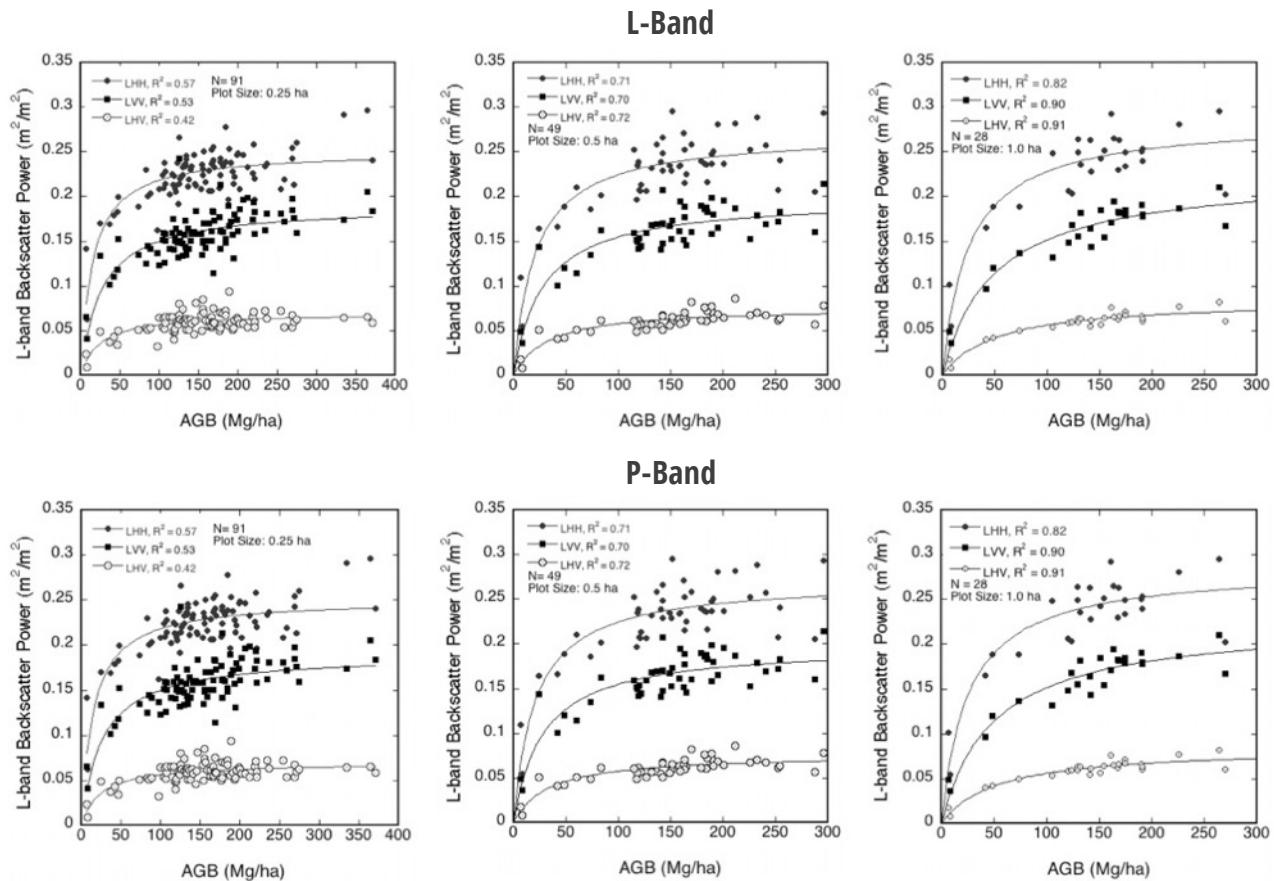
$$\log\log(AGB) = a_0 + a_1 \gamma_{HH}^0 + a_2 (\gamma_{HH}^0)^2 + b_1 \gamma_{HV}^0 + b_2 (\gamma_{HV}^0)^2 + c_1 \gamma_{VW}^0 + c_2 (\gamma_{VW}^0)^2 \quad (5.5)$$

These equations have a larger number of coefficients to determine but may perform better in ensuring that the asymptote in the radar backscatter-biomass relationship is well-represented in the model. **Figure 5.25** shows the model fit of Eq. (5.4) for L- and P-band data over tropical forest of Costa Rica, and **Table 5.4** shows the coefficients derived from the model fit.

The power-law relationship defined in Eq. (5.4) is the optimum fit to the normalized backscatter data with respect to the aboveground dry biomass in all cases. At both frequencies, the scale of analysis did influence the form of the model between AGB and backscatter; however, as the scale of analysis increased from 0.25 to 1.0 ha, the r-squared correlation between backscatter and AGB improved, largely due to the spatial averaging of the radar data and the reduction of speckle noise. The improvement from 0.25 to 0.5 ha is due to both the reduction in speckle noise

and the errors due to geolocation and orientation of the plots, whereas the improvement from 0.5- to 1.0-ha plots is mostly due to averaging a larger number of pixels, hence the reduction of speckle noise. Although all polarizations show similar trends with respect to increasing AGB, there are clear distinctions among them in terms of backscatter level and sensitivity to biomass. In both frequencies, the HV sensitivity to biomass is much higher, and the relationship improves much higher than other channels as the scale of measurement increases. However, at L-band, the sensitivity to biomass decreases rapidly at 100 Mg ha<sup>-1</sup> at 0.25 ha and with slightly higher values of 100–150 Mg ha<sup>-1</sup> at the 1-ha scale.

The P-band results show a very strong relationship to AGB over the entire range with gradual loss of sensitivity at AGB > 200 Mg ha<sup>-1</sup>. The r-squared correlation between P-band channels and AGB is almost above 0.7 in all cases and improves with the spatial scale.



**Figure 5.25** Model fits for the backscatter relationship with AGB for both L- and P-band at three polarizations of HH, HV, and VV at three different plot sizes of 0.25, 0.5, and 1.0 ha (Saatchi et al. 2011b).

RADAR CHANNELS	a0	a1	a2	a3
<b>0.25 ha Scale</b>				
LHH, LHV, LVV	-4.36	$-41.68 \pm 2.56$	$45.71 \pm 6.05$	$2.08 \pm 2.53$
PHH, PHV, PVV	-1.23	$64.11 \pm 6.05$	$235.41 \pm 22.80$	$119.41 \pm 8.73$
<b>0.5 ha Scale</b>				
LHH, LHV, LVV	-1.91	$16.49 \pm 3.32$	$63.76 \pm 11.28$	$39.26 \pm 3.93$
PHH, PHV, PVV	-0.31	$57.96 \pm 6.36$	$313.29 \pm 30.79$	$81.22 \pm 11.07$
<b>1.0 ha Scale</b>				
LHH, LHV, LVV	-0.67	$-7.35 \pm 4.87$	$106.63 \pm 21.96$	$48.11 \pm 6.72$
PHH, PHV, PVV	0.73	$42.13 \pm 13.49$	$323.02 \pm 64.41$	$71.51 \pm 18.74$

**Table 5.4** Parameters of the model fit to Eq. (5.4) using airborne radar backscatter at L-band P-band with three polarization combination and three spatial scales.

Model Parameters	A	B	C	D	$\alpha$	$\beta$
$\gamma_{\text{HV}}^0$	$0.04 \pm 0.01$	$0.06 \pm 0.03$	$0.04 \pm 0.1$	$0.00001$	$0.14 \pm 0.05$	$0.019 \pm 0.02$

**Table 5.5** Fitting coefficients and unknowns for a temperate/boreal mixed conifer forest site in Howland, Maine.

### 5.6.3 RADAR-BIOMASS PHYSICALLY BASED MODELS

Physically based models are often complex in formulation and include a large number of variables covering the remote sensing sensor characteristics, the geometry of measurements, and the forest characteristics. There are several types of physically based models that are used to simulate the radar measurements from forest canopies (Saatchi & McDonald 1997, Saatchi & Moghaddam 2000, Ulaby et al. 1990, Sun & Ranson 1995, Karam & Fung 1983, Karam et al. 1992, Oh et al. 1992). As discussed previously, the model is based on formulating the three dominant scattering mechanisms or radar backscattering power from vegetation layers. These include volume, volume-surface, and surface scattering:

$$\gamma^0 = \gamma_{\text{vol}}^0 + \Gamma_{\text{surf}} \gamma_{\text{vol-surf}}^0 + \Gamma_{\text{vol}} \gamma_{\text{surf}}^0, \quad (5.6)$$

where  $\gamma_{\text{vol}}^0$  is the volume backscattering from vegetation,  $\gamma_{\text{vol-surf}}^0$  is the volume forward scattering from vegetation,  $\gamma_{\text{surf}}^0$  is scattering from the soil surface,  $\Gamma_{\text{vol}} = \exp(-B \times \text{AGB})$  is the volume attenuation (absorption), and  $\Gamma_{\text{surf}}$  is the soil surface reflectivity.

The above equation and terms can be repeated for each polarization separately. Here, the model fits for only one polarization is represented, and the

methodology for developing semi-empirical models that include the physical formulation is provided. The HV polarization typically has better sensitivity to forest AGB than the HH or VV polarizations and has less sensitivity to the soil surface scattering and moisture. Therefore, for demonstration, this chapter focuses on the HV polarization. Another focus is on the L-band data as it appears to be the widely used data from space from the ALOS PALSAR data. However, the methodology can work for both L-band and C-band over different ranges of biomass depending on the sensitivity of each sensor (Fig. 5.26).

In forests where the canopy is not as dense, such as the temperate conifers and boreal forests, the magnitude, sensitivity to biomass, and the contributions of each scattering component may be different. For example, in boreal forests, L-band backscatter can possibly have significant contributions from the surface-volume term with potentially enhanced sensitivity to forest biomass and soil moisture (Sandberg et al. 2011). This possibility is explored by fitting a functional form that includes both volume and the volume-surface scattering term:

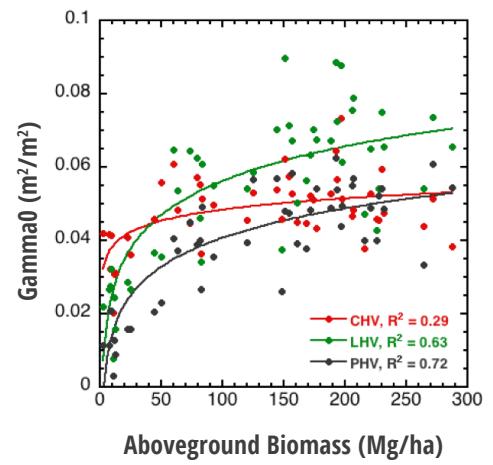
$$\gamma^0 = AW^\alpha (1 - e^{-BW}) + (CW^\beta + D)Se^{-BW}, \quad (5.7)$$

where  $A, B, C, D, a$ , and  $\beta$  are fitting coefficients, and the unknowns are  $W$  (as the AGB) and  $S$  (as the soil surface condition). The term  $S$  includes the reflectivity of the surface which depends primarily on soil surface moisture. The temperate/boreal mixed conifer forest was selected as an example to demonstrate the effect of other scattering contributions, as observations from ALOS PALSAR HV and AGB from ground data are available from the site in Howland, Maine (Table 5.5).

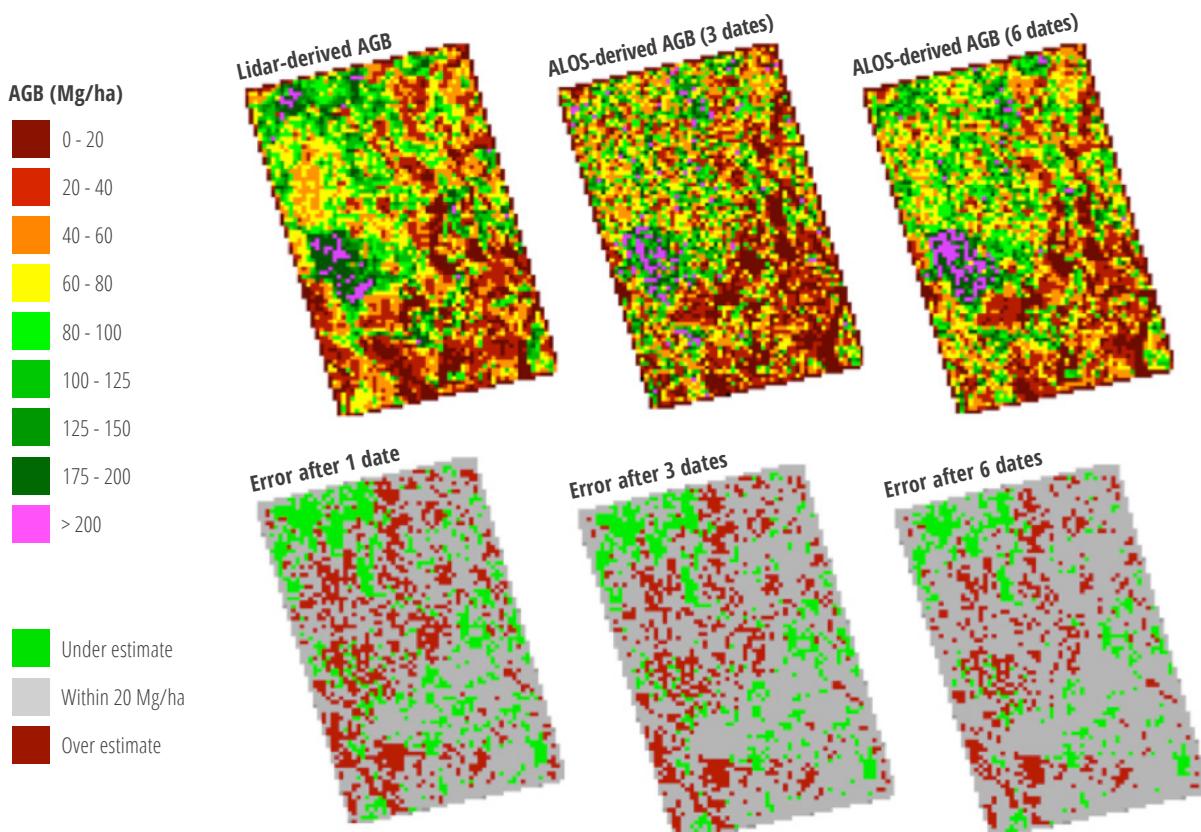
After developing the model for the study site, the next step is to estimate AGB from the model. The biomass estimation process from the model relies on a least-squared approach such that the function can be inverted to estimate  $W$  and  $S$ . If the only data available are from ALOS PALSAR, which provides dual-polarization imagery in HH and HV over most of the world, the least-squared method can be readily written as:

$$\text{Loss} = \sum_{t=1}^n \left\{ \left[ \eta_{\text{HH}} (f_{\text{HH}}(W, S_t) - \gamma_{\text{HV},t}) \right]^2 + \left[ \eta_{\text{HV}} (f_{\text{HV}}(W, S_t) - \gamma_{\text{HV},t}) \right]^2 \right\}, \quad (5.8)$$

where  $\eta = 1$  for HH, or 3.5 for HV. The Loss function will be minimized in a least-squared approach to es-



**Figure 5.26** Relationship of backscatter in HV polarization at three frequencies of C, L-, and P-band based on ground and backscatter data over Howland forest in Maine using biomass estimates from 0.1-ha plots AIRSAR data from one image. The backscatter values are different in magnitude and sensitivity, and the fit is based on a logarithmic model.



**Figure 5.27** Application of the semi-empirical model on the ALOS PALSAR HH and HV images over Howland, Maine forests and comparison with the lidar-derived AGB as a reference map. The results shows using multi-temporal ALOS imagery, the SAR estimation of biomass approaches the Lidar estimation. With 6 ALOS images, over 75% of the image pixels are estimated within 20 Mg/ha of the reference values. The accuracy increases to 86% after using 9 images.

timate  $W$  and  $S$  at each pixel. Note that with multiple SAR imagery acquired at different times  $t$ , the above equation is used by allowing the soil term  $S$  to change through time, but  $W(\text{AGB})$  remains constant to improve the estimation approach. This assumes that for a period of a season or few months to stack multitemporal radar data, AGB remains constant or has some undetectable change unless a major disturbance occurs.

The above formulation was designed to demonstrate that a semi-empirical or physically based algorithm can be derived for an ecoregion or forest type and applied over large areas. This is mainly due to the fact that statistical models developed based on field data over a small region may have large errors when applied over areas far from the original test area due to potential variations in landscape topography, soil moisture and roughness, and vegetation structure. However, the

semi-empirical algorithm relies more on the physics of the problem and compensates for the landscape and regional variation. In a more rigorous implementation of the problem, the Loss function can also be optimized locally by updating the coefficients of the model ( $A, B, C, D$ ) over a local window of pixels ( $3 \times 3$  or  $5 \times 5$  or larger) to allow for the model to be better adjusted to local variations of the forest structure within an ecoregion.

A simpler version of the physically based algorithm has been applied in few cases. Yu and Saatchi (2016) use a model that weights more on the volume scattering and combines the surface effect in both volume-surface interaction and surface scattering as an unknown term:

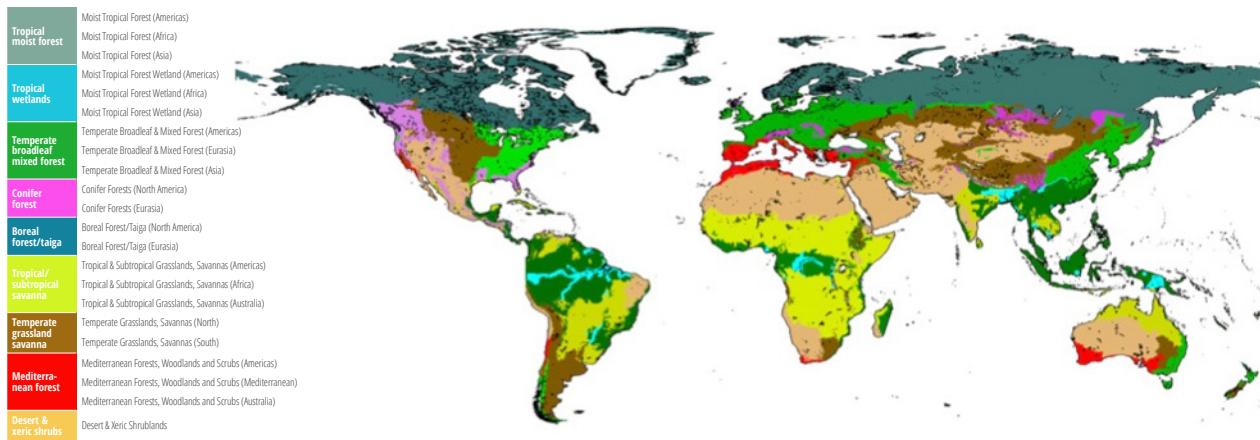
$$\gamma^0 = AW^\alpha(1 - e^{-BW}) + C , \quad (5.9)$$

where  $W$  is AGB in Mg/ha; and  $A, B, C$ , and  $\alpha$  are fitting coefficients. The above equation has been used

to model ALOS PALSAR variations over global vegetation biomass (Yu & Saatchi 2016). Bouvet et al. (2018) used a slightly simpler version of the model that can be inverted analytically, though with less flexibility for adjusting for the asymptote. In their version, the model is fit to the decibel values of the backscattering coefficient:

$$\gamma^0 = A(1 - e^{-BW}) + Ce^{-BW} . \quad (5.10)$$

This formulation ignores the volume-surface scattering and only considers the volume and surface scattering. The formulation corresponds to the modified Water Cloud Model (Santoro et al. 2002), an adaptation of the original Water Cloud Model (Attema & Ulaby 1978) that has been widely used for vegetation biomass estimation at higher frequencies such as C-band. The above model has already been used for L-band data by several authors (Cartus et al. 2012, Mermoz et al. 2014, Michelakis et al. 2014, Mitchard et al. 2011).



**Figure 5.28** Global ecoregions derived from the WWF ecoregion map (Olson, et al. 2001) by separating the ecoregions in different continents.

#### 5.6.4 RADAR-BIOMASS MODEL VARIATION ACROSS GLOBAL FORESTS

Radar-biomass algorithms will vary globally depending on the forest types. In order to provide a basic set of algorithms globally for rapid estimation of forest biomass in the absence of any reliable ground or lidar data, the development of preliminary algorithms for global vegetation ecoregions is summarized here. The primary objective of this section is to show how many different models can, on the average, represent the L-band radar sensitivity or relationship to biomass, starting with several datasets to develop the models:

- The radar backscatter data from ALOS/PALSAR, from the Japan Aerospace Exploration Agency (JAXA). JAXA has released the 2007–2010, 2015, 2016, and 2017 annual mosaics of ALOS/PALSAR data at fine-beam mode and dual polarization (HH, HV). The global mosaic is corrected for geometric distortion and topographic effects, with potentially significant residual distortions in areas of high slopes. HH and HV backscatter values from the 2007 mosaic have been used for this study. To achieve this, the backscatter digital numbers (DN) from the PALSAR product were converted to values of backscattering coefficient  in units of linear power then aggregated to a 50-m resolution using simple averaging within a 2×2-pixel window. The ALOS PALSAR data and instructions for converting the DN values to backscatter power are given [here](#).

- The GLAS, onboard the ICESat (2003–2008), can be used to make global estimates of forest height and vertical structure. In turn, these data are used to derive estimates of forest biomass at the GLAS effective footprint size of approximately 50 m (0.25 ha). All GLAS waveform data were filtered depending on the signal-to-noise ratio and cloud flag in the GLAS GLA14 dataset to develop more than 7 million clean waveforms distributed over the global forests. GLAS data can be downloaded from the [NASA DAAC website](#).
- Starting with 22 global ecoregions from the WWF biome map (Fig. 5.28), these were reduced to 15 regions that are vegetated and have significant structural diversity to affect the radar signature (Olson et al. 2001). These data were used as the main source for separating the ecoregions globally. This WWF-derived map is based on a suite of datasets such as climate, topography, and seasonality, separating the key global vegetation life-forms that have distinct structure, landscape, and climate features. Additional separation of ecoregions across continents was allowed because of distinct biogeography and plant distributions, as well as other factors including history of climate and human impacts.

The ICESAT GLAS waveform data provide vegetation vertical structure and different height metrics that were converted to AGB for each waveform (Lefsky 2010, Yu & Saatchi 2016). The ALOS PALSAR pixel values for HH and HV from the global mosaics of 2007 will be, and the associated ecoregions from

the WWF-based map were extracted for all latitude/longitude center coordinates of the GLAS footprints. Each AGB value was then associated with two ALOS polarized backscatter and one ecoregion class from WWF map. To demonstrate the relations between radar measurements and AGB, the AGB values were placed into 5 Mg/ha bins and the corresponding backscatter values were averaged. For every bin within each forest category, the mean and standard deviation of the ALOS-HH and HV backscatter values within the bin were calculated. The mid-AGB value was used to represent each bin. The extreme end of this bin distribution was terminated once the number of points within the bin fell below 500. The biomass values were also limited to 200–300 Mg/ha, a factor of 2–3 beyond the sensitivity limits of the L-band data with respect to the biomass to allow for developing the backscatter-biomass models and detecting the saturation region for each ecoregion (Fig. 5.29).

Using the models developed for HV-polarized backscatter for all 22 global ecoregions, a statistical F-test to compare the models between two or three models depending on each ecoregion using a pairwise statistical test was developed based on the statistical significance of extra sum-of-the-squares F test and the AIC approach. This analysis allows for the definition of a p-value to be small enough to give criteria necessary to separate statistically and significantly different models from other possible models. The process was first performed on models with similar ecoregions such as boreal forests of the two continents and then between the boreal forest mod-

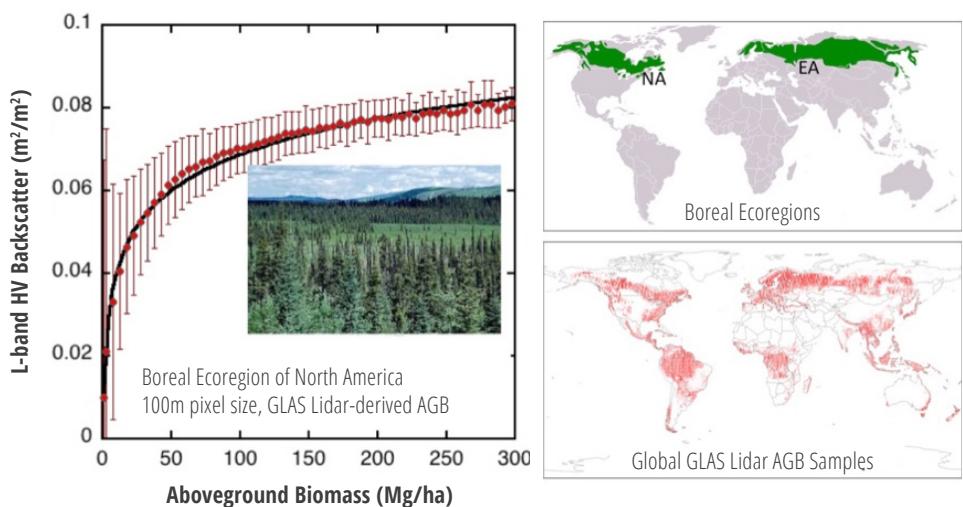
els and other ecoregions. The results of the statistical tests provided 15 distinct models across the global ecoregions (**Fig. 5.30**).

Both of the models in Eqs. (5.7) and (5.9) were used to fit the data and both performed well. Here, for simplicity and easier application of the model to the backscatter for biomass estimation, the coefficients for the global forests are provided for interested researchers or practitioners to apply on radar imagery to develop the biomass. The next sections provide specific instructions for practical implementation of the algorithms for different regions.

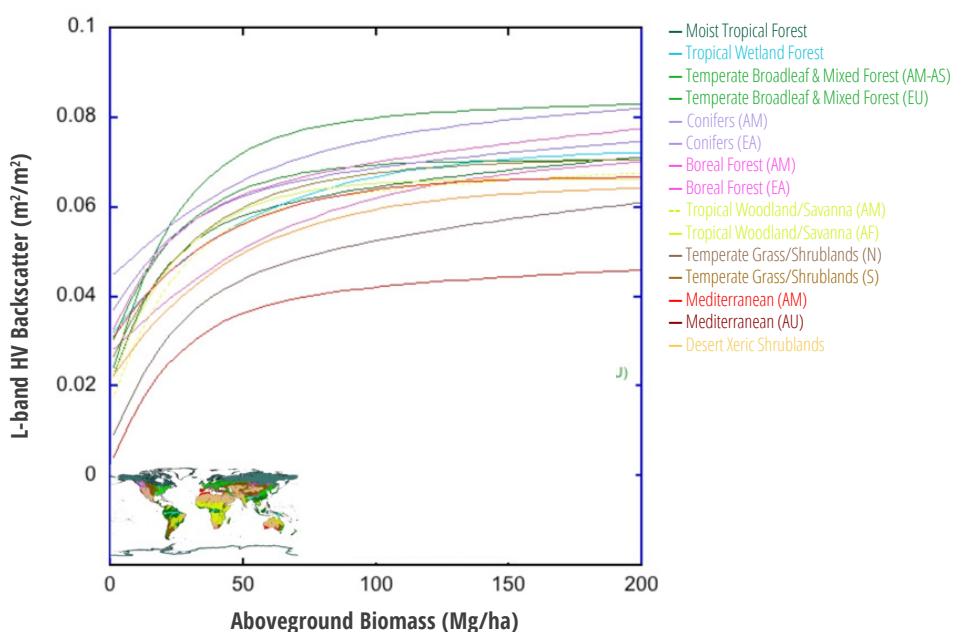
Note that because of the highly complex nonlinear nature of Eqs. (5.7) and (5.9), and the small number of fitting data points, the fitted coefficients may have large uncertainties for some ecoregions, as shown by the one-standard-deviation values of the fitted coefficients. While the more complex functional form of Eq. (5.7) may be closer to the physical representation of the full backscattering mechanism, caution must be taken in using these fitted coefficients to make physical interpretations because of the limited number of observational diversities. It is recommended that interested researchers develop more site-specific and regional models with improved ground or lidar data.

### 5.6.5 RADAR-BIOMASS NONPARAMETRIC MODELS

There are multiple ways of extrapolating samples of forest biomass data from ground or lidar measurement to a gridded map. These include parametric approaches such as the use of statistical regression models and semi-empirical models described previously that can be applied on individual radar pixels. However, maps of biomass have also been developed using a set of spatial environmental data from remote sensing and climate, and nonparametric approaches such as interpolation, co-kriging, classification, coloring by numbers, decision rule techniques, and machine-learning approaches as in the Random Forest (Xu et al. 2015), Maximum Entropy (MaxEnt) (Xu et al. 2015, Saatchi et al. 2011b), Super Vector Machine (Garcia et al. 2017), or Neural Networks (Del Frate & Solimini 2004). In some cases, the parametric models are not suitable for estimating biomass because the



**Figure 5.29** Sensitivity of L-band HV backscatter to AGB of boreal forests of North America dominated by conifers. The sensitivity is high up to 100 Mg/ha and starts declining for AGB >100 Mg/ha.



**Figure 5.30** Distinct L-band HV models for 15 ecoregions globally. The models are derived from binned backscatter and AGB data derived from GLAS lidar data.

models are developed with limited data over small regions and are used for large-scale biomass estimation. Factors such as landscape variability, forest structure, and variations of moisture and other environmental variables impact the applicability of a simple statistical model developed with limited data in and over a relatively non-representative landscape. Nonparametric models are found to be more suitable

in large-scale geospatial and geostatistical analyses because they are not affected as much by noise in the model or issues associated with multivariate normality. Nonparametric models can also integrate variables with different statistical distributions and provide more stable and relevant information. Furthermore, forest structure and biomass often exhibit complex, nonlinear variations, autocorrelation, and variable in-

teraction across temporal and spatial scales. In these cases, nonparametric approaches often greatly outperform the parametric methods. Proposed here are nonparametric and machine-learning approaches for mapping forest biomass over large areas using SAR imagery or fusion with other datasets such as topographical data from SRTM or even Landsat imagery for regional and national-scale biomass mapping for the improvement of biomass estimation.

Among nonparametric models, two common approaches used extensively for large-scale mapping were selected (Saatchi et al. 2011, Baccini et al. 2012, Xu et al. 2015): the MaxEnt estimation algorithm and the Random Forest estimation model. To apply these methods, two sets of data are required:

**(1) Training data**—the model or machine-learning training data can be selected from AGB estimated from inventory plot data or lidar measurements. The training data must be widespread to cover the range of landscape and biomass variations over the region of interest, with the number of samples covering the biomass range and representative of areas similar or compatible with the pixel size of the remote sensing imagery.

**(2) Spatial data layers**—Here, the spatial layer is SAR imagery from any airborne or satellite data such as Sentinel and ALOS. However, spatial data can be selected from a range of imagery such as SRTM to represent the topography or Landsat imagery to allow data fusion and improvement of estimation. In the case of SAR, the images are preferred to be terrain corrected, multi-looked, or speckle filtered and projected at the spatial scale compatible with the ground plots or lidar estimated biomass.

### 5.6.5.1 MaxEnt Model

MaxEnt is a probability-based algorithm that seeks the probability distribution by maximizing the information contained in the existing measurements (Berger et al. 1996, Phillips et al. 2006). The method is used as a classification approach, and each class has some probability of occurrence  $p(A_k)$ , where  $A$  is a measurement event of the response variable, while the measurements are from training samples that belong to class  $k$ . The following constraint assumes that

GLOBAL VEGETATION TYPE	A	B	C	a
<b>Africa Tropical Moist</b>	0.056492	0.064689	0	0.038247
<b>Asia Tropical Moist</b>	0.045409	0.060518	0	0.060518
<b>America Tropical Moist</b>	0.040546	0.068784	0	0.098841
<b>Temperate Conifer</b>	0.0092565	0.057336	0.04	0.27162
<b>Temperate Broadleaf/Mixed</b>	0.041469	0.034296	0.026406	0.012282
<b>Tropical Shrubland</b>	0.016429	0.11013	0	0.2675
<b>Tropical Dry Broadleaf</b>	0.021563	0.042324	0.027519	0.1117
<b>North America Boreal</b>	0.018911	0.019744	0.029106	0.15723
<b>Eurasia Boreal</b>	0.0091605	0.038506	0.04	0.26141
<b>Fresh Water Flooded</b>	0.047845	0.045581	0.022164	0.0058592
<b>Saline Water Flooded</b>	0.013682	0.051846	0.02192	0.21116

**Table 5.6** Proposed coefficients for simple ALOS PALSAR HV-based model as in Eq. (5.9) for several global ecoregions as examples for rapid estimation of forest biomass.

probabilities of all  $p(A_k)$  must sum to 1.

$$\sum_k p(A_k) = 1. \quad (5.11)$$

From information theory, the most uncertain probability distribution is the one that maximizes the entropy term:

$$E = -\sum_k p(A_k) \ln p(A_k). \quad (5.12)$$

This process will ensure that the distribution is estimated by keeping the randomness of samples for the largest entropy. Equation for  $E$  naturally gives the maximum value for the entropy when all probabilities are equal (randomness) assuming no other constraints are applied to the system except for the equation where the sum of the probabilities are 1. If additional information is available (i.e., some known AGB observations and corresponding measurements in  $X$  as in ALOS PALSAR or any SAR data; these are referred to as the training set), the probability distributions are “conditioned” on the available observations:

$$p(A_k | X) = p_k(X)p^0(A_k)/p(X). \quad (5.13)$$

The right part of Eq. (5.13) follows the Bayes’ theorem, meaning that the posterior probability  $p(A_k | X)$  depends on the distribution of  $X$  and equals to the product of prior probability  $p^0(A_k)$  and the probability distribution  $p_k(X)$  that finds  $X$  to be in the class  $k$ , and normalized by the probability distribution of

$X$  for the entire domain of measurement variables (here, satellite images). The maximization of the entropy term in Eq. (5.14) is equivalent to finding the probability distribution  $p_k(X)$  closest to  $p(X)$ , and the maximum entropy procedure gives the “raw” output:  $p_k^{raw}(X) = p_k(X)/p(X)$  (Elith et al. 2011). The prior probability  $p^0(A_k)$  is often unknown, as this quantity is the proportion of all observations over the entire scene that belongs to class  $k$ . Assuming that the training set is sampled randomly,  $p^0(A_k)$  can be estimated as  $p^0(A_k) = N_k/N_{total}$ , where  $N_k$  is the number of samples in the training set labeled as class  $k$ , and  $N_{total}$  is the total number of samples in the training set.

For the interested variable AGB, the numeric values can be categorized into a set of classes:  $k_1, k_2, k_3, \dots, k_n$ , where  $0 < k_1 \leq AGB_1 < k_2 \leq AGB_2 < \dots < k_n \leq AGB_{max}$ . And each class has a nominal value of AGB—usually the mean value of each class  $\overline{AGB}_k$ . To predict the AGB value for any pixel  $i$  with known measurements  $X_i$ , it is calculated as the expectation of all classes given the MaxEnt results retrieved from the training set:

$$\langle AGB_i \rangle = \frac{\sum_{k=1}^N [p(A_k | X_i)]^m p(A_k | X_i) \overline{AGB}_k}{\sum_{k=1}^N [p(A_k | X_i)]^m}. \quad (5.14)$$

Empirical tests have found that the model performs better by assigning higher weights to more probable classes. Therefore, an extra exponential

parameter is added to the raw output in the above equation that is determined to be approximately 3 (Saatchi et al. 2011a)

### 5.6.5.2 Random Forest Model

The Random Forest model is an ensemble model of decision trees trained from randomly selected subset features and random sampling of the training set using the bagging method (Breiman 2001). Random Forest can be a regression method when using regression trees, and for the  $j$ th regression tree, the regression model can be built as

$$AGB = f_j(x) + \varepsilon, \quad (5.15)$$

where  $x$ ,  $\varepsilon$ , and  $X$  are the bagged samples of the training set, and  $f_j(\cdot)$  is the nonparametric function determined by the  $j$ th regression tree. The final prediction of Random Forest regression is the unweighted average of the collection of trees:

$$\widehat{AGB}(X) = \frac{1}{J} \sum_{j=1}^J f_j(x). \quad (5.16)$$

This averaging process inevitably creates results biased towards the sample mean, and large/small values of AGB are often underestimated/overestimated. Various bias correction methods have been proposed to post-Random Forest results. Introduced here is a simple regression method on the Random Forest RF prediction to correct the biases, so that every 5 percentiles of the training data are grouped to have its own bias correction:

$$AGB = \alpha + \beta \widehat{AGB}(X) + \sum_{m=1}^M \gamma_m (\widehat{AGB}(X) - b_m) D_m + \varepsilon. \quad (5.17)$$

Here, the results of Random Forest prediction  $\widehat{AGB}(X)$  are further compared with the true AGB in the training set using segmented regression. Parameters  $\alpha$ ,  $\beta$ , and  $\gamma_m$  are all regression coefficients,  $b_m$  is the location of break points for the 5-percentile, 10-percentile, ..., and 95-percentile of AGB in the training set, and  $D$  is the dummy variable that equals to 1 when  $\widehat{AGB}(X) > b_m$ , and 0 otherwise. The bias-corrected Random Forest prediction is shown to have less underestimation of high AGB, which is important for biomass and carbon estimations.

To evaluate the performance of the machine-learning algorithms, recommend three statistical measures are recommended: (1) the coefficient of determination ( $R^2$ ), (2) the Root-Mean-Square

Error (RMSE), and (3) the Mean Signed Deviation (MSD). Once all of these measures are applied to an independent test set where the original AGB is obtained from ground data or airborne lidar, while the predicted AGB is derived using the SAR and other remote sensing data layers and the model trained from the training set. In addition, it is recommended to use the Moran's I statistics to quantify the spatial autocorrelation in the data. The local Moran's I index confirms the need to select more spatial samples in heterogeneous areas like forests, as it can identify spatial clusters and outliers (see Xu et al. 2015, Xu et al. 2017).

### 5.6.6 PRACTICAL CONSIDERATION FOR SAR BIOMASS ESTIMATION

The following is recommended for practical use of SAR imagery for biomass estimation:

**(1) Choice of SAR data**—Depending on the vegetation type and the scale of analysis and biomass range, the choice of radar data may be different. For all areas covered with low-vegetation biomass such as grasslands, shrublands, sparse woodlands, young secondary regeneration, and low-density wetlands, the C-band data from the Sentinel satellites are the most suitable datasets. If airborne SAR data are available for the study site, use of polarimetric C-band data at high spatial resolution is recommended. From low to moderately high biomass up to 100–150 Mg/ha, the use of L-band polarimetric or dual-pol data are recommended. ALOS-2 PALSAR imagery is the most suitable dataset because of its frequent observation (every 14 days), resolution (~20 m), and sensitivity to biomass. For all forests >150 Mg/ha of biomass, use of P-band data that are currently mainly from limited airborne sensors are recommended. P-band data can be used for estimating and monitoring tropical forest biomass.

**(2) InSAR observations**—Although, the subject was not covered in this chapter, the use of Interferometric SAR (InSAR) for measuring the forest structure across some vertical depth may help with improving the biomass estimation particularly beyond the saturation level

in some forests. Unfortunately, reliable InSAR data are not readily available. The future Biomass mission (and to some extent the NISAR mission) may provide some InSAR data. However, the use of Sentinel, ALOS, and Terra-X SAR data have been used in InSAR models in different studies to explore the use of vertical structure derived from radar for biomass estimation.

**(3) Multitemporal observations**—Due to the sensitivity of radar imagery to soil moisture, and to some extent variations of vegetation moisture seasonality, the use of time series images for reducing the effect of environmental factors for biomass estimation is recommended. The SAR biomass model often performs poorly if it is developed based on one SAR image and applied on an image acquired at a different season or date. One practical approach is to collect as many SAR images over the study areas as possible and average the data temporally to reduce the effect of the moisture before developing the model as shown in **Figure 5.25**.

**(4) Map unit and pixel size**—Choose map units of 100 m or more for improved results from the biomass estimation. It is recommended that SAR biomass models are developed with plots of at least 1 ha in size for a relatively unbiased estimate of the biomass within the range of biomass allowed for the SAR data. If reliable models are developed at smaller pixels or plot sizes (e.g., 0.25 ha) for some forest types (dry forests, woodlands, boreal), it is recommended to estimate the biomass at 0.25 ha and aggregate the result to 1 ha or more for applications. The error of biomass estimation will reduce with a factor of slightly less than  $\sqrt{n}$ , where  $n$  is the number of pixels for averaging (Weisbin et al. 2014).

**(5) SAR measurement diversity**—Most models shown in this section were based on L-band HV backscatter measurements due to its improved sensitivity to biomass and moderate effects of moisture or other environmental factors. However, use of models that include

several polarizations or even measurements of two or three frequencies improve the model performance and the accuracy of the biomass estimation. In most practical studies, access to some ALOS PALSAR and Sentinel satellite imagery is possible. Combining the datasets in a statistical model as shown in this section can improve the accuracy of the biomass across different ranges.

## 5.7 Uncertainty Analysis

Uncertainty analysis has become an important ingredient of forest biomass estimation from both ground and remote sensing data due to Intergovernmental Panel on Climate Change (IPCC) guidelines (IPCC 2006). This section provides a summary of different types of uncertainty analysis from simple to more complex inferences of mean or total biomass (carbon stocks) of forests at regional or national scales. According to the IPCC ([Chapter 3](#)) and the [Carbon Fund Methodological Framework](#), all forest biomass and carbon assessments at the project, jurisdictional, and national levels must address the uncertainty related to the biomass estimation and all derived products such as emissions from deforestation and degradation by:

- (1) Identifying and assessing sources of uncertainty
- (2) Minimizing uncertainty where feasible and cost effective
- (3) Quantifying remaining uncertainty

The sources of uncertainty are identified in both the land-use or activity data and the biomass estimations for different land use and land cover classes and emission factors. Once the sources are identified, their relative contribution to the overall uncertainty of the biomass estimation at the regional for land cover types and hence emissions and removals can be quantified and reported. Here, the uncertainty analysis is summarized in three steps: (1) Cross Validation (CV) approach for developing uncertainty for SAR-biomass models and local area estimation of biomass when ground or reference data are available, (2) error propagation approach showing how uncertainty from different sources of errors can be combined to provide total uncertainty on the biomass estimation at the biomass map units or on the aver-

age for a region, and (3) inference of forest biomass at regional scales by calculating both the mean and the variance or uncertainty around the mean using uncertainty of sources of errors and spatial correlation of map units or derived biomass pixels.

### 5.7.1 CROSS VALIDATION

Cross validation is a modeling technique used to check the statistical learning consistency with independent data from the training set itself. Not only can it be used to check the performance of the SAR biomass model or spatial modeling by making predictions on new data that are never used in the training, but it is also often used as a technique of parameter tuning to avoid “overfitting.” For regression-based analysis, the mean-squared-error (MSE) is normally used as the scoring parameter in the CV process. There are several ways of cross validation commonly used to evaluate the performance, including k-fold approach, leave-one-out CV, repeated random subsampling (or Monte Carlo CV), and so on. Interested readers can consult several references for the use of validation approaches for quantitative remote sensing products (Browne 2000, Hawkins et al. 2003, Arlot & Celisse 2010).

### 5.7.2 ERROR PROPAGATION

The overall sources of uncertainty for estimating forest biomass from SAR or any remote sensing data can be summarized as follows:

- **Measurement Errors**—This error can be either random or systematic and results from errors in measuring, recording, and transmitting the information.
  - In ground data, there are several sources of the error that can impact the biomass estimation at the plot level (Chave et al. 2005). In addition, in measurements of trees, the size and location of the plot can introduce significant errors in biomass estimation as a reference data to be compared to the SAR measurement.
  - SAR measurements also may have errors associated with the absolute calibration of the system, the RTC method for removing topography and incidence angle effects, and geolocation of pixels when using the data to compare with
- **Statistical Random Error**—This source of error often appears in inventory data that are supposed to be a random sample of a finite size depending on the variance of the population. Here, the sample size is a key source of uncertainty.
- **Misclassification and missing data**—This uncertainty is due to incomplete, unclear, or faulty definition of data, and allometric models leading to bias in estimation of biomass. This will often occur when working with the ground plots.

ground plots. Together these effects can cause bias and random errors in estimation of emissions. Discussion has already been presented as how to improve the errors associated with the SAR data processing and radiometric corrections.

- **Statistical and Sampling Errors**—To develop models of SAR backscatter with ground reference data or airborne lidar-derived biomass may also have errors associated with the sampling and the statistical representation of the plots and pixels. In general, plot data need to represent the landscape variations of biomass from low to high biomass and must follow the requirements of the size, orientation, and geolocation and number of samples.

- **Lack of representativeness of data**—This source of uncertainty is associated with a lack of complete correspondence between ground and SAR data. In addition, to develop the SAR biomass models, the pixels must be spatially representative in the SAR image and not all are from a certain incidence angle, at a certain elevation and slope in order to make sure that the relationship developed between the SAR data and ground are representative. Any errors in sample size and sampling characteristics can introduce both systematic and random errors.

- **Models**—Models developed from SAR data and biomass often have uncertainty due to both the choice of the model function and the fit of the model parameters. If data are noisy, the model fits may have large errors that include both the systematic (choice of wrong model equation) and random errors.

- **Statistical Random Error**—This source of error often appears in inventory data that are supposed to be a random sample of a finite size depending on the variance of the population. Here, the sample size is a key source of uncertainty.

- **Misclassification and missing data**—This uncertainty is due to incomplete, unclear, or faulty definition of data, and allometric models leading to bias in estimation of biomass. This will often occur when working with the ground plots.

For example, using biomass data from plots with different type of measurements, lack of availability of allometric models to estimate biomass (e.g., allometry for tropical wetland forests).

- **Missing data**—This uncertainty may result when measurements are below the detection limit causing a nondetected data that can, in turn, introduce both bias and random errors. By assuming that there are several sources of errors that introduce uncertainty in the pixel-level estimation of biomass, the total uncertainty associated with estimating AGB at the pixel level can be calculated by assuming all errors are independent and random, by using

$$\varepsilon_{AGB} = \sqrt{\varepsilon_{\text{measure}}^2 + \varepsilon_{\text{model}}^2 + \varepsilon_{\text{sampling}}^2 + \varepsilon_{\text{prediction}}^2}, \quad (5.18)$$

where each of the terms are the relative errors at that pixel scale. Using the above equation, the errors at the pixel level will be propagated and a map of the uncertainty at the pixel level will be created. The main requirement for a pixel-level map of uncertainty is to be able to have a pixel-level prediction error from the model. The prediction error for SAR estimation of biomass at the pixel level is often developed through a boot-strapping approach where the model errors are simulated to generate different predictions for the pixel scale AGB and to produce the mean and use the variance as the prediction uncertainty or error.

### 5.7.3 REGIONAL INFERENCE OF BIOMASS

The goal of regional estimate of biomass and forest carbon stocks is to be able to develop emission factors for calculating emissions and removals from different types of human-induced disturbances in the forest such as deforestation, degradation, regeneration, or agroforestry. The problem then is to be able to use the estimates of biomass at the map units (pixel scale) from SAR data to estimate the mean and variance (uncertainty) of the biomass at large scales. The mean is estimated by the average of the biomass of all pixels for a region. However, for estimating the variance several components of errors must be included in the calculation, such as the errors associated with the spatial correlation of

biomass estimates at the pixel level.

#### 5.7.3.1 Spatial Autocorrelation

To demonstrate the existence of spatial autocorrelations among the biomass estimates at the pixel level, the use of semivariogram analysis (Isaaks & Srivastava 1990) is recommended. The variogram-based approaches assume that the spatial autocorrelation of variables only depends on the distance  $h$ , while it has no other directional or locational dependence. The variogram  $\gamma(h)$  is defined as

$$\gamma(h) = \frac{1}{2} E \left[ \left( y_{x_i} - y_{x_j} \right)^2 \right] = [C(0) - C(h)], \quad (5.19)$$

where  $||(x_i - x_j)|| = h$  and  $C(h)$  is the covariogram depending on the distance  $h$ . In addition to above uncertainty at the pixel scale, to calculate the uncertainty at the regional level for forest biomass, the spatial correlation of the errors at the pixel level much be considered. The spatial correlation derived from semivariogram analysis will provide the variance to the estimate of the error using the following model

(see VT0005) (Weisbin et al. 2014).

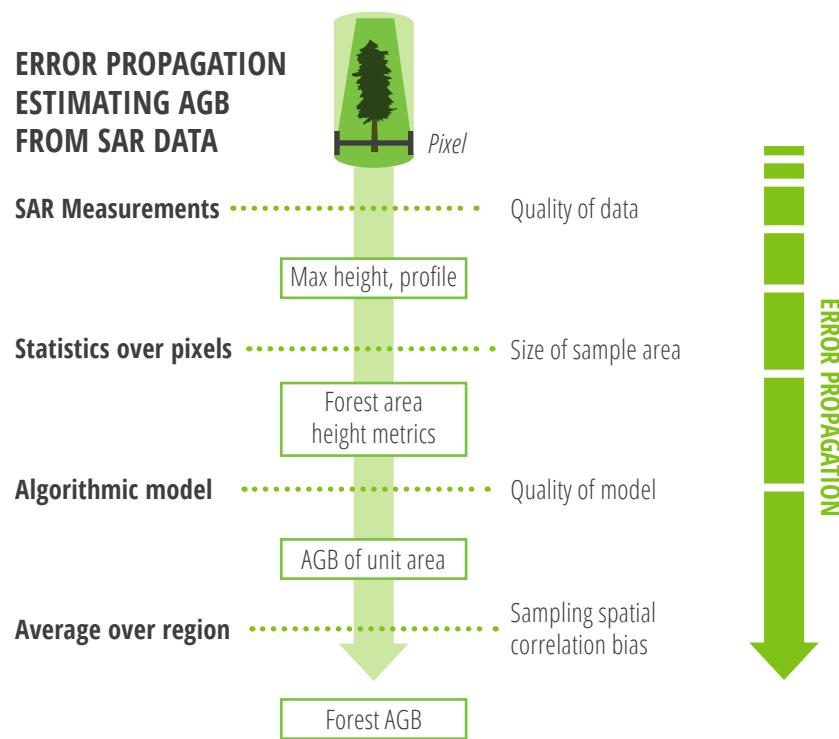
$$\sigma_L^2 = P^{-1} \frac{1}{m} \left( \sum_{i=1}^m \sigma_{ui}^2 + 2 \sum_{i=1}^m \sum_{j=1, j < i}^m p(d) \sigma_{ui} \sigma_{uj} \right) \quad (5.20)$$

and

$$p(d) = \exp \left( -\frac{d}{\alpha} \right), \quad (5.21)$$

where

- $P = 1$  (representing the size of the pixel as 1 ha)
- $i, j$  are the generic indices representing pixels in the biomass map
- $n$  is the number of pixels within each LULC or stratum
- $r$  is the range from semivariogram estimating the spatial correlation of errors associated with the AGB pixel level errors
- $c$  is the parameter of fit for exponential spatial correlation function derived from semivariogram analysis.  $c = 1/3$  is the default value (Chilès & Delfiner 2012) (unitless)
- $d$  is the distance between pixels  $i$  and  $j$  within  $m$  (pixels)



**Figure 5.31** Schematic showing the main sources of uncertainty in SAR estimation of AGB and the process of error propagation for total uncertainty assessment.

- $p(d)$  is the spatial correlation function in terms of distance  $d$  based on exponential semivariogram model (unitless)
- $\sigma_L^2$  is the variance derived from a priori RS data, a pilot study, or default values of AGB density for the LULC class
- $m$  is a dummy large number representing pixels in the map for each LULC. The number can be arbitrarily large or at least twice the default value of range  $r$
- $\sigma_{w,j}^2$  is the estimated variance associated with AGB values for each 1-ha pixel of the map

By assuming a pixel level uncertainty that is derived from the boot-strapping approach of SAR-biomass relationship or from the machine-learning algorithm of ( $\sigma$ ) at each pixel, the uncertainty of the mean biomass at the regional scale can be evaluated using:

$$\begin{aligned}\underline{\sigma}^2 = & \frac{1}{N^2} \left( \sum_{i=1}^N \sum_{j=1}^N \text{cov}(\sigma_{\varepsilon,i}, \sigma_{\varepsilon,j}) \right) \\ & + \frac{1}{N^2} \left( \sum_{i=1}^N \sum_{j=1}^N \text{cov}(\sigma_{f,i}, \sigma_{f,j}) \right) \\ & + \frac{1}{N^2} \left( \sum_{i=1}^N \sum_{j=1}^N \text{cov}(\sigma_{z,i}, \sigma_{z,j}) \right),\end{aligned}\quad (5.22)$$

where  $N$  is the total number of pixels,  $\sigma_\varepsilon$ ,  $\sigma_f$ , and  $\sigma_z$  are the pixel-level errors from (1) spatial mapping uncertainty, (2) allometric equation uncertainty, and (3) uncertainty of predictor variables from SAR backscatter, respectively. The three sources of errors are assumed independent, so that the overall uncertainty of regional estimates comes from the three covariance terms.

The first covariance using spatial autocorrelation is modelled

$$\begin{aligned}\underline{\sigma}_e^2 = & \frac{1}{N^2} \left( \sum_{i=1}^N \sigma_i^2 + \sum_{i=1}^N \sum_{j(j \neq i)}^N \rho_{ij} \sigma_i \sigma_j \right) \\ & + \sum_{p=1}^m \sum_{q=1}^m \left( g_p \text{cov}(\phi_p, \phi_q) g_q \right),\end{aligned}\quad (5.23)$$

where  $\rho_{ij}$  is the correlation coefficient between pixels  $i$  and  $j$ , and it can be approximated from the variogram (Eq. (5.19)) normalized  $C(h)$  under the assumption that spatial autocorrelation only changes with distance  $h$ .

The second covariance is related only to the allome-

tric model coefficients and can be reformulated as

$$\begin{aligned}\underline{\sigma}_f^2 = & \sum_{p=1}^m \sum_{q=1}^m \left( \frac{1}{N} \sum_{i=1}^N \frac{\partial f}{\partial \phi_p} \right) \text{cov}(\phi_p, \phi_q) \left( \frac{1}{N} \sum_{j=1}^N \frac{\partial f}{\partial \phi_q} \right) \\ = & \sum_{p=1}^m \sum_{q=1}^m \left( g_p \text{cov}(\phi_p, \phi_q) g_q \right),\end{aligned}\quad (5.24)$$

where  $g_p = \frac{1}{N} \sum_{i=1}^N \frac{\partial f}{\partial \phi_p}$  is the mean of first derivative with respect to the allometric model coefficient  $\phi_p$ , and  $m$  is the total number of coefficients in the allometric model—or in lidar-AGB model,  $m$  equals to 2. In the case that the biomass values used to calibrate the SAR data are from ground plots and not the lidar-AGB model, one can assume a certain fixed value as the uncertainty of the biomass from the tree allometry (Chave et al. 2014, Chave et al. 2005), or assume the value is 0.

The third covariance is related to the measurement errors. In the case of the lidar-AGB model,  $\sigma_z$  is the error associated lidar mean canopy height. Without in-situ validation of height measurements, it is impossible to evaluate this type of error. Discussion in the main paper has shown that at least the model-based height interpolation is very accurate and the error in 1-ha resolution is negligible. In the case of SAR, this error may be related to backscatter error associated with the radiometric calibration. This uncertainty can be a fixed value (1 dB for all pixels) or a value that varies depending on the SAR range and azimuth or the local incidence angle as a result of the terrain topographical complexity. Calculating the measurement errors of terrain-corrected SAR backscatter may be difficult and beyond the scope of this chapter. Therefore, the use of a calibration error recommended by the SAR processing team or available in the literature is recommended. For further readings on the uncertainty of inference of biomass at the regional level see McRoberts et al. (2017), Ene et al. (2017), Naesset et al. (2016), and Xu et al. (2017).

#### 5.7.4 PRACTICAL CONSIDERATION FOR UNCERTAINTY CALCULATION

For validation of SAR-derived biomass maps, methodologies that can help improve the uncertainty

estimates or reduce the uncertainty are identified as part of the IPCC good practice guidelines. The biomass map can be distributed to the community to be used for land use planning, REDD+ projects, and the Emission Reduction (ER) programs; and in all applications, formal uncertainty assessments are required. Regional evaluation of the map can be performed by using inventory plots or airborne lidar data and site-specific lidar biomass allometry that together allow the estimation of the potential bias and the evaluation of the spatial consistency of the map. The methodology to develop regional estimates of forest biomass must follow one of the many standard protocols established by forest inventory techniques or the IPCC guidelines. For lidar sampling, a certified methodology that can be used for regional forest biomass estimation has already been developed. The methodology was developed by SASSAN Saatchi and recently developed as a VCS tool with the collaboration of Terra Global Capital as VT0005 (see the attached appendix). At the time of releasing the map, ground data was too limited to have a comprehensive evaluation of the map regionally or locally. Here, a set of protocols is recommended for those who are interested in evaluating the map further at the local or regional scale.

- The biomass map derived from SAR is considered to have both systematic and random errors. The uncertainty of the map depends strongly on the input biomass data used for training and evaluating the results. Any errors in ground-based allometry and the lidar derived biomass as reference can impact the estimation and the map's accuracy.
- The biomass estimates in the map can be readily updated and improved when more data becomes available. Increasing the number of ground-estimated forest biomass, use of forest specific models, use of multitemporal SAR images for biomass estimation are critical steps to improve the accuracy of the map.
- For evaluating the map at regional scales (>10,000 ha), samples of lidar and plot data can be used. Any plots that are designed statistically to estimate the mean biomass with a high confidence interval (90%) can be used to compare with the

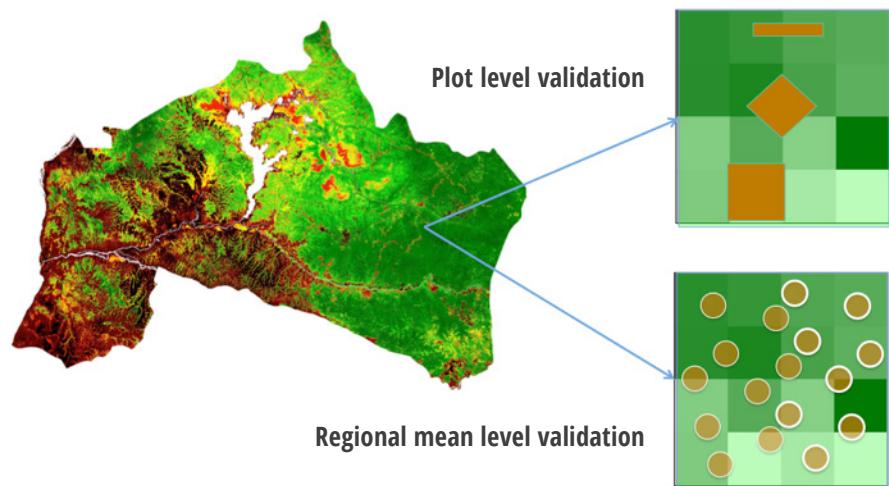
map pixel values. At the scale of parks, concessions, and communities, the biomass map can be used along with any land cover map to develop mean biomass density and can be compared with independent inventory data available for the same region.

- The map can also be evaluated at the pixel scale (e.g., 1 ha). However, for the comparison, the following precautions must be considered:
  - The plots have to be a minimum of 1 ha or larger. Using smaller plots is not recommended, as the biomass of the forest is extremely heterogeneous, particularly at scales of less than 1 ha.
  - The 1-ha plots chosen for comparison with the map have to be aligned with the map pixel orientation. Any plots with different orientation may have large uncertainty when compared to the map because of the variations of the biomass at the 1-ha scale.
  - The map also includes an uncertainty number associated with the biomass of each pixel. Any comparison with the pixel value biomass should consider the error provided for the pixel.
  - The number of plots must be larger than a few. Comparison only makes sense when it is statistically designed. Using one or a few plots will not provide any realistic and fair comparison of the map. It is recommended that at least 20–30 plots be used in statistically evaluating the map and comparing the results with the uncertainty provided.

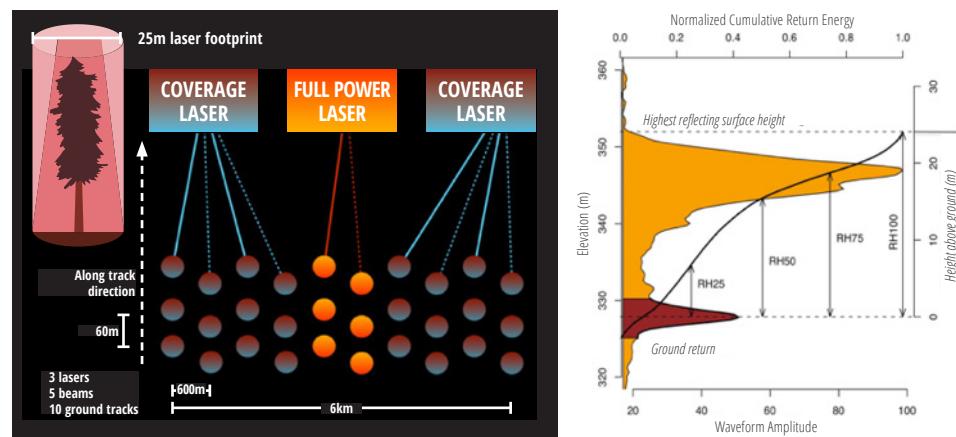
## 5.8 Future Biomass Missions

### 5.8.1 GEDI (LAUNCH 2018–2019)

The scientific goal of GEDI is to characterize the effects of changing climate and land use on ecosystem structure and dynamics to enable improved quantification and understanding of the Earth's carbon cycle and biodiversity. Focused on tropical and temperate forests from its vantage point on the International Space Station (ISS), GEDI uses a lidar sensor (near infrared 1,064-nm wavelength) to provide the first global, high-resolution (25 m) sampling observations



**Figure 5.32** Methodology to evaluate the forest biomass map locally or at the pixel level using field inventory plots. For local evaluation, inventory plots must follow a statistical design to allow accurate mean AGB values. For pixel-level evaluation, plots must be equal to or larger than the pixel and must be oriented to maximize the overlay of the best spatial match between the pixel and the plot biomass.



**Figure 5.33** Distribution of GEDI footprints across the landscape from the three lasers and multiple beams (left panel) and the typical distribution of forest vertical structure captured by the GEDI footprint level waveforms.

of forest vertical structure. GEDI addresses three core science questions: (1) What is the aboveground carbon balance of the land surface? (2) What role will the land surface play in mitigating atmospheric CO<sub>2</sub> in the coming decades? (3) How does ecosystem structure affect habitat quality and biodiversity? Answering these questions is critical for understanding the future path of global climate change and the Earth's biodiversity.

GEDI informs these science questions by collecting ~12 billion cloud-free land-surface lidar waveform (vertical profile) observations over a two-year

mission lifetime. The instrument uses three laser transmitters split into five beams that are dithered to produce 10 parallel ground tracks of 25-m footprints (Fig. 5.33). GEDI will produce estimates of canopy height, elevation, and vertical canopy profile measurements. The 25-m (~0.0625 ha) footprint measurements are used to model AGB and then used to derive mean AGB and variance on a 1-km grid.

#### 5.8.1.1 GEDI CAL/VAL Requirements

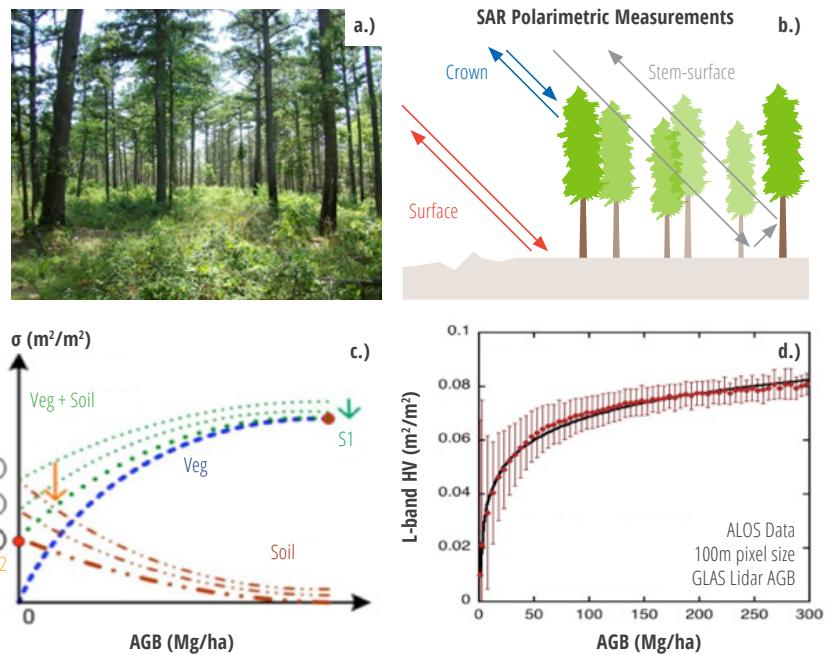
From its vantage point on the ISS, GEDI is focused on tropical and temperate forests between 51.5°S

and 51.5°N. The GEDI biomass calibration strategy is to develop globally representative pre-launch models for footprint AGB using near-coincident airborne laser scanning (ALS) data and plot inventory data. Mean and standard error of AGB for 1-km grid cells are then estimated from the modelled footprint AGB via statistical inference. The baseline requirement for GEDI is that the standard error of AGB estimates within 80% of Level 4B gridded product at 1-km cells will be <20 Mg ha<sup>-1</sup> or 20%, whichever is greater. The GEDI science products are developed using a series of airborne lidar and ground plots globally and models to estimate biomass from GEDI waveforms. These datasets are sampled globally to be representative of major forest types.

### 5.8.2 NISAR MISSION (LAUNCH 2021)

NISAR is a joint project between [NASA](#) and [ISRO](#) to co-develop and launch the first dual-frequency SAR satellite. NASA will provide the L-band (24-cm wavelength), and ISRO will provide the [S-band](#) (12-cm wavelength). The mission will acquire polarimetric and interferometric observations at an unprecedented coverage in space and time, which is optimized for studying changes of the global Earth surface.

NISAR will focus on the most dynamic ecosystems such as disturbed and recovering forests, inundated wetlands, and croplands. NISAR will measure aboveground woody vegetation biomass and its disturbance and recovery globally at the hectare scale; biomass accuracy shall be 20 Mg/ha or better for areas of woody biomass  $\leq$  100 Mg/ha over at least 80% of these areas. Therefore, the mission will focus on areas of low biomass, covering a significant portion of boreal, temperate, and savanna woodlands. It will provide seasonal to annual observations of biomass change in the most dynamic forests impacted by AGB disturbance and recovery. The NISAR mission will be able to provide L-band dual pol (HH, HV) observations every 12 days in ascending and descending orbits covering global forests every 6 days. These observations will be used to produce maps of the distribution of forest biomass at 1-ha grid cells. The NISAR radar is designed for global InSAR measurements, but the science products produced do not



**Figure 5.34** Schematic showing a typical northern conifer forest (a) simulated to an ensemble of trees with stems, branches, and leaves (b) exposed to L-band radar energy with dominant scattering from forest components, (c) suggesting the combined influence of structure and soil moisture on radar backscatter with reduced sensitivity to biomass at higher AGB values. The last panel (d) shows the sensitivity of radar backscatter at L-band HV polarization showing the sensitivity to biomass values  $<$  100 Mg/ha with sample data from the entire northern coniferous forests (Yu & Saatchi 2016).

include direct information on the vertical structure of forests. Rather, AGB is estimated from backscatter measurements and exploits either empirical statistical approaches or inversion of physically-based scattering models that must be calibrated over study sites globally to capture the structural and composition differences of forests in different ecoregions.

The NISAR algorithm is based on an analytical semi-empirical model with coefficients that are calibrated with structure and biomass information from ground measurements. The forest inventory data available in a network of calibration plots distributed globally in different ecoregions (15 ecoregions as discussed in [Sec. 5.6](#)) and accompanied by airborne lidar observations to extend the ground observations and enable validation of the spatial variations of AGB. The size of plots used for calibration of the NISAR algorithm must be either  $>1$  ha if used directly with the SAR data or smaller ( $\sim 0.25$  ha) if used in conjunction with the ALS observations. In addition, forest inventory data can be used to evaluate and report the

uncertainty of NISAR AGB at the national or regional scale and for carbon accounting and assessments.

### 5.8.3 BIOMASS (LAUNCH: 2022)

Biomass, the ESA's seventh Earth Explorer mission will be launched in the 2020–2021 timeframe and has the aim of providing crucial information about the state of the forests and how they are changing globally. The mission goal is to provide estimates of height and AGB in the world's forests. The science case on which Biomass was selected is based on its ability to provide estimates of AGB within dense tropical forests to monitor their storage and changes from disturbance at seasonal and annual frequency. The requirement for the Biomass mission is to estimate forest biomass with an accuracy of  $\leq 20\%$  for more than 67% of areas with biomass  $> 50$  Mg/ha on a 4-ha spatial grid cell (200-m x 200-m pixels) every six months for a period of five years of the mission duration. This requirement is achieved by using a P-band (70 cm wavelength) SAR sensor, because

of its unique capabilities to penetrate even dense tropical forest. The measurements will provide radar polarimetric backscatter (HH, HV, VH, VV) and interferometric observation with PolInSAR capability for forest height estimation and TomoSAR capability for backscatter vertical profile measurements.

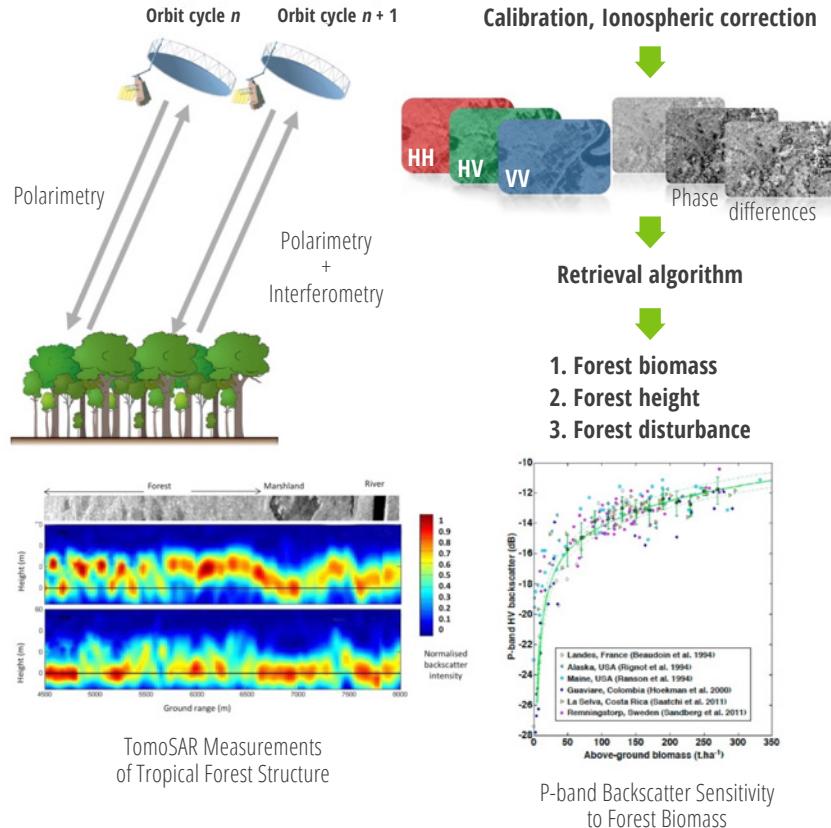
In addition, the Biomass mission will provide global maps of forest height at the same 4-ha spatial scale for all forests >10-m height with 30% accuracy and include a 50-x-50-m deforestation map globally every six months. These measurements together, will significantly improve the ability to reduce the uncertainty in the global carbon cycle by providing spatially refined and temporally frequent observation of carbon fluxes in forest ecosystems.

The coverage of Biomass is global with a restriction, imposed by the U.S. Department of Defense Space Objects Tracking Radar (SOTR) stations, over Europe and the North and Central Americas. Under these restrictions, only 3% of AGB carbon stock coverage is lost in the tropical forest biome, which constituted 66% of global AGB carbon stocks in 2005. The loss is more significant in the temperate (72%), boreal (37%), and subtropical (29%) biomes. The calibration/validation (CAL/VAL) requirements of Biomass are primarily focused in tropical forest ecosystems, where the bulk of mission observations are located. The biomass and structure algorithms require large ground plots (>4 ha) or lidar-derived AGB estimates from airborne observations. These measurements must represent the variations of tropical forest structural types and allometric characteristics and must be repeated during the mission to allow validation of both biomass stocks and changes from disturbance and recovery.

#### 5.8.4 CROSS-MISSION SYNERGISM

All three missions have significant overlaps in science objectives and products but focus on different observations, covering different regions, and retrieving different components of AGB at different spatial and temporal scales. The cross-mission synergism is based on the following observations and assessments from the breakout sessions:

- **Area coverage** and the science products from the space missions are immediately rec-



**Figure 5.35** Biomass mission P-band SAR measurements showing the configuration of space measurements and the sensitivity of backscatter power and interferometry to forest structure.

ognized as complementary such that without the data from all the missions, wall-to-wall coverage, and estimation of the global forest biomass are impossible. Biomass focuses on tropical and subtropical woodlands at 4 ha, while NISAR is global but limited to areas of low forest biomass at 1 ha, and GEDI not limited by AGB, but with limited coverage collecting sample footprints within  $\pm 50$  degrees latitude.

#### • Differences in biomass components

retrieved by each space mission suggest that a synergistic global AGB product cannot be mechanistically produced by combining the maps, but rather requires a systematic data fusion approach. For reference, BIOMASS will estimate AGB when woody biomass is > 50 Mg/ha, NISAR will estimate AGB when woody and leafy biomass is < 100 Mg/ha, and GEDI will estimate AGB for the entire range from height measurements within each 25 m footprint.

- **Leverage the sensitivities of each measurement approach to cross-calibrate space mission products** can be achieved by using the measurements and products of one mission to CAL/VAL the algorithm or products of other missions. Although every space mission has a different method for estimating AGB, thus making it difficult to directly compare between products, an approach could be used that compares either similar lower level products or leverages different algorithm sensitivities (e.g., NISAR can provide more robust estimates for forests with 20 Mg/ha than for grasslands with  $\leq 5$  Mg/ha). For example, GEDI forest height may be used to develop and verify algorithms for the Biomass tomography-derived tree height. Similarly, height or backscatter products from NISAR and Biomass missions can provide information on the spatial variability of forest structure and biomass to improve the algorithm and resolu-

tion of GEDI height and biomass gridded products.

- **NISAR and Biomass measurements spatially overlap**, thus enabling data fusions such as (1) the combined measurements of L-band P-band for improving the estimates of low-biomass forests, (2) the use of higher temporal frequency NISAR observations to reduce the effects of soil moisture and vegetation

phenology on the BIOMASS estimation approach, (3) the deployment of a two-frequency algorithm to enable estimation of leaf, branch, and stem biomass, and (4) the use of combined measurements to increase sensitivity of the observations for detecting low-impact forest degradation and slow recovery.

- **Additional ecological science products** may be produced from synergistic integration

of data that enables exploration of the physical characteristics of each measurement. For example, other forest variables such as basal area, volume, branch, leaf, and stem biomass, and forest stand wood density may be derived using the combined sensitivity of radar observations to dielectric constants and tree stem and crown volumes, and ability of lidar waveforms to measure the vertical distributions and canopy gaps.

Mission	Measurement	Product	Area Coverage	Grid Cell	Accuracy	Pre-launch Cal/Val Mission	Cal/Val needs	Post-launch Cal/Val	Sites
GEDI	Height	Height Metrics	50 deg. Latitude	25m footprint; 500m grid	~1m (canopy top footprint level)	ALS & LVIS flights RT modeling	LVIS samples globally	No validation requirement	International; crowd-sourced
	Waveform	Aboveground biomass (entire range)	50 deg. Latitude	25m footprint; 1 km grid	20 Mg or 20% at 1km, 80% px std. err.; mission	Footprint calibration equations; sampling simulations, ALS & LVIS flights	ALS-derived biomass from ground plot at 1 km grids	No validation requirement	NFI data supersites with ALS-derived biomass > 100 ha
BIOMASS	HH, HV, VH, VV Backscatter	Aboveground biomass (entire range)	Global (excluding North/Central America, Europe)	200 m (4 ha)	20% or 10 Mg/ha for biomass < 50 Mg/ha annual	Combined radar backscatter and TomoSAR & POLinSAR algorithm	Plots > 4 ha & ALS; across ecoregions	NFI & regional samples	NFI data CTFS; ForestGeo
	POLinSAR	Forest Height	Global (excluding North/Central America, Europe)	200 m (4 ha)	20% of total height	POLinSAR height algorithm	ALS & LVIS data distributed across ecoregions	Same approach as pre-launch	
	TomoSAR Vertical Profile	TBD	Global (excluding North/Central America, Europe)	200 m (4 ha)	TBD	TomoSAR vertical structure	ALS & LVIS data distributed across ecoregions	Same approach as pre-launch	ALS and LVIS data
	Time Series	Forest Disturbance	Global (excluding North/Central America, Europe)	50 m (0.25 ha)	90% of pixels, annual	Optical imagery time series over selected sites	Landsat time series data, high-resolution imagery	Same approach as pre-launch	Distributed globally at deforestation hot spots
NISAR	HH & HV Backscatter	Aboveground biomass < 100 Mg/ha	Global low biomass areas	100 m (1 ha)	20 Mg/ha, 80% px < 100 Mg/ha annual	Radar biomass equations, algorithm	Plots > 1 ha & ALS data across ecoregions	NFI & regional samples	NFI data Distributed large plots & ALS data
	Time Series	Disturbance > 50% change in canopy cover	Global forests	100 m (1 ha)	80% of pixels, annual	High-res optical & ALOS/SACOM time series over selected sites	Landsat time series data, high-resolution imagery	Same approach as pre-launch	

**Table 5.7** Overall characteristics of the NASA and ESA missions to quantify the global forest structure and biomass that can be used to develop synergistic biomass products

## 5.9 References

- Andersen, H. E., S. E. Reutebuch, R. J. McGaughey, M. V. N. d’Oliveira, M. Keller. (2014). Monitoring selective logging in western Amazonia with repeat lidar flights. *Remote Sensing of Environment*, 151: 157 – 165.
- Arlot, S., and A. Celisse. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4, 40-79.
- Asner, G. P., J. Mascaro. (2014). Mapping tropical forest carbon: Calibrating plot estimates to a simple Lidar metric. *Remote Sensing of Environment*, 140, 614 – 624.
- Asner, G. P., G. V. N. Powell, J. Mascaro, D. E. Knapp, J. K. Clark, J. Jacobson, T. Kennedy-Bowdoin, A. Balaji, G. Paez-Acosta, E. Victoria, L. Secada, M. Valqui, and R. F. Hughes. (2010). High-resolution forest carbon stocks and emissions in the Amazon. *Proceedings of the National Academy of Sciences of the United States of America*, 107(38), 16738 – 16742.
- Attema, E. P. W., and F. T. Ulaby. (1978). Vegetation modeled as a water cloud. *Radio Science*, 13(2), 357 – 364.
- Baccini, A., S. J. Goetz, W. S. Walker, N. T. Laporte, M. Sun, D. Sulla-Menshe, J. Hackler, P. S. A. Beck, R. Dubayah, M. A. Friedl, S. Samanta, and R. A. Houghton. (2012). Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nature Climate Change*, 2, 182-185.
- Baghdadi, N., C King, A. Bourguignon, and A. Remond. (2010). Potential of ERS and Radar-sat data for surface roughness monitoring over bare agricultural fields: Application to catchments in Northern France. *International Journal of Remote Sensing*, 23(17), 3427 – 3442.
- Berger, A. L., and V. J. D. Pietra. (1996). A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1), 39-71.
- Bonan, G. B. (2008). Forests and climate change. *Forcings, Feedbacks, and the Climate Benefits of Forests*. *Science*, 320(5882), 1444-1449.
- Boudreau, J., Nelson, R. F., Margolis, H. A., Beaudoin, A., Guindon, L., & Kimes, D. S. (2008). Regional aboveground forest biomass using airborne and spaceborne lidar in Québec. *Remote Sensing of Environment*, 112(10), 3876-3890.
- Bouvet, A., Mermoz, S., Le Toan, T., Villard, L., Mathieu, R., Naidoo, L., & Asner, G. P. (2018). An above-ground biomass map of African savannahs and woodlands at 25m resolution derived from ALOS PALSAR. *Remote Sensing of Environment*, 206, 156-173.
- Breiman, L. (2001) *Machine Learning* 45: 5.
- Brown, S. (1997). Estimating biomass and biomass change of tropical forests: a primer (Vol. 134). *Food & Agriculture Org.*
- Brown, S., 2002. Measuring carbon in forests: current status and future challenges. *Environmental pollution*, 116(3), pp.363-372.
- Bryant, R., M. S. Moran, D. P. Thoma, C. D. Holifield Collins, S. Skirvin, M. Rahman, K. Slocum, P. Starks, D. Bosch, M. P. Gonzalez Dugo. (2007). Measuring surface roughness height to parameterize radar backscatter models for retrieval of surface soil moisture. *IEEE Geoscience and Remote Sensing Letters*, 4(1), 137 – 141.
- Canadell, J. G., and M. R. Raupach. (2008). Managing forests for climate change mitigation. *Science*, 320(5882), 1456 – 1457.
- Cartus, O., M. Santoro, J. Kellndorfer. (2012). Mapping forest aboveground biomass in the northeastern United States with ALOS PALSAR dual-polarization L-band. *Remote Sensing of Environment*, 124, 466-478.
- Change, I. P. O. C. (2006). 2006 IPCC guidelines for national greenhouse gas inventories. [2013-04-28]. <http://www.ipcc-nggip.iges.or.jp/public/2006gl/index.html>.
- Chambers, J. Q., J. dos Santos, R. J. Ribeiro, N. Higuchi. (2001). Tree damage, allometric relationships, and above-ground net primary production in central Amazon forest. *Forest Ecology and Management*, 152(1-3), 73 – 84.
- Chauhan, N. S., D. M. Le Vine, R. H. Lang. (1994). Discrete scatter model for microwave radar and radiometer response to corn: Comparison of theory and data. *IEEE Geoscience and Remote Sensing Society*, 32(2), 416 – 426.
- Chauhan, N. S., R. H. Lang, K. J. Ranson. (1991). Radar modeling of a boreal forest. *IEEE Transactions on Geoscience and Remote Sensing*, 29(4), 627 – 638.
- Chave, J., Andalo, C., Brown, S., Cairns, M. A., Chambers, J. Q., Eamus, D., ... & Yamakura, T. (2005). Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia*, 145(1), 87-99.
- Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S., & Perez, R. (2004). Error propagation and scaling for tropical forest biomass estimates. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1443), 409–420.
- Chave, J., Réjou-Méchain, M., Bürquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B., ... & Vieilledent, G. (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. *Global change biology*, 20(10), 3177-3190.
- Chen, Q., Laurin, G. V., & Valentini, R. (2015). Uncertainty of remotely sensed aboveground biomass over an African tropical forest: Propagating errors from trees to plots to pixels. *Remote Sensing of Environment*, 160, 134–143.
- Chilès, J. P., and P. Delfiner. (2012). *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons, Inc: New York.

- Ciais, P., C. Sabine, G. Bala, L. Bopp, V. Brovkin, J. Canadell, A. Chhabra, R. DeFries, J. Gallo-way, M. Heimann, C. Jones, C. Le Quere, R. B. Myreni, S. Piao, and P. Thornton. (2013). Carbon and other biogeochemical cycles. In *Climate Change 2013. The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T.F., D Qin, GK Plattner, M Tignor, SK Allen, J Boschung, A Nauels, Y Xia, V Bex, and PM Midgley (eds)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Cloude, S. (2010). *Polarisation: applications in remote sensing*. Oxford University Press.
- Clark, D. B., and J. R. Kellner (2012). Tropical forest biomass estimation and the fallacy of misplaced concreteness. *Journal of Vegetation Science*, 23, 1191 – 1196.
- Cugois, P. C., J. van Zyl, T. Engman. (1995). Measuring soil moisture with imaging radars. *IEEE Transactions on Geoscience and Remote Sensing*, 33(4), 915 – 926
- Del Frate, F., & Solimini, D. (2004). On neural network algorithms for retrieving forest biomass from SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 42(1), 24-34.
- Dobson, M. C., F. T. Ulaby, L. E. Pierce, T. L. Sharik, K. M. Bergen, J. Kellndorfer, J. R. Kendra, E. Li, Y. C. Lin, A. Nashashibi, K. Sarabandi, P. Siqueira. (1995). Estimation of forest biophysical characteristics in northern Michigan with SIR-C/X-SAR. *IEEE Transactions on Geoscience and Remote Sensing*, 33(4): 877 – 895.
- Dobson, M. C., F. T. Ulaby, T. LeToan, A. Beaudoin, E. S. Kasischke, N. Christensen. (1992). Dependence of radar backscatter on coniferous forest biomass. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2), 412-415.
- Dobson, M. C., F. T. Ulaby. (1986). Preliminary Evaluation of the SIR-B Response to Soil Moisture, Surface Roughness, and Crop Canopy Cover. *IEEE Transactions on Geoscience and Remote Sensing*, GE-24(4): 517 – 526.
- Drake, J. B., Dubayah, R. O., Clark, D. B., Knox, R. G., Blair, J. B., Hofton, M. A., ... & Prince, S. (2002). Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sensing of Environment*, 79(2), 305-319.
- Dubayah, R. O., & Drake, J. B. (2000). lidar remote sensing for forestry. *Journal of Forestry*, 98(6), 44-46.
- Dungan, J. L., Perry, J. N., Dale, M. R. T., Legendre, P., Citron-Pousty, S., Fortin, M. J., ... & Rosenberg, M. (2002). A balanced view of scale in spatial statistical analysis. *Ecography*, 25(5), 626-640.
- Elith, J., S. T. Phillips, T. Hastie, M. Dudik, Y. E. Chee, C. J. Yates. (2010). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17(1), 43 – 57.
- Ene, L. T., Næsset, E., Gobakken, T., Bollandsås, O. M., Mauya, E. W., & Zahabu, E. (2017). Large-scale estimation of change in aboveground biomass in miombo woodlands using airborne laser scanning and national forest inventory data. *Remote Sensing of Environment*, 188, 106-117.
- Englhart, S., Jubanski, J., & Siegert, F. (2013). Quantifying dynamics in tropical peat swamp forest biomass with multi-temporal lidar datasets. *Remote Sensing*, 5(5), 2368-2388.
- FAO. (2015). *Global Forest Resources Assessment 2015*, FAO, Rome, Italy.
- Ferraz, A., S. Saatchi, C. Mallet, V. Meyer. (2016). Lidar detection of individual tree size in tropical forests. *Remote Sensing of Environment*, 183, 318 – 333.
- Freeman, A. and S. L. Durden. (1998). A three-component scattering model for polarimetric SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 36(3), 963 – 973.
- Fung, A. K., and K. S. Chen. (1992). Dependence of the surface backscattering coefficients on roughness, frequency, and polarization states. *International Journal of Remote Sensing*, 13(9), 1663 – 1680.
- García, M., Saatchi, S., Casas, A., Koltunov, A., Ustin, S. L., Ramirez, C., & Balzter, H. (2017). Extrapolating forest canopy fuel properties in the California Rim Fire by combining airborne lidar and Landsat OLI Data. *Remote Sensing*, 9(4), 394.
- Gibbs HK, Brown S (2000) Geographical distribution of woody biomass carbon stocks in tropical Africa: An updated database for 2000. Available at <http://cdiac.ornl.gov/epubs/ndp/ndp055/ndp05b.html> from the Carbon Dioxide Information Center, Oak Ridge National Laboratory, Oak Ridge, TN. Accessed 2009.
- Gibbs, H. K., S. Brown, J. O. Niles, J. A. Foley. (2007). Monitoring and estimating tropical forest carbon stocks: Making REDD a reality. *Environmental Research Letters*, 2(4), 045023.
- Gibbs, H. K., S. Yui, R. Plevin. (2014). New Estimates of Soil and Biomass Carbon Stocks for Global Economic Models. *Working Paper*: Purdue University.
- Hajnsek, I., Kugler, F., Lee, S. K., & Papathanassiou, K. P. (2009). Tropical-forest-parameter estimation by means of Pol-InSAR: The INDREX-II campaign. *Geoscience and Remote Sensing, IEEE Transactions on*, 47(2), 481-493.
- Hallé, F., and R. A. A. Oldeman. (1975). An essay on the architecture and dynamics of growth of tropical trees.
- Hallé, F., R. A. A. Oldeman, and P. B. Tomlinson. (1978). Opportunistic tree architecture. In *Tropical Trees and Forests*, 269 – 331. Springer-Verlag, Berlin, Heidelberg.
- Hallikainen, M. T., F. T. Ulaby, M. C. Dobson, M. A. El-rayes, L. Wu. (1985). Microwave dielectric behavior of wet soil – part 1: empirical models and experimental observations. *IEEE Transactions on Geoscience and Remote Sensing*, GE-23(1), 25 – 34.
- Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., ... & Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *science*, 342(6160), 850-853.
- Harding, D. J., & Carabajal, C. C. (2005). ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophysical research letters*, 32(21).

- Harris, N. L., Brown, S., Hagen, S. C., Saatchi, S. S., Petrova, S., Salas, W., ... & Lotsch, A. (2012). Baseline map of carbon emissions from deforestation in tropical regions. *Science*, 336(6088), 1573–1576.
- Hawkins, D. M., Basak, S. C., & Mills, D. (2003). Assessing model fit by cross-validation. *Journal of chemical information and computer sciences*, 43(2), 579–586.
- Hayashi, M., Saigusa, N., Yamagata, Y., & Hirano, T. (2015). Regional forest biomass estimation using ICESat/GLAS spaceborne lidar over Borneo. *Carbon Management*, 1–15.
- Heath, L. S., Smith, J. E., Skog, K. E., Nowak, D. J., & Woodall, C. W. (2011). Managed Forest Carbon Estimates for the US Greenhouse Gas Inventory, 1990–2008. *Journal of Forestry*, 109(3), 167–173.
- Hensley, S., Oveisgharan, S., Saatchi, S., Simard, M., Ahmed, R., & Haddad, Z. (2014). An error model for biomass estimates derived from polarimetric radar backscatter. *IEEE Transactions on Geoscience and Remote Sensing*, 52(7), 4065–4082.
- Hoekman, D. H., & Quiriones, M. J. (2000). Land cover type and biomass classification using AirSAR data for evaluation of monitoring scenarios in the Colombian Amazon. *Geoscience and Remote Sensing, IEEE Transactions on*, 38(2), 685–696.
- Houghton, R. A., A. Baccini, W. S. Walker. (2018). Where is the residual terrestrial carbon sink? *Global Change Biology*, 24(8), 3277 – 3279.
- Hyde, P., Dubayah, R., Walker, W., Blair, J. B., Hofton, M., & Hunsaker, C. (2006). Mapping forest structure for wildlife habitat analysis using multi-sensor (lidar, SAR/InSAR, ETM+, Quickbird) synergy. *Remote Sensing of Environment*, 102(1), 63–73.
- Isaaks, E., and R. Srivastava. (1990). *An Introduction to Applied Geostatistics*. Oxford University Press, Oxford, UK.
- Karam, M. A., A. K. Fung, R. H. Lang, N. S. Chauhan. (1992). A microwave scattering model for layered vegetation. *IEEE Transactions on Geoscience and Remote Sensing*, 30(4), 767–784.
- Karam, M. A., A. K. Fung, R. H. Lang, N. S. Chauhan. (1992). A microwave scattering model for layered vegetation. *IEEE Transactions on Geoscience and Remote Sensing*, 30(4), 767–784.
- Karam, M. A., and A. K. Fung. (1983). Scattering from randomly oriented circular discs with application to vegetation. *Radio Science*, 18(4), 557 – 565.
- Keller, M., M. Palace, G. Hurtt. (2001). Biomass estimation in the Tapajos National Forest, Brazil: Examination of sampling and allometric uncertainties. *Forest Ecology and Management* 154(3), 371–382.
- Ketterings, Q.M., Coe, R., Noordwijk, M., Ambagau, Y., Palm, CA (2001) Reducing uncertainty in the use of allometric biomass equations for predicting aboveground tree biomass in mixed secondary forests. *Forest Ecology and Management*, 146, 199–209.
- Lang, R. (1981). Electromagnetic backscattering from a sparse distribution of lossy dielectric scatterers. *Radio Science*, 16(1), 15 – 30.
- Lang, R. H., and J. S. Sighu. (1983). Electromagnetic backscattering from a layer of vegetation: A discrete approach. *IEEE Transactions on Geoscience and Remote Sensing*, GE-21(1), 62 – 71.
- Lawrence, M, RF McRoberts, E Tomppo, T Gschwantner, K Gabler. 2010. Comparisons of National Forest Inventories. In: Tomppo E., Gschwantner T., Lawrence M, R McRoberts (eds) *National Forest Inventories*. Springer, Dordrecht.
- Lee, J. S., M. R. Grunes, G. de Grandi. (1999). Polarimetric SAR speckle filtering and its implication for classification. *IEEE Geoscience and Remote Sensing*, 37(5), 2363 – 2373.
- Le Quere, C., R. M. Andrew, P. Friedlingstein, S. Sitch, J. Hauck, J. Pongratz, P. Pickers, J. I. Korsbakken, G. P. Peters, J. G. Canadell, A. Arneth, V. K. Arora, L. Barbero, A. Bastos, L. Bopp, F. Chevallier, L. P. Chini, P. Clais, S. C. Doney, T. Gkritzalis, D. S. Goll, I. Harris, V. Haverd, F. M. Hoffman, M. Hoppema, R. A. Houghton, T. Illyina, A. K. Jain, T. Johnsen, C. D. Jones, E. Kato, R. F. Keeling, K. K. Goldewijk, P. Landschutzer, N. Lefevre, S. Lienert, D. Lombardozzi, N. Metzi, D. R. Munro, J. E. M. S. Nabel, S. Nakaoka, C. Meill, A. Olsen, T. Ono, P. Patra, A. Peregon, W. Peters, P. Peylin, B. Pfeil, D. Pierrot, B. Poulter, G. Rehder, L. Resplandy, E. Robertson, M. Rocher, C. Rodenbeck, U. Schuster, J. Schwinger, R. Seferian, I. Skjelvan, T. Steinhoff, A. Sutton, P. P. Tans, H. Tian, B. Tilbrook, F. N. Tubiello, I. T. van der Laan-Luijkx, G. R. van der Werf, N. Viovy, A. P. Walker, A. J. Wiltsshire, R. Wright, S. Zaehle. (2018). Global carbon budget 2018. *Earth System Science Data*, 10, 2141–2194.
- Le Toan, T., A. Beaudoin, J. Riou, D. Guyon. (1992). Relating forest biomass to SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2), 403–411.
- Le Toan, T., A. Beaudoin, J. Riou, D. Guyon. (1992). Relating forest biomass to SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2), 403–411.
- Le Toan, T., S. Quegan, M. W. J. Davidson, H. Balzter, P. Paillou, K. Papathanassiou, S. Plummer, F. Rocca, S. Saatchi, H. Shugart, L. Ulander. (2011). The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sensing of Environment*, 115(11), 2850 – 2860.
- Lefsky, M. A. (2010). A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System. *Geophysical Research Letters*, 37(15).
- Lefsky, M. A., Cohen, W. B., Parker, G. G., & Harding, D. J. (2002). lidar Remote Sensing for Ecosystem Studies lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists. *BioScience*, 52(1), 19–30.
- Lefsky, M. A., Keller, M., Pang, Y., De Camargo, P. B., & Hunter, M. O. (2007). Revised method for

- forest canopy height estimation from Geoscience Laser Altimeter System waveforms.* *Journal of Applied Remote Sensing*, 1(1), 013537-013537.
- Mascaro, J., Asner, G. P., Muller-Landau, H. C., van Breugel, M., Hall, J., & Dahlin, K. (2011b). Controls over aboveground forest carbon density on Barro Colorado Island, Panama. *Biogeosciences*, 8(6), 1615-1629.
- Matzler, C. (1994). Microwave (1-100 GHz) dielectric model of leaves. *IEEE Transactions on Geoscience and Remote Sensing* 32(4): 947 – 949.
- McMahon, T. (1973). Size and shape in biology. *Science* 179(4079), 1201 – 1204.
- McRoberts, R. E., Chen, Q., & Walters, B. F. (2017). Multivariate inference for forest inventories using auxiliary airborne laser scanning data. *Forest Ecology and Management*, 401, 295-303.
- McRoberts, R. E., Næsset, E., & Gobakken, T. (2013). Inference for lidar-assisted estimation of forest growing stock volume. *Remote Sensing of Environment*, 128, 268-275.
- Mermoz, S., Réjou-Méchain, M., Villard, L., Le Toan, T., Rossi, V., & Gourlet-Fleury, S. (2015). Decrease of L-band SAR backscatter with biomass of dense forests. *Remote Sensing of Environment*, 159, 307-317.
- Mermoz, S., T. L. Toan, L. Villard, M. Rejou-Mechain, J. Seifert-Granzin. (2014). Biomass assessment in the Cameroon savanna using ALOS PALSAR data. *Remote Sensing of Environment*, 155, 109-119.
- Meyer, V., Saatchi, S. S., Chave, J., Dalling, J. W., Bohlman, S., Fricker, G. A., ... & Hubbell, S. (2013). Detecting tropical forest biomass dynamics from repeated airborne lidar measurements. *Biogeosciences*, 10(8), 5421-5438.
- Michelakis, D., N. Stuart, G. Lopez, V. Linares, and I. H. Woodhouse. (2014). Local-scale mapping of biomass in tropical lowland pine savannas using ALOS PALSAR. *Forests*, 5(9), 2377-2399.
- Minh, D. H. T., Tebaldini, S., Rocca, F., Le Toan, T., Villard, L., & Dubois-Fernandez, P. C. (2015). Capabilities of BIOMASS tomography for investigating tropical forests. *Geoscience and Remote Sensing, IEEE Transactions on*, 53(2), 965-975.
- Mironov, V. L., M. C. Dobson, V. H. Kaupp, S. A. Komarov, V. N. Kleshchenko. (2004). Generalized refractive mixing dielectric model for moist soils. *IEEE Transactions on Geoscience and Remote Sensing*, 42(4), 773 – 785.
- Mitchard, E. T. A., Saatchi, S. S., Lewis, S. L., Feldpausch, T. R., Woodhouse, I. H., Sonké, B., ... & Meir, P. (2011). Measuring biomass changes due to woody encroachment and deforestation/degradation in a forest-savanna boundary region of central Africa using multi-temporal L-band radar backscatter. *Remote Sensing of Environment*, 115(11), 2861-2873.
- Mitchard, E. T. A., S. S. Saatchi, I. H. Woodhouse, G. Nangendo, N. S. Ribeiro, M. Williams, C. M. Ryan, S. L. Lewis, T. R. Feldpausch, and P. Meir. Using satellite radar backscatter to predict above-ground woody biomass: A consistent relationship across four different African landscapes. *Geophysical Research Letters*, 36, L23401.
- Mitchard, E. T., Feldpausch, T. R., Brienen, R. J., Lopez-Gonzalez, G., Monteagudo, A., Baker, T. R., ... & Pardo Molina, G. (2014). Markedly divergent estimates of Amazon forest carbon density from ground plots and satellites. *Global ecology and biogeography*, 23(8), 935-946.
- Mitchard, E. T., Saatchi, S. S., Baccini, A., Asner, G. P., Goetz, S. J., Harris, N. L., & Brown, S. (2013). Uncertainty in the spatial distribution of tropical forest biomass: a comparison of pan-tropical maps. *Carbon balance and management*, 8(10), 1-13.
- Mokany, K., Raison, R., & Prokushkin, A. S. (2006). Critical analysis of root: Shoot ratios in terrestrial biomes. *Global Change Biology*, 12(1), 84–96.
- Morel, A. C., Saatchi, S. S., Malhi, Y., Berry, N. J., Banin, L., Burslem, D., Nilus, R., & Ong, R. C. (2011). Estimating aboveground biomass in forest and oil palm plantation in Sabah, Malaysian Borneo using ALOS PALSAR data. *Forest Ecology and Management*, 262, 1786–1798.
- Moussavi, M. S., W. Abdalati, T. Scambos, A. Neuenschwander. (2014). Applicability of an automatic surface detection approach to micro-pulse photon-counting lidar altimetry data: implications for canopy height retrieval from future ICESat-2 data. *International Journal of Remote Sensing*, 35(13), 5263 – 5279.
- Næsset, E., Gobakken, T., Solberg, S., Gregoire, T. G., Nelson, R., Ståhl, G., & Weydahl, D. (2011). Model-assisted regional forest biomass estimation using lidar and InSAR as auxiliary data: A case study from a boreal forest area. *Remote Sensing of Environment*, 115(12), 3599-3614.
- Næsset, E., Ørka, H. O., Solberg, S., Bollandsås, O. M., Hansen, E. H., Mauya, E., ... & Gobakken, T. (2016). Mapping and estimating forest area and aboveground biomass in miombo woodlands in Tanzania using data from airborne laser scanning, TanDEM-X, RapidEye, and global forest maps: A comparison of estimated precision. *Remote sensing of Environment*, 175, 282-300.
- Neigh, C. S., Nelson, R. F., Ranson, K. J., Margolis, H. A., Montesano, P. M., Sun, G., ... & Andersen, H. E. (2013). Taking stock of circumboreal forest carbon with ground measurements, airborne and spaceborne lidar. *Remote Sensing of Environment*, 137, 274-287.
- Nelson, R., Ranson, K. J., Sun, G., Kimes, D. S., Kharuk, V., & Montesano, P. (2009). Estimating Siberian timber volume using MODIS and ICESat/GLAS. *Remote Sensing of Environment*, 113(3), 691-701.
- Neumann, M., S. Saatchi, L. M. H. Ulander, J. E. S. Fransson. (2012). Assessing performance of L- and P-band polarimetric interferometric SAR data in estimating boreal forest above-ground biomass. *IEEE Transactions on Geoscience and Remote Sensing* 50(3), 714 – 726.

- Ngomanda, A., N. L. E. Obiang, J. Lebamba, Q. M. Mavououlou, H. Gomat, G. S. Mankou, J. Loumeto, D. M. Ipanga, F. K. Ditsouga, R. Z. Koumba, K. H. B. Bobe, C. M. Okouyi, R. Nyangadouma, N. Lepengue, B. Mbatchi, N. Picard. Site-specific versus pantropical allometric equations: Which option to estimate the biomass of a moist central African forest? *Forest Ecology and Management*, 312, 1-9.
- Nogueira, E. M., Nelson, B. W., Fearnside, P. M., França, M. B., & de Oliveira, A. C. A. (2008). Tree height in Brazil's 'arc of deforestation': Shorter trees in south and southwest Amazonia imply lower biomass. *Forest Ecology and Management*, 255(7), 2963–2972.
- Oh, Y., K. Sarabandi, F. T. Ulaby. (1992). An empirical model and an inversion technique for radar scattering from bare soil surfaces. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2), 370 – 381.
- Olofsson, P., Foody, G. M., Stehman, S. V. and Woodcock, C. E. (2013). Making better use of accuracy data in land change studies: estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, 129:122-131
- Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'Amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettenberg, W. W., Hedao, P., Kassem, K. R. (2001). Terrestrial ecoregions of the world: a new map of life on Earth. *Bioscience*, 51(11), 933-938.
- Peplinski, N. R., F. T. Ulaby, M. C. Dobson. (1995). Dielectric properties of soils in the 0.3 – 1.3-GHz range. *IEEE Transactions on Geoscience and Remote Sensing*, 33(3), 803 – 807.
- Phillips, S. J., R. P. Anderson, R. E. Schapire. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3-4), 231-259.
- Raney, R. K. (2007). Hybrid-polarity SAR architecture. *IEEE Transactions on Geoscience and Remote Sensing*, 45(11), 3397-3404.
- Ranson, K. J., G. Sun, J. F. Weishampel, R. G. Knox. (1997). Forest biomass from combined ecosystem and radar backscatter modeling. *Remote Sensing of Environment*, 59(1), 118 – 133.
- Ranson, K. J., G. Sun. 1994. Mapping biomass of a northern forest using multifrequency SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 32(2), 388 – 396.
- Reese, H., M. Nilsson, T. G. Pahlen, O. Hagner, S. Joyce, U. Tingelof, M. Egberth, and H. Olsson. Countrywide estimates of forest variables using satellite data and field data from the national forest inventory. *AMBIO: A Journal of the Human Environment*, 32(8).
- Rosette, J., P. North, J. Suarez. (2008). Stemwood Volume Estimates for a Mixed Temperate Forest using Satellite Lidar. *Japan Society of Forest Planning*, 13, 205 – 214. Rosette, J., J. Suarez, R. Nelson, S. Los, B. Cook, and P. North. (2012). Lidar remote sensing for biomass assessment. In *Remote Sensing of Biomass – Principles and Applications* [L. Fatoyinbo (ed)], Intech.
- Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T., Salas, W., ... & Morel, A. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences*, 108(24), 9899-9904.
- Saatchi, S. S., Houghton, R. A., Dos Santos Alvala, R. C., Soares, J. V., & Yu, Y. (2007). Distribution of aboveground live biomass in the Amazon basin. *Global Change Biology*, 13(4), 816-837.
- Saatchi, S. S., R. H. Land. (1989). Self consistent approach to average waves in a one dimensional discrete random medium. *PIERS: Progress in Electromagnetics Research Symposium. Proceedings of the symposium held July 25-26, 1989 in Boston, Massachusetts, USA*. P. 277.
- Saatchi, S. S., D. M. Le Vine, R. H. Lang. (1994). Microwave backscattering and emission model for grass canopies. *IEEE Transactions on Geoscience and Remote Sensing*, 32(1), 177 – 186.
- Saatchi, S., Marlair, M., Chazdon, R. L., Clark, D. B., & Russell, A. E. (2011b). Impact of spatial variability of tropical forest structure on radar estimation of aboveground biomass. *Remote Sensing of Environment*, 115(11), 2836-2849.
- Saatchi, S., Mascaro, J., Xu, L., Keller, M., Yang, Y., Duffy, P., ... & Schimel, D. (2015). Seeing the forest beyond the trees. *Global Ecology and Biogeography*, 24(5), 606-610.
- Saatchi, S. S., and K. C. McDonald. (1997). Coherent effects in microwave backscattering models for forest canopies. *IEEE Transactions on Geoscience and Remote Sensing*, 35(4), 1032- 1044.
- Saatchi, S.S. and M. Moghaddam. (2000). Estimation of crown and stem water content and biomass of boreal forest using polarimetric SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 38(2), 697 – 709.
- Sabine, C. L., M. Heimann, P. Artaxo, D. C. E. Bakker, CA Chen, C. B. Field, N. Gruber, C. Le Quere, R. G. Prinn, J. E. Richey, P. R. Lankao, J. A. Sathaye, R. Valentini. (2004). Current status and past trends of the global carbon cycle. In *The Global Carbon Cycle: Integrating Humans, Climate, and the Natural World* [C. B. Field and M. R. Raupach (eds)], Island Press, Washington, D.C.
- Sandberg, G., L. M. H. Ulander, J.E. S. Fransson, J. Holmgren, T. Le Toan. (2011). L- and P-band backscatter intensity for biomass retrieval in hemiboreal forest. *Remote Sensing of Environment* 115(11), 2874 – 2886.
- Santoro, M., Askne, J., G. Smith, J. E. S. Fransson. (2002). Stem volume retrieval in boreal forests from ERS-1/2 interferometry. *Remote Sensing of Environment*, 81(1), 19-35.
- Schimel, D., B. B. Stephens, J. B. Fisher. (2015). Effect of increasing CO<sub>2</sub> on the terrestrial carbon cycle. *Proceedings of the National Academy of Sciences of the United States of America*, 112(2), 436 – 441.

- Schimel, D., R. Pavlick, J. B. Fisher, G. P. Asner, S. Saatchi, P. Townsend, C. Miller, C. Frankenberger, K. Hibbard, P. Cox. (2014). Observing terrestrial ecosystems and the carbon cycle from space. *Global Change Biology*, 21(5), 1762 – 1776.
- Sellers, P. J., D. S. Schimel, B. Moore III, J. Liu, A. Eldering. (2018). Observing carbon cycle-climate feedbacks from space. *Proceedings of the National Academy of Sciences of the United States of America*, 115(31), 7860 – 7868.
- Shi, J., J. Wang, A. Y. Hsu, P. E. O'Neill, E. T. Engman. (1997). Estimation of bare surface soil moisture and surface roughness parameter using L-band SAR image data. *IEEE Transactions on Geoscience and Remote Sensing*, 35(5), 1254 – 1266.
- Shugart, H. H., Saatchi, S., & Hall, F. G. (2010). Importance of structure and its measurement in quantifying function of forest ecosystems. *Journal of Geophysical Research: Biogeosciences* (2005–2012), 115(G2).
- Small, D. (2011). Flattening gamma: Radiometric terrain correction for SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 49(8), 3081 – 3093.
- Small, D., F. Holecz, E. Meier, D. Nuesch. (1998). Absolute radiometric correction in rugged terrain: A plea for integrated radar brightness. *IGARSS '98. Sensing and Managing the Environment. 1998 IEEE International Geoscience and Remote Sensing Symposium Proceedings*.
- Stinson, G., Kurz, W. A., Smyth, C. E., Neilson, E. T., Dymond, C. C., Metsaranta, J. M., . . . & Blain, D. (2011). An inventory-based analysis of Canada's managed forest carbon dynamics, 1990 to 2008. *Global change biology*, 17(6), 2227–2244.
- Sun, G., and K. J. Ranson. (1995). Three-dimensional radar backscatter model of forest canopies. *IEEE Transactions on Geoscience and Remote Sensing*, 33(2), 372 – 382.
- Tittmann, P., & Saatchi, S. S. (2015). Tool for measuring ALBF using remote sensing v1.0, (March), 1–32. Accessed Dec 2018: <https://verra.org/wp-content/uploads/2018/03/VT0005-Tool-for-measuring-ALBF-using-remote-sensing-v1.0.pdf>
- Tomppo, E., M. Gschwantner, M. Lawrence, and R. E. McRoberts. (2010). *National Forest Inventories: Pathways for Common Reporting*. Springer, Dordrecht.
- Tomppo, E., J. Heikkinen, H. M. Henttonen, A. Ihlainen, M. Katila, H. Makela, T. Tuomainen, N. Vainikainen. (2011). *Designing and Conducting a Forest Inventory of Finland*. Springer, Dordrecht.
- Tsang, L., & Kong, J. A. (2004). *Scattering of electromagnetic waves: advanced topics* (Vol. 26). John Wiley & Sons.
- Tsang, L., J. A. Kong, and R. T. Shin. 1985. *Theory of Microwave Remote Sensing*. John Wiley, New York.
- Ulaby, F. T., A. Aslam, M. C. Dobson. (1982). Effects of vegetation cover on the radar sensitivity to soil moisture. *IEEE Transactions on Geoscience and Remote Sensing*, GE-20(4), 476 – 481.
- Ulaby, F. T., and M. A. El-rayes. (1987). Microwave dielectric spectrum of vegetation – Part II: Dual-dispersion model. *IEEE Transactions on Geoscience and Remote Sensing*, GE-25(5), 550 – 557.
- Ulaby, F. T., M. C. Dobson. (1989). *Handbook of Radar Scattering Statistics for Terrain*. Artech House, Norwood MA.
- Ulaby, F. T., T. H. Bengal, M. C. Dobson, J. R. East, J. B. Garvin, D. L. Evans. (1990). Microwave dielectric properties of dry rocks. *IEEE Transactions on Geoscience and Remote Sensing*, 28(3), 325 – 336.
- Ulander, L. M. H. (1996). Radiometric slope correction of synthetic-aperture radar images. *IEEE Transactions on Geoscience and Remote Sensing*, 34(5), 115 – 1122.
- Villard, L., & Le Toan, T. (2015). Relating P-Band SAR Intensity to Biomass for Tropical Dense Forests in Hilly Terrain:  $y_0$  or  $t_0$ . *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(1), 214-223.
- Weisbin, C. R., Lincoln, W., & Saatchi, S. (2014). A systems engineering approach to estimating uncertainty in above-ground biomass (AGB) derived from remote-sensing data. *Systems Engineering*, 17(3), 361–373.
- West, G. B., and J. H. Brown. (2005). The origin of allometric scaling laws in biology from genomes to ecosystems: towards a quantitative unifying theory of biological structure and organization. *Journal of Experimental Biology* 208: 1575-1592.
- Wigneron, J. P., P. Ferrazzoli, A. Olioso, P. Bertuzzi, A. Chanzy. (1999). A simple approach to monitor crop biomass from C-band radar data. *Remote Sensing of Environment*, 69(2), 179-188.
- Wulder, M. A., and S. E. Franklin. (2012). *Remote Sensing of Forest Environments: Concepts and Case Studies*, Springer, New York.
- Xu, L., Saatchi, S. S., A. Shapiro, V. Meyer, A. Ferraz, Y. Yang, J. Bastin, N. Banks, P. Boeckx, H. Verbeeck, S. L. Lewis, E. T. Muanza, E. Bongwele, F. Kayembe, D. Mbenza, L. Kalaf, F. Mukendi, F. Ilunga, D. Ebuta. (2017). Spatial distribution of carbon stored in forests of the Democratic Republic of Congo. *Scientific Reports*, 7, 15030.
- Xu, L., Saatchi, S. S., Yang, Y., Myneni, R. B., Frankenberger, C., Chowdhury, D., Bi, J. (2015). Satellite observation of tropical forest seasonality: Spatial patterns of carbon exchange in Amazonia. *Environmental Research Letters*, 10, 084005.
- Xu, L., Saatchi, S. S., Yang, Y., Yu, Y., & White, L. (2016). Performance of non-parametric algorithms for spatial mapping of tropical forest structure. *Carbon balance and management*, 11(1), 18.

- Yu, Y. and S. Saatchi. (2016). Sensitivity of L-band SAR Backscatter to aboveground biomass of global forests. *Remote Sensing*, 8(6), 522.
- Yu, Y., Saatchi, S. Fore, A. et al. (*In Review*) Carbon Stocks in World's Forests and Shrublands.
- Yueh, S. H., J. A. Kong, J. K. Jao, R. T. Shin, T. L. Toan. (1992). Branching model for vegetation. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2), 390 – 402.
- Yun, S.-H., Hudnut, K., Owen, S., Webb, F., Simons, M., Sacco, P., Gurrola, E., Manipon, G., Liang, C., and Fielding, E., 2015b, Rapid Damage Mapping for the 2015 Mw 7.8 Gorkha Earthquake Using Synthetic Aperture Radar Data from COSMO-SkyMed and ALOS-2 Satellites: *Seismological Research Letters*, v. 86, no. 6, p. 1549-1556.
- Zeng, W., E. Tomppo, S. P. Healey, and K. V. Gadow. (2015). The national forest inventory in China: history-results-international context. *Forest Ecosystems* 2:23.
- Zolkos, S. G., S. J. Goetz, R. Dubayah. (2013). A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. *Remote Sensing of Environment*, 128, 289 – 298.
- Zhang X., Ni-meister W. (2014) *Remote Sensing of Forest Biomass*. In: Hanes J. (eds) *Biophysical Applications of Satellite Remote Sensing*. Springer Remote Sensing/Photogrammetry. Springer, Berlin, Heidelberg

# APPENDIX D

## Mapping Forest Biomass with Radar Remote Sensing – Chapter 5 Training Module

### 1 SOFTWARE AND DATA SOURCES

Software:

- QGIS, Microsoft Excel

Remote Sensing Data:

- ALOS 2 / PALSAR-2 RTC annual mosaic product
- Lidar data that overlaps with some of the ALOS 2 / PALSAR-2 data
- Forest inventory plot data that overlaps with some of the lidar data.

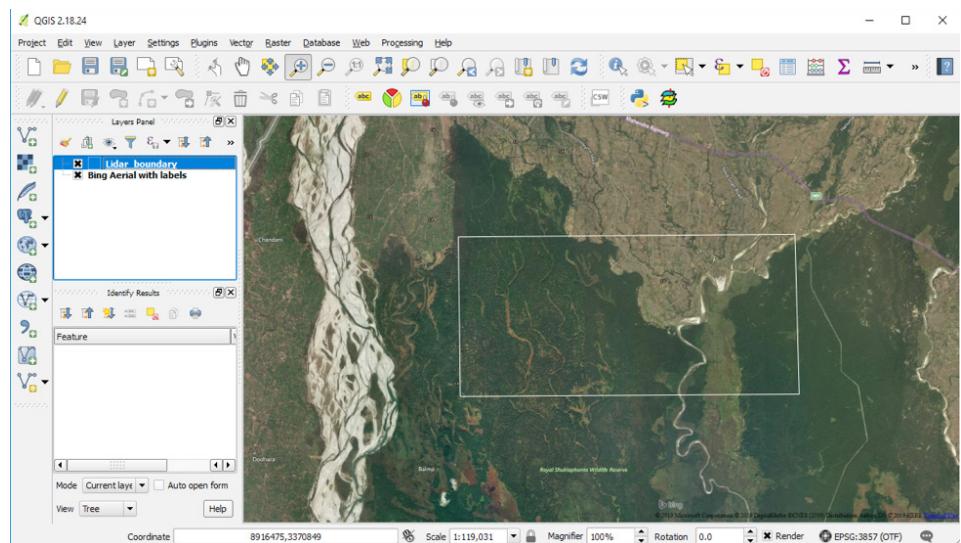
**Note:** All of the data to run the tutorial is included in the data.zip file for this chapter hosted on the SAR Handbook website. If you would like to do the same processing on your own data, you will need a SAR RTC product (steps for downloading from ASF or JAXA can be found in the tutorial for Chapter 6), lidar data covering part of your area of interest, and forest inventory data that overlaps with the lidar data.

### 2 AIRBORNE LIDAR AND INVENTORY PLOT DATA

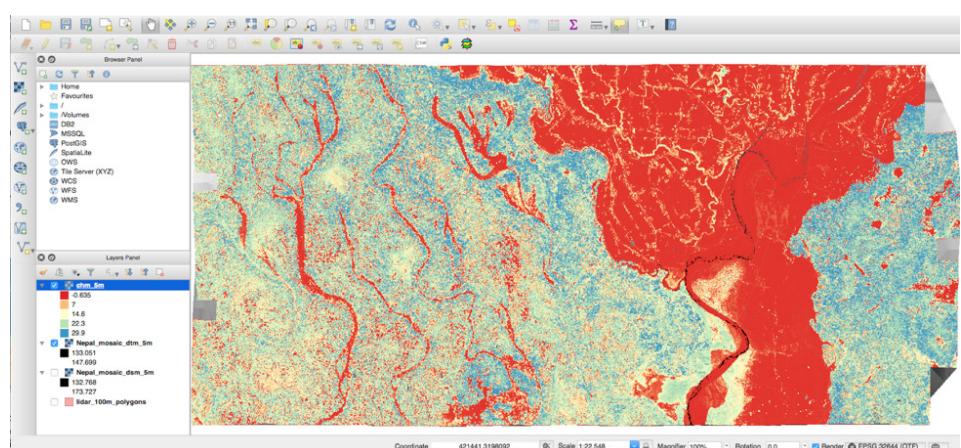
Download the Data.zip file for Chapter 5 from the SAR Handbook website and unzip it in the desired location. The zipped file consists of three folders: Lidar, Ground\_Plots, and ALOS with example data from Nepal.

**Step 1:** Open QGIS and add the shape file D:\NepalData\Lidar\Lidar\_boundary. Add Google Aerial with Labels as the background (Web > OpenLayers plugin > Google Maps). Examine the location of the Lidar\_boundary file to understand the study site location and the landscape across Western Nepal.

**Step 2:** Add the LIDAR DTM (Lidar > dtm\_5m.tif) and DSM (Lidar > dsm\_5m.tif) files to QGIS (Layer > add Layer > add Raster Layer). The images are provided as geotiffs with 5 meter spatial resolution.



**Figure 1.1** QGIS interface displaying the study area in western Nepal. The white line represents the lidar boundary.



**Figure 1.2** Canopy Height Model created using the lidar-derived DSM and DTM. The image shows the Spectral color band stretched using min/max values.

**Step 3:** In this step, we will produce a Canopy Height Model (CHM) by subtracting the DTM from DSM. Using the raster calculator tool in QGIS (Raster > Raster Calculator), enter the following equation:

$$\text{Dsm\_5m} - \text{Dtm\_5m}$$

Since you will be using this file multiple times, you may want to create a results folder and save the CHM to this new folder under the file name chm\_5m.tif.

**Step 4:** Double click the CHM image name (chm\_5m) in the layers panel of QGIS and explore the Layer

Properties, including projection, display, and other image characteristics. The projection is in UTM Zone 44 N, Datum: WGS-84.

**Step 5:** From here you can also apply a color scale to the CHM that will highlight short to tall forests in the study region. With Layer Properties still open, go to Style > Render type > Singleband pseudocolor > Load min/max values > Min/max > Load > Color > Spectral (or any other color scale you like) > Apply. Your result should look similar to **Figure 1.2**.

Step 6: Use the Profile Tool to explore height distribution in the data. The Profile Tool is a Plugin and needs to be installed by going to Plugins > Manage and Install Plugins > Profile Tool > Install Plugin. After installation, the Profile icon  will appear in the toolbar. Click on the Profile icon > Select the CHM image in the Layers Panel > Click Add Layer > Draw a line at any place over the image (double click to end the line). Depending on where you draw your line, your result should look something like **Figure 1.3**.

Note: This image shows a typical example of a CHM profile which can be achieved drawing a line. The profile image can be saved for future use. Also be careful in interpreting your profile chart. It will start wherever you draw your line and move in whatever direction you ended your line. If you draw your line from east to west (instead of west to east), your profile will start in the east and move toward the west. In Figure 1.3, the line was drawn from east to west; therefore, the right part of the profile starts in the east and moves westward.

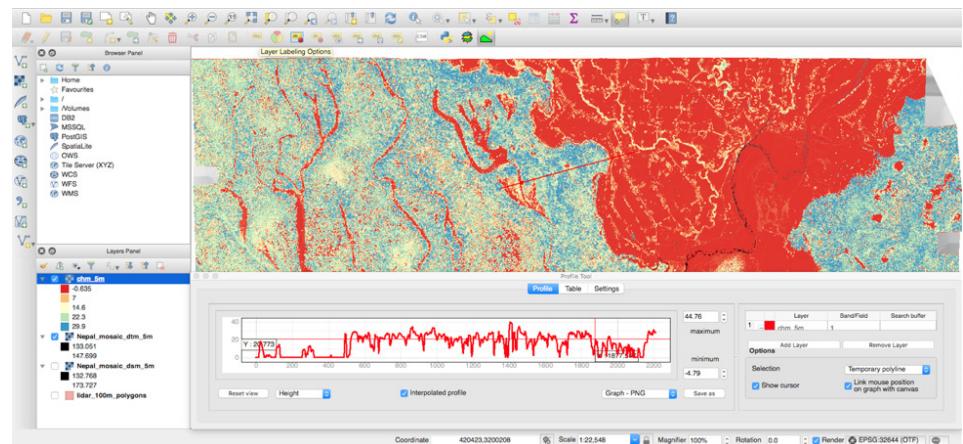
Step 7: Play around with drawing different lines across different parts of the scenes. Where do you see the highest canopy height? The lowest? How does this pattern change across the landscape?

### 3 LIDAR BIOMASS MODEL DEVELOPMENT

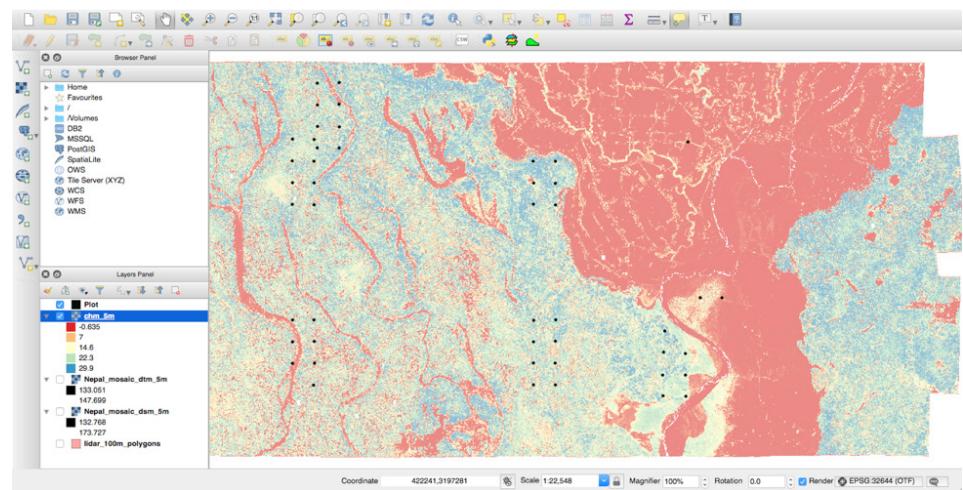
Step 1: Add the shapefile containing the ground plot data (Ground\_Plots > plot.shp) to the QGIS (Layer > Add Layer > Add Vector Layer). There are 47 small plots available for the study area. Each plot represents a 20 m radius ground footprint.

Step 2: In this step we use the Zonal Statistics Tool to extract the lidar-derived mean canopy height from each plot. Go to Processing > Tools > Search for Zonal Statistics. Once you open up the Zonal Statistics Tool, set your raster layer to chm\_5m.tif and your vector layer to plot.shp and run the tool.

Step 3: A new shapefile (named Zonal Statistics) will be added to your Layers Panel. You can right click the new layer and select Open Attribute Table to view the data associated with each plot. If you scroll all the way



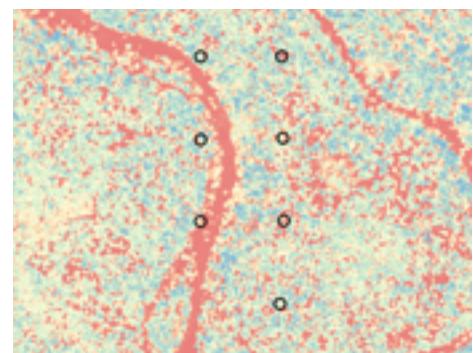
**Figure 1.3** Results of the profile tool showing canopy height variation across the red line drawn in the center of the scene.



**Figure 1.4** Distribution of field plots (denoted by black dots) across the study area.

to the right of the attribute table, you will see the zonal statistics you just calculated. Most important for the next steps is the \_mean column, which contains the average canopy height for each plot.

Step 4: Next, you want to create an Excel spreadsheet from the Zonal Statistics attribute table. One method you can use is to install the XYTools plugin (Plugins > Manage and Install Plugins > search for XYTools > Install Plugin). Next, make sure the Zonal Statistics layer is highlighted in your Layers Panel. Go to Vector > XYTools > Save attribute table as Excel file. Check the following fields: Object ID, AGB, \_std, and \_mean. Alternatively, you can right click the Zonal statistics layer in the Layers Panel and select Save As to save the data as a csv file,



**Figure 1.5** A zoomed view of one set of field plots. Later in the exercise, we will compute the average of the pixel values that fall within the field plot set using the zonal histogram tool.

OBJECTID	_mean	_std	AGB
1	7.85	7.49	97.77
2	16.91	10.45	290.54
3	18.54	9.26	263.21
4	18.09	8.51	163.86
5	9.73	9.12	172.78
6	14.20	8.93	210.74
7	10.09	11.72	68.58
8	19.36	5.97	176.68
9	18.65	4.99	277.04
10	16.30	7.34	146.01
11	15.67	10.48	268.26
12	14.44	11.25	250.54
13	23.47	4.73	693.69
14	12.61	10.09	103.73
15	20.79	5.25	325.56
16	12.06	9.66	109.76
17	19.56	7.90	810.71
18	23.26	3.70	687.63
19	19.08	6.65	293.98
20	17.72	7.76	399.16
21	20.73	4.74	258.01
22	23.15	7.44	591.53
23	24.49	2.96	579.44
24	24.98	5.60	530.78
25	18.08	5.29	344.34
26	25.93	5.96	657.66
27	21.93	6.20	764.19
28	21.94	5.97	441.79
29	25.34	6.43	582.33
30	16.16	8.53	238.40
31	23.44	5.13	674.51
32	10.39	7.11	196.14
33	20.86	8.31	249.41
34	17.41	9.66	209.03
35	17.85	6.91	243.84
36	11.23	11.06	257.55
37	29.32	4.51	858.85
38	20.25	5.90	139.54
39	22.96	5.01	412.32
40	21.82	5.74	667.84
41	21.78	4.40	287.50
42	19.47	1.95	233.49
43	19.46	4.95	330.82
44	20.05	1.84	395.18
45	3.88	3.39	63.51
46	8.17	4.75	138.14
47	3.11	2.41	80.94

**Table 1** Plot values for \_mean, \_std, and AGB.

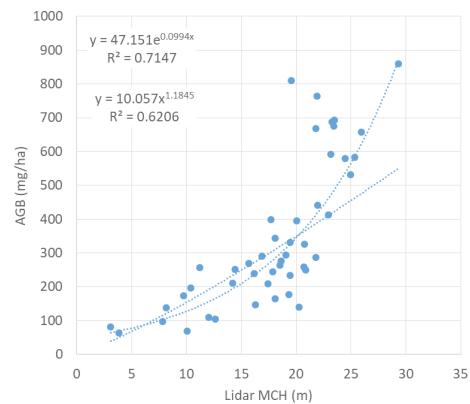
which can be opened in Excel. The values in your table should look similar to those in **Table 1**.

**Step 5:** Open your Zonal Statistics data in Excel. Here we want to plot the aboveground biomass (AGB) of the plots with respect to the MCH (\_mean) values. You can do this by selecting the data in the AGB and \_mean columns and creating a scatterplot using the scatterplot tool (Insert > Scatterplot  )

**Step 6:** Now we want to fit the best model to present the data. You can do this by right clicking one of the points in the scatterplot and selecting add trendline. From the format trendline pane, you can evaluate various trendline options for the best fit and display the equations that go with each trendline. In this case, the best model is a power-law. However, as the plots are small and the sensitivity of height to capture the high biomass values of small plots may saturate, use other functions such as exponential (as shown in the inset graph on this page). During the training given at SERVIR-HKH, there was a consensus among the participants with local knowledge for limiting the maximum biomass in the study site to 1000 Mg/ha.

**Step 7:** Here we create a CHM image of 40 m pixel resolution by performing an 8x8 resampling of the chm\_5m image (There are many ways to do this in QGIS, but one is to go to Processing > Toolbox > SAGA > Raster tools > Resampling, select chm\_5m as your Grid, leave upscaling and downscaling method as Nearest Neighbor, make the Cellsize 40, and run the tool). Save this file as chm\_40m. This process will produce a 40 m average MCH (mean top canopy height) image.

**Step 8:** Now we are going to create an AGB map for the entire area covered by LiDAR data by using the relationship we identified between the plot AGB and the LiDAR canopy height. Using Raster Calculator (Raster > Raster Calculator) apply the best fit equation you derived in step 6:  $47.151 * \exp(0.0994 * \text{MCH})$  to the 40 m resolution LiDAR MCH data to develop AGB map from the Lidar image as shown below. Save the result as LiDAR\_agb\_40m.tif

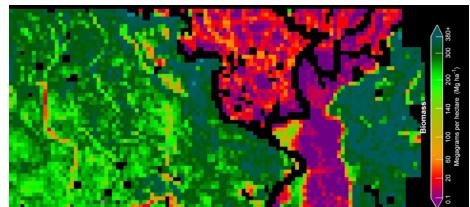


a. Note: MCH is the resampled lidar map created in step 6 (chm\_40m).

b. Note: The Raster Calculator tool in QGIS does not have an "exp" function; therefore, you can replace the equation with:  $47.151 * (2.718282 ^ (0.0994 * \text{MCH}))$ .

**Step 9:** The output of step 8 will be a AGB map that covers the same area as your LiDAR data. Now we want to resample the LiDAR AGB map from 40 m to 100 m (1-ha) spatial resolution (There are many ways to do this in QGIS, but one is to go to Processing > Toolbox > SAGA > Raster tools > Resampling, select LiDAR\_agb\_40m as your Grid, leave upscaling and downscaling method as Nearest Neighbor, make the cellsize 100, and run the tool). Save the output as LiDAR\_agb\_100m.

**Step 10:** Display the final image and apply a color scale as part of QGIS color ranges from dark red to green from low to high biomass (below).



**Step 11:** Refer to Chapter 5 for forming the LiDAR biomass models and the uncertainty depending on the plot size and LiDAR pixel size. In this exercise, a simple method was used to develop the model. The sources of uncertainty and the quantification and propagation of errors are discussed in more detail in the chapter.

## 4 ALOS BIOMASS MAPPING (NEPAL)

In this exercise, the LiDAR estimated biomass map will be used to train the SAR ALOS image to develop AGB map from the ALOS data for the larger study area.

### 4.1 Radar Processing

**Step 1:** In this exercise, we use the LiDAR derived biomass map (lidar\_agb\_100m) as the reference data to develop a model for radar estimation of biomass. The inventory plot data are small and are not suitable to extract data from ALOS PALSAR data. Therefore, we use the 1-ha resolution LiDAR based AGB map for both developing model and testing the results.

**Step 2:** For this exercise, the ALOS 2/PALSAR 2 annual mosaic for 2015 is provided in the folder Data.zip folder (ALOS). Data are in HH and HV polarizations. For practice, you may want to download the ALOS PALSAR data for 2015, 2016, 2017 directly from the JAXA website. See steps for downloading in the Chapter 6 Training Appendix.

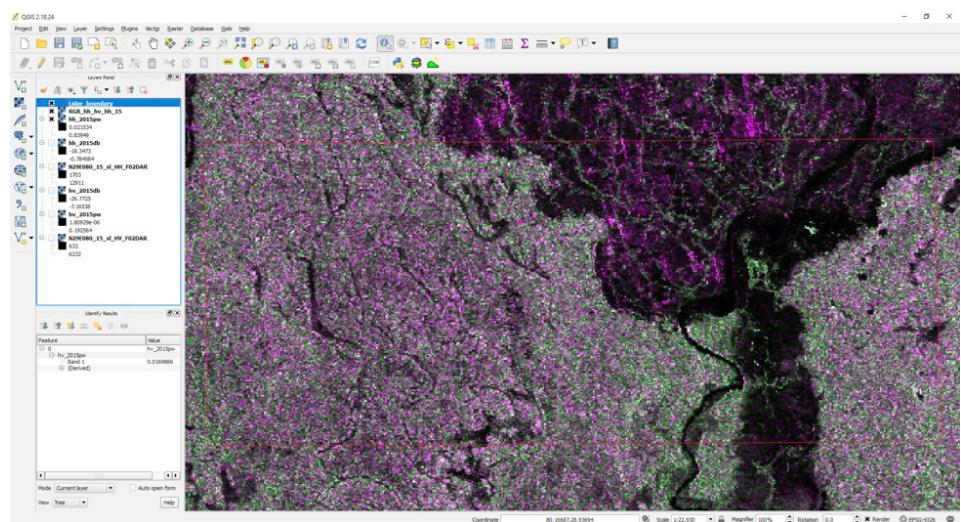
a. Note: If you download the imagery directly from JAXA, the website provides data in a grid of 1-degree tiles; you will need to select N29E080 for this site.

b. Note: If you compare the images for 2015 – 2017, you will notice some temporal variability in the backscatter due to variations in environmental conditions such as soil moisture or phenology. In this exercise, we will use data for 2015.

**Step 3:** Open the N29E080\_15\_sl\_HV\_F02DAR and the N29E080\_15\_sl\_HH\_F02DAR files (Data > ALOS) in QGIS. Note that these ALOS RTC annual mosaics are ready to use at source. Radiometric terrain correction and precise geometric corrections have already been performed.

**Step 4:** The backscatter data in the ALOS RTC annual mosaic comes as a digital number (DN) and needs to be converted to gamma naught dB for analysis. To convert the DN to dB values apply the following equation using Raster Calculator (Raster > Raster Calculator):

$$\text{Gamma\_dB} = 10 * \log_{10} [(\text{DN})^2] - 83.0$$



**Figure 1.6** SAR RGB image (HH, HV, HH) derived from ALOS PALSAR data.

Next, convert the dB values to power backscatter by applying the following equation using Raster Calculator (Raster > Raster Calculator):

$$\text{Gamma\_pw} = 10^{(0.1 * \text{Gamma\_dB})}$$

You will need to do this step for both the HV (N29E080\_15\_sl\_HV\_F02DAR) and HH (N29E080\_15\_sl\_HH\_F02DAR) images.

Save the Gamma\_pw result as gamma\_pw\_HV (or gamma\_pw\_HH for the HH polarized data). At the end of this step, you should have created two new files, gamma\_pw\_HV and gamma\_pw\_HH.

**Step 5:** Now we will create an RGB image for visualizing the backscatter power in color. You may consider a three-band composite, where R: HH (gamma\_pw\_HH), G: HV (gamma\_pw\_HV), B: HH (gamma\_pw\_HH). You could also use a ratio of HV/HH as the blue band (Calculate the HV/HH ratio using Raster Calculator). To create a multiband image, go to Raster > Miscellaneous > Merge > Edit > Type. Another option is to copy and paste following gdal command into the Edit box. Note that your data folders may be different:

```
gdal_merge.bat -ul_lr 80.0 29.0 81.0 28.0
-separate -of GTiff -o D:/Data/Results/RGB_hh_
hv_hh_15.tif D:\Data\Results\hh_2015pw.tif D:\
Data\Results\hv_2015pw.tif D:\Data\Results\
hh_2015pw.tif
```

Since the HV polarization is most sensitive to forest structure, areas that have high backscatter in HV (showing up as green in **Figure 1.6**) are likely to have higher AGB values as well.

### 4.2 Radar Biomass Model

Next, we produce samples from Lidar data to compare with radar measurements and develop a best-fit model. Refer to Chapter 5 for more detail on how to choose the appropriate LiDAR samples and issues related to the differences in date and the changes that occur between radar and LiDAR data. Any changes of landscape can easily introduce large discrepancies between LiDAR derived biomass and radar backscatter measurements.

**Step 1:** Open the resampled LiDAR biomass map at 1-ha (lidar\_agb\_100m). We can create a random or systematic sample dataset. To facilitate a systematic sample, we created a shapefile with horizontal and vertical polygons (lidar\_100m\_polygons.shp) which we used to create a systematic sample of points (lidar\_systematic\_sampling.shp). Open “lidar\_100m\_polygons.shp” (Data > Lidar) in QGIS.

**Step 2:** Use the lidar\_systematic\_sampling points to extract all of the 1-ha values from the LiDAR biomass map(lidar\_agb\_100m) (Processing > Toolbox > SAGA > Vector > Raster > Add Raster values to points) and save

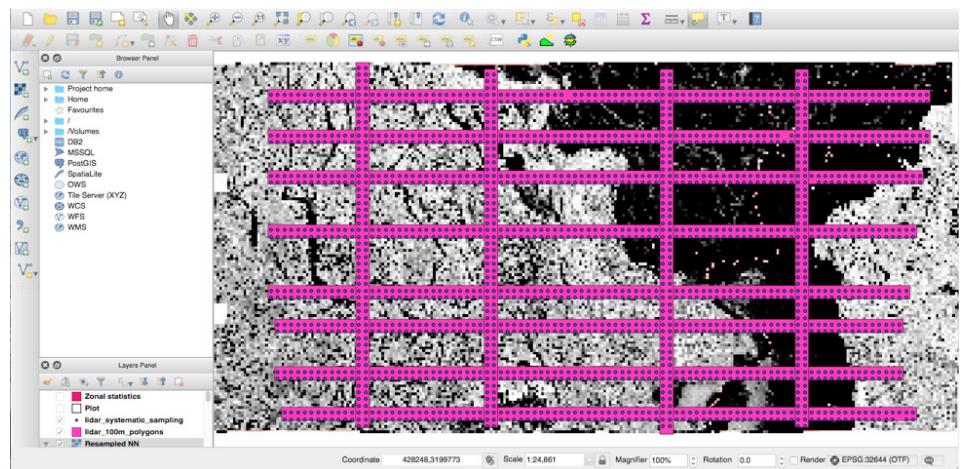
the data as a shapefile with the file name lss\_100m\_agb. This sampling strategy produces 1822 sample points that we will use to develop the model and quantify the uncertainty.

**Step 3:** Since the LiDAR based biomass data represent plot size of 1-ha, now we are going to resample the power backscatter radar data into 1-ha (100m) pixel size to match. In QGIS, resample the gamma\_pw\_HV and gamma\_pw\_HH files to 100m (Processing > Toolbox > SAGA > Raster tools > Resampling). Save these files as gamma\_pw\_HV\_100m and gamma\_pw\_HH\_100m.

**Step 4:** Next, we extract the resampled HH and HV power backscatters (gamma\_pw\_HV\_100m and gamma\_pw\_HH\_100m) from the radar images to the sampling points (lss\_100m\_agb) (Processing > Toolbox > SAGA > Vector > Raster > Add Raster values to points). Save this result as lss\_100m\_agb\_sar. When you save the shapefile, it will also create a .dbf file, which easily can be opened in Excel.

**Step 5:** In Excel, open the lss\_100m\_agb\_sar.dbf file. Before doing any analysis, we need to clean our data by removing all rows that have lidar AGB values that are negative or equal to zero (likely representing water pixels or erroneous data) from the spreadsheet. We also need to eliminate any rows where the lidar AGB value has missing data or no data (NAN, -9999, or 9999). After all cleaning (removing rows where AGB is zero or bad points), you should have a spreadsheet with 1649 data points.

**Step 6:** Now create two scatterplots, one that shows HV vs. AGB and a second that shows HV vs. AGB. You can use the same methods described in Section 3, step 5. Note that you should see a large spread of values, in part due to differences between the lidar and ALOS PALSAR acquisition times. Since the data collection did not occur on the same date, there may be some land cover or soil moisture change that could cause error in your model. Additional issues could include georeferencing discrepancies, topographical effects, errors due to speckle, and potential incidence angle variations. Be sure to consider these limitations as you work to improve your model and interpret your results.



**Figure 1.7** Horizontal and vertical polygons (pink) used to create the systematic sample points (dark blue) as inputs to generate the AGB model are displayed.

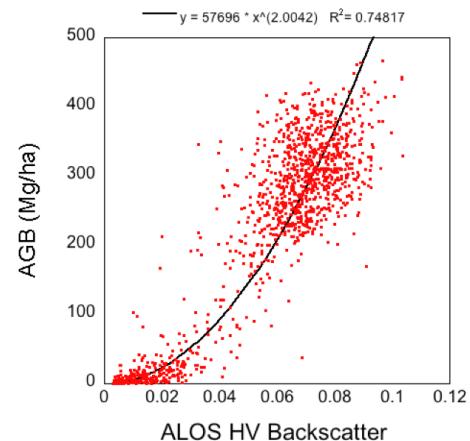
**Step 7:** Fit a logarithmic or a power-law to the both HH and HV SAR data to see the strength or weakness of the relationship between radar backscatter and lidar derived biomass (See section 3, step 6).

### 4.3 Radar Biomass Mapping

**Step 1:** In the previous section, we looked at the relationship between AGB and HH and HV backscatter. In this section, we focus on the relationship between AGB and HV backscatter only, as HV polarization has the strongest sensitivity to biomass. However, other radar polarization measurements and model fits are discussed in the text of Chapter 5.

**Step 2:** In section 4.2, step 7, we generated a best-fit model based on a power-law:  $AGB = 57696 * (HV^{2.0042})$ . We will use this equation to model biomass from backscatter. Using Raster Calculator (Raster > Raster Calculator), apply this equation to the HV backscatter image at 100 m spatial resolution (gamma\_pw\_HV\_100m). Save this file as HV\_biomass\_100m. Note that the equation is developed from 1-ha (100m) LiDAR derived map and should only be applied at the same resolution radar image. One cannot apply this equation to any other resolution (smaller or larger) radar image without introducing additional uncertainty.

**Step 3:** Evaluate the saturation in this model. Although the fit shows no saturation, however, the data shows



**Figure 1.8** An example of a best fit model between AGB and HV backscatter.

that HV backscatter has almost no sensitivity to biomass above 200 Mg/ha for these forests (See Saatchi et al. 2011, or Saatchi et al., 2007 for other alternative equations and saturation of the radar data).

**Step 4:** Display the map of forest biomass from the 2015 ALOS PALSAR2 image (HV\_biomass\_100m) using a color range to show the variation of biomass across the image. **Step 5:** Read values of the biomass from the image and visually compare it to the reference biomass map derived from the lidar image (lidar\_agb\_100m).

**Step 6:** Mask out all pixels above 200 Mg/ha to show that the map has large uncertainty over areas of above 200 Mg/ha and cannot be trusted (Raster > Raster

Calculator > "HB\_agb\_100m >=200" to create mask). Although you can leave the map untouched by explaining the fact that the map has large uncertainty in areas where AGB is > 200 Mg/ha.

**Step 7:** Clip the radar AGB map (HV\_agb\_100m) to the same extent as the LiDAR AGB map (lidar\_agb\_100m) using Raster > Extraction > Clipper in QGIS. Save this file as HV\_agb\_100m\_clp.

**Step 8:** Here we are going to calculate the percent difference between the clipped radar-derived AGB map (HB\_agb\_100m\_clp) and the lidar AGB map (lidar\_agb\_100m) using the raster calculator (Raster > Raster Calculator). Note that the lidar and backscatter AGB maps should be the same size for this step. Use the following equation in Raster Calculator to calculate the percent difference in AGB:

$$\text{Diff} = 100 * (b1 - b2) / b1$$

Where b1 is the LiDAR map (lidar\_agb\_100m) and b2 is the radar map (HV\_agb\_100m\_clp).

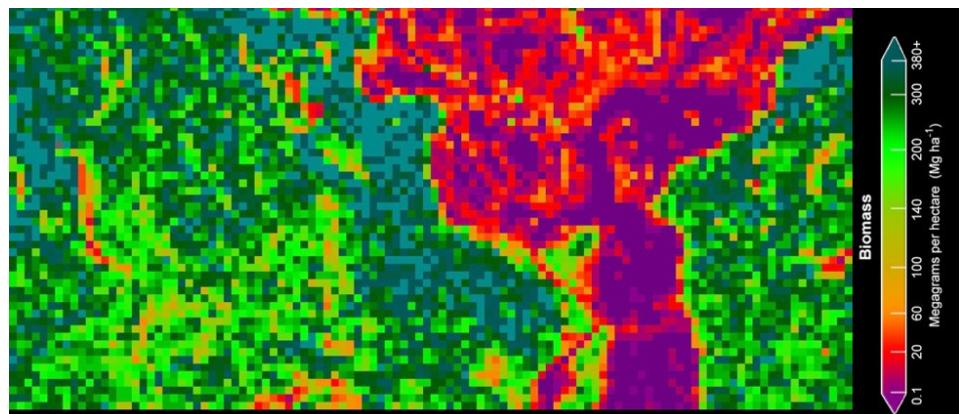
**Step 9:** Display the difference map in percentage and provide a color range to show the range of values and include the color range on the side for presentation of the results (**Figure 1.10**).

#### 4.4 Improving the AGB Map

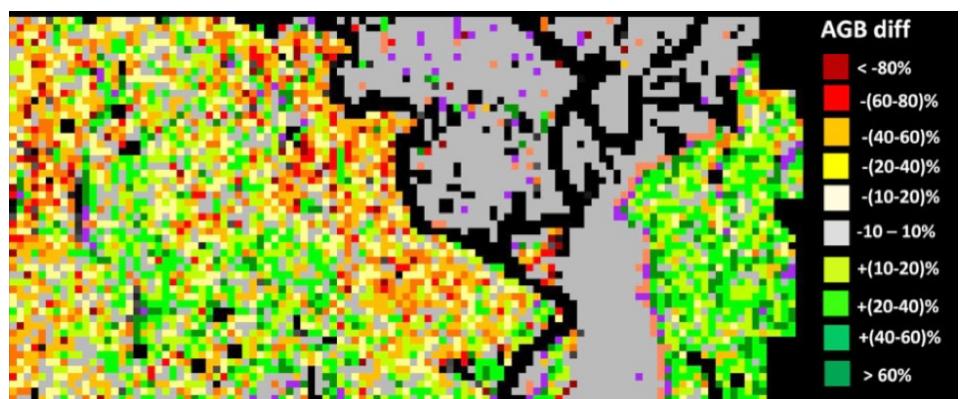
Here, we improve the radar biomass model and AGB mapping by using multi-temporal radar imagery. In an ideal scenario, ALOS PALSAR data from different seasons and over time from the same or multiple years can be downloaded and used to reduce the effects of soil moisture and phenology and improve biomass mapping.

**Step 1:** Download the ALOS PALSAR mosaic data for 2015, 2016, and 2017 from JAXA website (You can use the data included in Data > ALOS or see the Chapter 6 training appendix for information on downloading ALOS PALSAR data from the JAXA website).

**Step 2:** Using the steps provided previously, for each year, calculate the gamma power of the HV backscatter (Section 4.1, step 4) and resample each image to 100m (Section 4.2, step 3). At the end of this step, you should



**Figure 1.9** Example results showing variation in biomass derived from HV backscatter.



**Figure 1.10** An example of an AGB percent difference map between lidar-derived AGB and backscatter (HV)-derived AGB.

have three HV backscatter images resampled to 100m: one for each year.

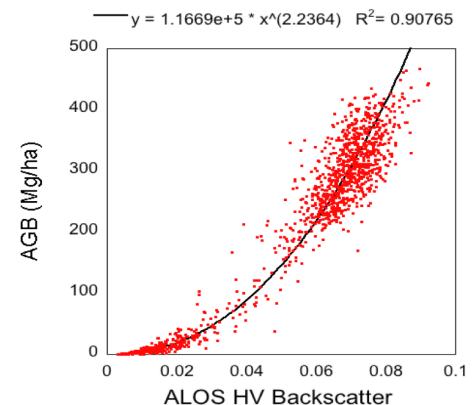
**Step 3:** Use the lidar systematic sample points (lss\_100m\_agb) to extract the HV backscatter values for each year (See Section 4.2, step 4). Save the output as lss\_100m\_agb\_sar15\_17.

**Step 4:** Open the lss\_100m\_agb\_sar15\_17.dbf in Excel. Remember to clean the data as described in Section 4.2 step #5). Next, create a new column where you average the backscatter values from 2015, 2016, and 2017 to create a mean backscatter value in the spreadsheet.

**Step 5:** Create a scatterplot with the lidar-derived AGB and the three year HV mean in the spreadsheet. Develop a new model using the power-law function for simplicity.

**Step 6:** Before we apply the model we created in step 5, first we need to average the three backscatter imag-

es to create one single image of HV. Remember to use your gamma power images that have been resampled to 100m. You can use Raster Calculator to average the three images in QGIS.



**Figure 1.11** An example of a best-fit model between AGB and a 3-year average of HV backscatter.

**Step 7:** Based on your results from step 5, apply the best-fit equation  $AGB = 116690 * (HV^{2.2364})$  to the averaged HV image using Raster Calculator. Display the results with an appropriate color scale.

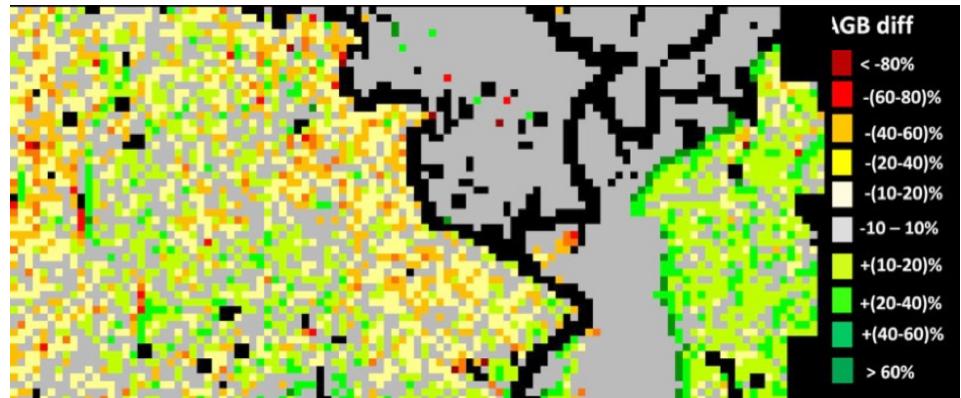
**Step 8:** Develop a percent difference map between LiDAR AGB and the new HV AGB map and color the range of biomass difference in percentage and display it (See Section 4.3 steps 7 and 8).

**Step 9:** Compare the new percent difference map with the earlier version derived just from the 2015 backscatter. Where do you notice differences in the overall negative and positive percent differences?

#### 4.5 Evaluating Uncertainty in the AGB Map

By assuming that we have several sources of errors that introduce uncertainty in the pixel level estimation of biomass, we can calculate the total uncertainty associated with estimating AGB at the pixel level by assum-

$$\varepsilon_{AGB} = \sqrt{\varepsilon_{measure}^2 + \varepsilon_{model}^2 + \varepsilon_{sampling}^2 + \varepsilon_{prediction}^2}$$



**Figure 1.12** Example of an AGB percent difference map between lidar AGB and a three-year average of HV

ing all errors were independent and random, by using: where each of the terms are the relative errors at that pixel scale. Detailed description of error analysis and uncertainty assessment of the map are given in the Chapter 5. Here we examine three steps for evaluating the uncertainty of the map.

**Step 1:** For pixel level prediction, use model fit parameter uncertainty to simulate several biomass maps by bootstrapping the coefficients using the range of pa-

rameter uncertainty.

**Step 2:** Generate several maps (100 if the image is small as in the Nepal case) or about 20-30 if the image is large.

**Step 3:** Calculate the mean and variance of the bootstrapping approach and show the variance as a new map.



**DR. MARC SIMARD** is a Senior Scientist with NASA-Jet Propulsion Laboratory's Radar Science & Engineering Section. He obtained a PhD in Geomatics in 1998 at the Université Laval, Québec. Simard is also a Project Scientist within the Joint Institute for Regional Earth System Science at the University of California in Los Angeles (UCLA). He is also an Adjunct Faculty member at the Department of Oceanography and Coastal Sciences of Louisiana State University (LSU). He is an active member of the International Blue Carbon Working Group (the Blue Carbon Initiative is sponsored by Conservation International, UNEP and IUCN) and JAXA's Global Mangrove Watch program.

Since 2004, Dr. Simard has served as the Principal or Co- Investigator of several NASA funded projects in the Carbon Cycle and Ecosystems (CC&E) program. His research has a strong multi-disciplinary dimension with a focus on coastal wetlands regions. He is currently, among other NASA projects, a co-investigator in a NASA Carbon Monitoring System (CMS) for mangrove forests entitled: "Estimating Total Ecosystem Carbon in Blue Carbon and Tropical Peatland Ecosystems".

Marc Simard was the Principal Investigator for the JPL funded 2016 multi-aircraft campaign in the Mississippi River Delta, a precursor to Delta-X. Finally, and highly relevant, Simard is a member of the Science Teams of two upcoming NASA spaceborne missions: NISAR (NASA-ISRO Synthetic Aperture Radar) and SWOT (Surface Water and Ocean Topography) both planned for launch in 2021. Dr. Simard expects that, in addition to providing deep understanding of accretion processes, the Delta-X data will be used for the NISAR and SWOT cal/val, and enhance these NASA missions.

# CHAPTER 6

## Radar Remote Sensing of Mangrove Forests

Marc Simard, Senior Scientist, Radar Science and Engineering Section, Jet Propulsion Laboratory, California Institute of Technology

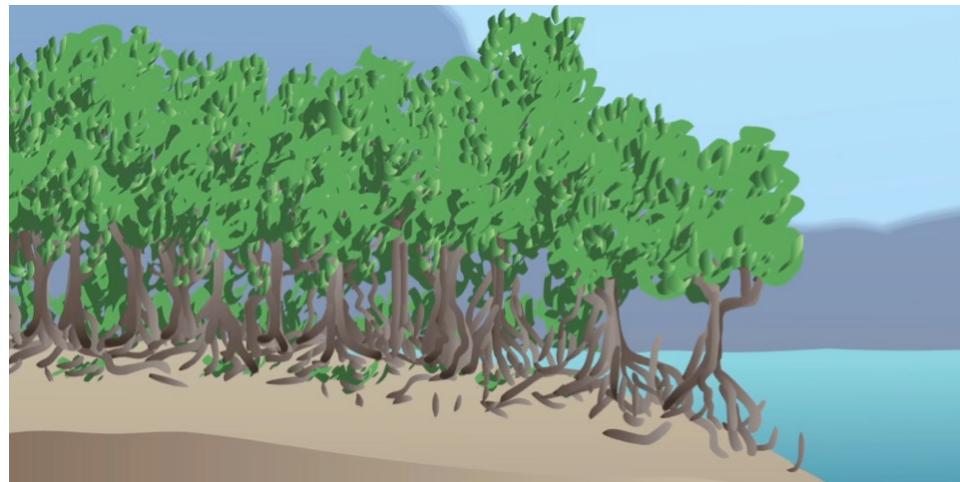
### ABSTRACT

Mangrove forests are highly productive ecosystems providing critical ecosystem services. It is estimated that about a third of mangrove forests have been lost during the last century and are still being decimated at a rate of about 0.4% per year. Mangrove forests are thin strips of forests living along tropical coasts. These environments are generally cloudy; optical remote sensing instruments provide limited temporal coverage to ensure consistent monitoring of mangrove health and status. Radar remote sensing enables all-weather monitoring of mangrove forest gain and loss in extent. In addition, it can observe several parameters related to vertical canopy structure and biomass. The first part of this chapter introduces mangrove forests and the state-of-the-art radar remote sensing techniques to measure and monitor mangrove forest structure. The second part of this chapter presents a step-by-step tutorial on the use of radar remote sensing to make these measurements. After this chapter, the reader will be able to perform analysis of radar images of mangrove forests and more.

### 6.1 Introduction to Mangrove Forests

Mangrove forests are some of the most productive ecosystems in the world. They thrive within the intertidal zone along the coasts of tropical and subtropical regions (**Fig. 6.1**). Mangrove trees can sustain salt water and soils with low oxygen availability through root adaptations. They were recently included in the Intergovernmental Panel on Climate Change (IPCC) climate mitigation strategy through the Wetland supplements. While mangrove forests cover a small land area (<1%), they may be responsible for 10% of global carbon export to oceans (Jennerjahn and V. Ittekot 2002). Most importantly, mangrove forests provide numerous ecosystem services that sustain the livelihood of millions of people (Barbier et al. 2011). Some of these services, in addition to carbon sequestration, include protection of coastline and infrastructure against severe storms and tsunamis, nursery of fish and crustaceans, and the production of lumber and charcoal.

Mangrove forests occupy terrestrial and marine environments, enabling them to support a very broad range of biodiversity. This biodiversity ranges from organisms within the soil to high numbers of fish spe-



**Figure 6.1** Mangrove forests occupy the intertidal zone between approximately mean water to high tide. Assuming a flat ground at sea level is generally a good approximation (Image credit: Maria Raykova).

cies to terrestrial species including reptiles, mammals, birds, and insects. Mangroves provide both direct and indirect services to local populations that inhabit the coastal zone. Several products can be directly sourced from the mangrove that have both a subsistence use and economic value. Primary benefits include sediment trapping; the production of nutrients and organic matter through detritus; a sink for carbon, nitrogen, and phosphorus; maintaining water quality; provision of food and habitat for biodiversity; and

providing shoreline protection from storms and rising sea levels (e.g., Quoc Tuan et al. 2012). Mangroves have been demonstrated to provide greater long-term economic benefits to local households through their preservation, compared to the short-term economic gains through their destruction for products (McNally et al. 2011). An important reason for conversion of mangroves is shrimp farming, driven by economic incentives as mangrove is often regarded as wasteland (Primavera 2000). The level of services often depends

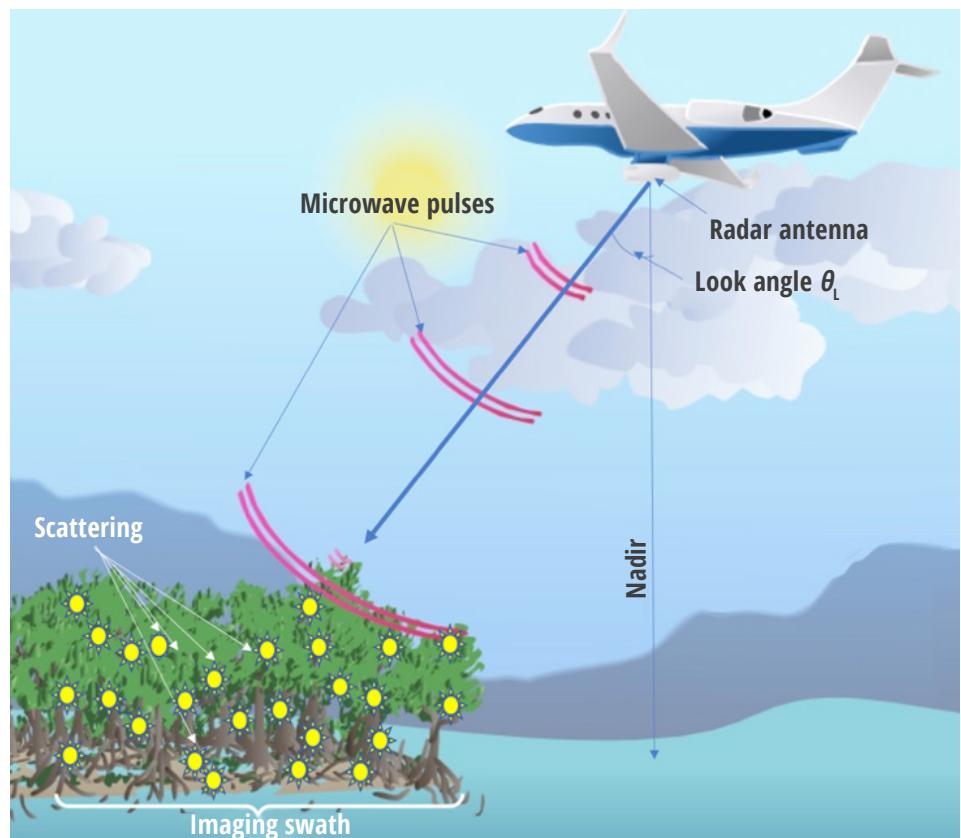
on forest structure and, therefore, structural parameters such as height and basal area. For example, tall forests provide more protection against strong winds, store more carbon, and can provide more lumber and coal. Mangrove structure should therefore be considered when mangrove economical value is assessed.

Mangroves can exceed 60m in height and are subsequently able to attain high values of above ground biomass (AGB) around the  $800 \text{ Mg ha}^{-1}$  mark (Simard et al. 2019). These values rival observations in other types of tropical forests. Their elevated Net Primary Productivity (NPP) contribute large amounts of organic carbon inputs into the underlying soil. Thanks to slow anaerobic decomposition, mangrove forests store disproportionate amounts of carbon in their soils. Mangrove ecosystems are estimated to be amongst the most carbon rich ecosystems within the tropics, storing an average of  $1,023 \text{ Mg C ha}^{-1}$  in shallow soil depth ranging from 0.5–3 m (Donato et al. 2011). At current mangrove loss rates, between 0.02–0.12 Pg per year is released to the atmosphere, that is 10% of total carbon emissions from deforestation despite accounting for less than 1% of tropical forest area (Donato et al. 2011). Hamilton et al. (2016) estimate current global loss between 0.16 and 0.39%. Thus, they play a significant role in the carbon cycle and consequently are becoming economically viable to protect (Murray 2012, Jerath 2012, Pendleton et al. 2012). Initiatives, such as the Reducing Emissions from Deforestation and forest Degradation plus (REDD+) are carbon accrediting programs, whereby the carbon stored within natural ecosystems is valued by its ability to offset anthropogenically produced  $\text{CO}_2$ . This payment for ecosystem services (PES) initiative extends beyond the worth of carbon to other services such as their resilience to hazards and role in maintaining fish biodiversity and water quality (Locatelli et al. 2014). The inclusion of mangroves in such initiatives can make them profitable environments, enhancing the socioeconomic benefits of mangrove forests beyond subsistence use to an asset for all global citizens.

There are several environmental drivers of mangrove structure, mainly, precipitation and temperature (Simard et al. 2019) and the availability of nutrients and salinity (Castañeda-Moya et al. 2013). Local geophysical characteristics such as microtopography



**Figure 6.2** Global distribution of mangrove forests (From Giri et al. 2010).



**Figure 6.3** Radar imaging, whether airborne or spaceborne, is performed by transmitting a microwave pulse ‘sideways’ that scatters with land features such as forest branches, trunks roots, and ground. The imaging swath illuminated by one pulse determines the image size in the cross-track (or range) direction. (Background image credit: Maria Raykova) Additional information on how SAR images the world can be found in [Chapter 2](#).

and freshwater availability strongly control salinity and hydroperiod, and thus mangrove NPP (Castañeda-Moya et al. 2013). However, microtopography is very difficult to estimate at the landscape scale and cannot be observed directly from spaceborne remote sensing. On the other hand, remote sensing can measure the ecosystem response to these geophysical variables, reflected in observed canopy height, spe-

cies distribution, and spatial patterns.

Despite their importance for carbon storage, biodiversity, and supporting indigenous local communities, mangrove forests are threatened across their entire range (Thomas et al. 2017). Mangrove areal extent for the nominal year 2000 was 13.7 million ha, far below the previous 1980 estimate of 18.8 million ha (FAO 2007) (**Fig. 6.3**). The rate of mangrove

loss throughout the 1990s was estimated as  $1\text{ yr}^{-1}$ ; a rate twice that of terrestrial rainforest over the same period (Mayaux et al. 2005). Recently, Hamilton and Casey (2016) found the yearly loss of mangrove extent between 2000–2012 varying between 0.16 and 0.39%. Comparatively, 30% of total tropical terrestrial forest has been lost as a consequence of anthropogenic activity since monitoring began, while it is estimated that one third of total mangrove forest has been lost over the last half-century alone (Alongi 2002). Unfortunately, the loss of mangroves across the globe is deemed critical enough that 11 of the true 70 mangrove species have met the criteria of the Red List categories of threat (Polidoro et al. 2010).

The largest driver of this loss has been the conversion of mangrove forests to aquaculture. In an assessment of mangrove forest extent and loss at a variety of locations in the Americas, Africa, Asia, and Australia, the greatest cause of mangrove loss was evaluated to be due to mariculture practices. 52% of  $36 \times 10^3 \text{ km}^2$  of the estimated loss from within countries containing 66% of the total area of mangrove forests was caused by shrimp cultivation (Valiela et al. 2001). Aquaculture is the fastest growing animal-food sector in the world. In 2011, fish from aquaculture practices accounted for nearly half of the total fish consumed worldwide (45.6%). Aquaculture has a plethora of direct and indirect detrimental impacts upon a mangrove forest. These include the immediate loss of mangroves for pond construction alongside the alteration of natural tidal flows, release of toxic wastes, reduced water quality and alterations to sedimentation rates, and turbidity. Additional pressures upon mangroves include the development of the coastal zone that causes the direct replacement of mangrove and a suite of associated environmental problems, such as pollutants in runoff. Furthermore, the coastal zones of the world are becoming increasingly populated, and the current trends of increasing global population will put further demands upon mangrove forests (FAO 2013, FAO 2012).

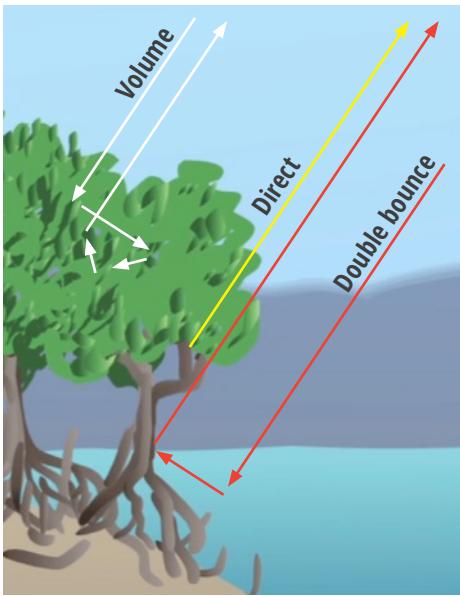
As we move increasingly through an era of unprecedented climate change, Earth's climate will undergo changes that are currently not known with certainty. Sea level rise is expected to continue and accelerate over the coming century, with an increase in mean

sea level by as much as 1 m by 2100 (CITATION). Although sea level rise will not be uniform throughout the oceans, 70% of the world's coastlines are estimated to experience sea level rise within 20% of the global mean (IPCC AR5). Mangroves are known to accrete sediment (Cheong et al. 2013) by trapping sediment suspended when inundated or through the build-up of peat through the decomposition of organic matter (Krauss et al. 2014). The survival of mangrove forests in the face of sea level rise is therefore dependent on whether sediment is accreted at the same rate as sea level rise (McKee et al. 2002, Hashimoto et al. 2006). Should sea levels rise above the rate of the terrestrial surface, mangroves will either face periods of longer inundation or will migrate landwards into new areas. It is not currently known how climate change will affect the atmosphere and subsequent terrestrial processes, making the extrapolation of all the effects of climate change on mangrove forests difficult. Precipitation is expected to be spatially variable, with increasing climate change and growing contrast between wet and dry regions and between seasons (IPCC AR5). The impact of increasing precipitation is expected to have a positive effect on growth rates, biodiversity, and mangrove extent as they migrate into previously drier environments (Eslami Andargoli et al. 2009). An increase in precipitation will also decrease the salinity of the environment and is expected to lead to an increase in species richness and diversity (Asbridge et al. 2015). In contrast, a decrease in precipitation will increase the salinity of mangrove environments and cause an overall decrease in mangrove area as freshwater influxes become too saline to support growth (Duke et al. 1998, Gilman et al. 2008). Decreases in precipitation are also likely to cause a reduction in photosynthesis as a consequence of increased aridity (Arreola Lizarraga et al. 2004). Atmospheric temperatures will increase by as much as  $2^\circ\text{C}$  by the end of 2100 and are forecast to increase further thereafter (IPCC AR5). Changes in atmospheric temperature can be expected to cause an expansion of mangroves into higher latitudes and change the species composition and distribution of mangrove forests (Soares et al. 2012, Wilson and Saintilan 2012, Saintilan et al. 2014). The effects of climate change on mangrove forests are difficult to

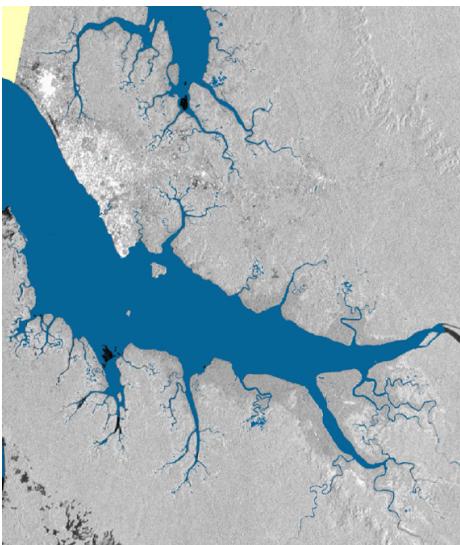
accurately predict due to the complexity of the natural system and complex feedbacks.

Traditionally, large-scale mangrove mapping was limited to sketch maps, fieldwork maps, and the digitizing of digital datasets (Spalding et al. 1997). Yet over the past years, the number of studies on mangrove extent, change, ecosystem structure, ecosystem services, and vulnerability derived from remote sensing have proliferated (Kuenzer et al. 2011). The first global map exclusive to mangrove forests that used remotely sensed data alone was that of Giri et al. (2011). This work processed over 1,000 Landsat scenes gathered over the period 1997–2000, and estimated the total mangrove extent to be 13,776,000 ha despite the methodology suffering from a number of limitations. Since then, changes have occurred and been detected at global scales (Lucas et al. 2014, Thomas et al. 2017, and in maps by Hansen et al 2013). Products that incorporated annual estimates of mangrove extent followed, such as the CGMFC-21 (Hamilton and Casey 2016, Hutchison et al. 2014), with a much higher temporal resolution. In addition, there have been many regional, country-scale, and project-scale assessments of mangrove extent, change, and three-dimensional structure using both optical and passive remotely-sensed data (Fatoyinbo and Simard 2008, Fatoyinbo et al. 2013, Spalding et al. 2010, Simard et al. 2019).

There is a plethora of remotely sensed data available for mapping mangrove extent and change, which has the potential for long-term monitoring of land-use change and the identification of proximate drivers of change. These include the Landsat time-series, now enhanced by the European Space Agency (ESA) Sentinel-2 platforms, that, when combined, offer an unprecedented quantity of data. Similarly, radar data is available from ESA via the Sentinel-1 satellites and annual mosaics from the Japan Aerospace Exploration Agency's (JAXA's) Advanced Land Observing Satellite 2 (ALOS-2) platform. Historic radar imagery to aid in time-series mapping is freely available via ALOS and Japanese Earth Resources Satellite 1 (JERS-1) mosaics (See **Table 2.5, Chapter 2**). These will soon be followed by the joint NASA-ISRO Synthetic Aperture Radar (NISAR) mission planned for launch in December 2021. It is therefore timely to develop



**Figure 6.4** Radar scattering mechanisms. In mangrove forests with aerial roots (*Rhizophora*), the microwave signal is attenuated, decreasing backscatter at high biomass.



**Figure 6.5** Radar backscatter at L-HH, acquired by ALOS-2 over the Gabon Estuary, Gabon. Bright areas to the North West result from strong double-bounce effect in urban structures of the city of Libreville. Large medium backscatter (mid-gray tones) are due to volume scattering in forests. Water was easily masked based on darker backscatter (in particular at LHV). Interestingly, due to strong attenuation from roots, tall mangrove forests with red mangrove trees reaching several tens of meters exhibit lower backscatter along the South Eastern portion of the Estuary.

new remote sensing algorithms that combine optical and radar data for long-term monitoring of mangroves forests in Southeast Asia.

## 6.2 Radar Remote Sensing

A radar instrument generates its own electromagnetic signal by transmitting a microwave pulse that enables observation of Earth's surface (or other planets and moons) day and night. In order to generate an image, the pulse is focused in a direction away from nadir (Fig. 6.3). When the pulse is transmitted at nadir, the instrument is called a radar altimeter. The microwave pulse typically illuminates ground areas of tens of kilometers, and only the portion of energy reflected toward the radar is measured. The angular reflection pattern depends on the target properties such as roughness (differs greatly between plants, water surface, urban structures) and geometry. The geometry is determined by the look angle and the terrain slope. The former is the angle subtended by the line of sight between the radar and a target on the ground. Thus, it varies greatly across an image. The look angle and terrain slope can be combined into the incidence or projection angle. These angles are often used to perform terrain radiometric corrections (sometimes called ‘terrain flattening’), which is intended to remove image artifact due to geometry. In mangroves, topographic effects are generally neglected due to their unique setting of very flat areas.

### 6.2.1 MICROWAVE BANDS

Radars are active instruments with a definite advantage over optical sensors: they can see through clouds, day and night. This is a particularly prized attribute along the tropical coastlines, and its free public availability is continuously rising. They transmit a microwave pulse and measure the portion of the energy that is reflected back. The measured return is called “backscatter” and is generally presented in decibels ( $10\log_{10}(\text{Intensity})$ ). Some Radar instruments come in several “colors” (i.e., wavelength bands): Ka and Ku-bands, X-, C-, S-, L-, and P-bands. Those are denominations introduced during the development of radar during World War II, and they simply refer to a range of frequencies as defined by

the Institute of Electrical and Electronics Engineers (IEEE). See **Table 2.3 in Chapter 2** to see common applications of SAR bands.

### 6.2.2 SCATTERING MECHANISMS IN MANGROVES

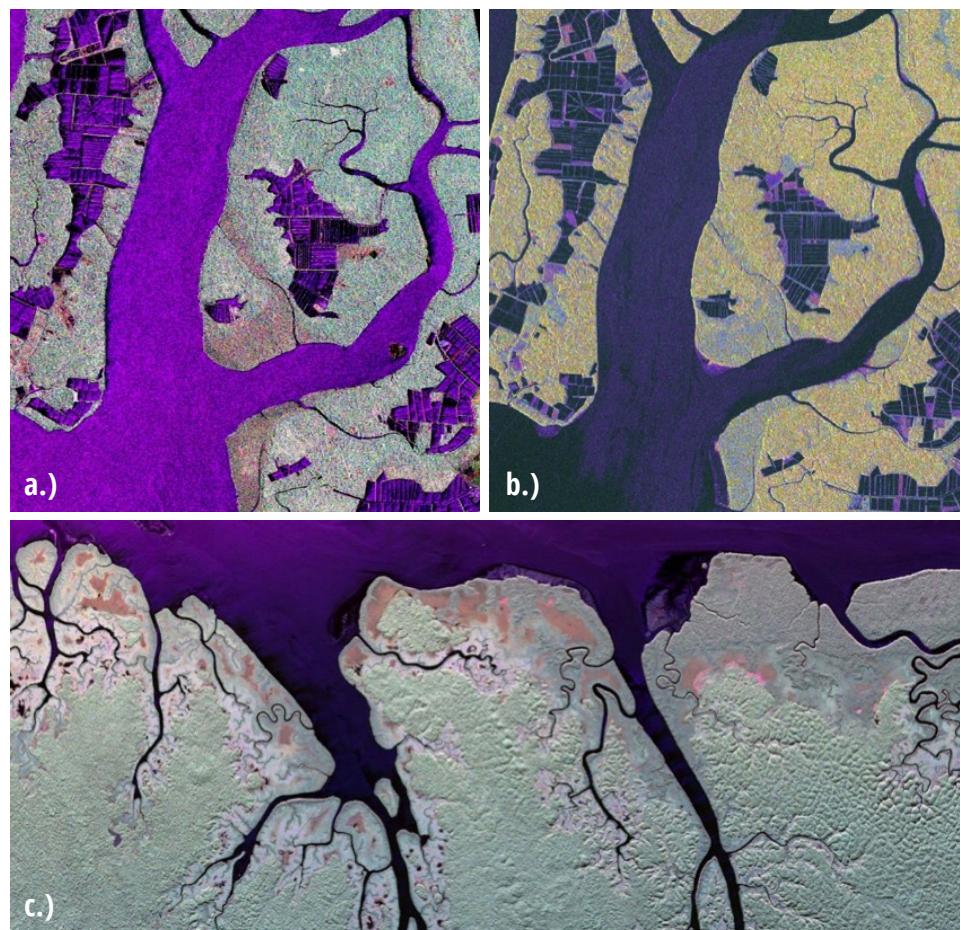
There are three types of scattering mechanisms: (1) direct (or single bounce), (2) double-bounce, and (3) volume scattering (see Fig. 6.4). In mangrove forests, the “double-bounce” term that strongly impacts the HH channel (see **Chapter 2, Table 2.3**, relative scattering strength by polarization, and the subsequent section) may be reduced by the presence of aerial roots as microwaves are scattered and attenuated (Fig. 6.5). The dominant scattering mechanism in mangrove forest strongly depends on canopy structure. Trends in volume and double-bounces’ signatures vary much more than in other types of forests. In particular, in mangroves the volume scattering decreases and double-bounce scattering increases in closed and open canopies, respectively (see following section on polarimetry). Inundation at the time of data acquisition impacts radar signals in open mangrove forests.

### 6.2.3 POLARIMETRY

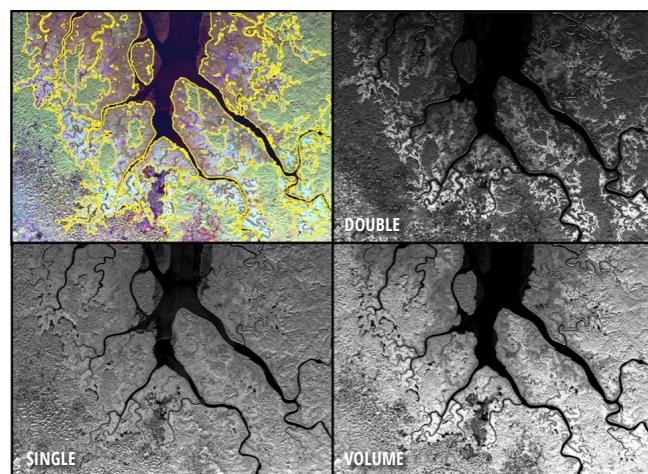
The radar measurement can also be characterized through polarimetry. Generally, radar instruments are enabled for several orthogonal polarimetric modes, transmitting horizontal (H) or vertical (V) polarization, and receiving either H or V. For example, an L-band radar transmitting a horizontally polarized microwave and receiving its vertical polarization would be identified as L-HV. A single instrument can collect data in several polarimetric mode by alternating pulses. A quad-pol L-band radar collects four channels: L-HH, L-HV, L-VH, and L-VV (e.g., Fig. 6.6). Upcoming instruments like the Radarsat constellation suite of instruments will provide circular polarization, indicating that polarization state changes in time. Each polarimetric configuration can be considered as an image band in the radar dataset, each sensing various characteristics of the forest canopy through the variety of scattering mechanisms. While the HV measurement is dominated by the volume scattering reflections, the HH and VV contain a significant ground contribution.

In particular, the HH polarization is strongly impacted by the occurrence of double-bounce scattering, which is greatly enhanced by the presence of water. As such, it is important to note the contribution from double-bounce scattering in mangrove forests can be greatly reduced by the presence of aerial roots. *More about polarimetry in Chapters 2 and 3.*

Early work on polarimetry at C-, L-, and P-band (e.g., Mougin et al. 1999, Proisy et al. 2002) has shown that the relative impact of the scattering mechanisms changes significantly with radar wavelength, with ground and double-bounce contributions increasing wavelength. Polarimetry has been demonstrated as a powerful method to classify wetland types, including mangroves at X-band (e.g., Hong et al. 2015), and mangrove species and structure at L-band (e.g., Brown et al. 2016) and C-band (e.g., Kovacs et al. 2013, Cougo et al. 2015). There are several models to obtain the relative contribution of the three scattering mechanisms. A popular one is the Freeman-Durden decomposition (available in SNAP and PolSARpro software) used in Proisy et al. (2002). **Figure 6.7** shows the decomposition of a fully polarimetric radar image acquired by the UAVSAR's L-band airborne radar instrument. In the top left image of **Figure 6.7**, the brown areas represent low double interaction found in tall *Rizophora* mangrove forests, and shades of blue are found in more open and shorter shrub mangroves. Green tones, representing dominance of volume scattering, are found in inland forests. Otherwise red tones, representing dominance of single bounce, occur over open land surfaces. The individual contribution of each scattering mechanism is shown in grey-scale images in **Figure 6.7**. Contrary to inland forests, the volume component is reduced in tall mangrove forests and increased with shorter ones. While the volume component (or even at HV polarization) may become similar to that of inland forests, the apparent texture of mangrove forests is much smoother, in part due to inland topography and to overall homogeneity of mangrove canopy structure. Thus, polarimetric signature can be used to identify mangrove forests from other landcover types, particularly at longer wavelengths (e.g., L-band), and also differentiate mangrove structural attributes and species.



**Figure 6.6** Color composite images of mangrove forests of the Guayas Estuary in Ecuador from: a) ALOS-PALSAR-1 HH, HV, and VV in RGB, respectively; b) Sentinel-1 VV, VH, and VV in RGB, respectively. Volume scattering dominates at all polarization and at both L- and C-band. Lower backscatter is observed for the younger, low-density forest (<100t/ha) found along the coast at the bottom of the images. Figure 6(c) shows a color composite (HH, HV, VV) of a 6-m resolution L-band airborne UAVSAR image acquired in the Gabon Estuary showing the distinct signature of mangrove forest whether tall (east) or short shrub forest (west). Green indicates the dominance of HV in inland forest is more significant than in mangrove forest where all polarization configurations (HH, HV, and VV) behave similarly, resulting in the observed grey level intensities.



**Figure 6.7** Polarimetric representation of mangrove forests in Akanda, Gabon. Top left image shows an RGB color composite image of single, volume, and double-scattering components based on the Freeman-Durden decomposition. The yellow polygons show the mangrove extent.

## 6.2.4 INTERFEROMETRY

Radar can operate in interferometric mode, which means the measurement is obtained from several individual observations of the same target, viewed from different radar antenna positions. These individual observations can be collected simultaneously with pairs of radar antennas, or with a single antenna operating in repeat-pass mode. When observations are from a slightly different point of view, as in the pairs of radar antennas formation, the elevation can be estimated through geometry. The latter technique involves repeating an observation using the same instruments and combining measurements through interferometry (complex multiplication) to obtain coherence (a measure of the similarity of two images) and phase (relative location of scattering phase center between the two images). In addition, a pair of repeat-pass observations obtained from the same location (same instrument from same location but different time) enables measurements of changes in elevation or displacement of targets such as glacial flows, landslides, ground subsidence, etc. In the case of forest, zero-baseline repeat-pass interferometry is strongly impacted by ‘temporal decorrelation,’ which is a result of the motion of branches, changes in moisture, or growth. The zero-baseline interferometric measurement has been used to classify forest-age and structure of other types of forests (Simard et al. 2012, Pinto et al. 2012) as it strongly depends on forest height (Lavalle et al. 2012). Temporal decorrelation can impact repeat-pass non-zero baseline observation and must be compensated to obtain canopy height (Denbina et al. 2018).

Single-pass, non-zero baseline radar interferometry has been used to map mangrove canopy height (Simard et al. 2006, Simard et al. 2008, Fatoyinbo et al. 2013, Lagomasino et al. 2015, Lee et al. 2015, Simard et al. 2019). These maps express mangrove forest height using single-pass interferometric data obtained by the Shuttle Radar Topography Mission (SRTM) and TanDEM-X data (e.g., Fig. 6.7). The mangrove canopy height can then be translated into above ground biomass through in situ allometry relating biomass to canopy height (see section on mapping mangrove forest structure). On the other hand, repeat-pass interferometry can be used to identify structural attribute such as canopy stature and closure (e.g., Fig. 6.9).

RADAR BAND	SHRUB MANGROVES	TALL MANGROVES
P-HH	Around -17dB	Around -8dB (may increase with AGB)
P-HV and P-VH	Around -22dB	Around -14dB (may increase with AGB)
P-VV	Around -10dB	Around -7dB (may increase with AGB)
L-HH	-25dB to -15dB	Reduces from -5dB to -18dB with AGB
L-HV and L-VH	-25 to -20dB	Reduces from -15 dB to -22dB with AGB
L-VV	-20dB to -12 dB	Reduces from -8 to -16dB with AGB
C-HH	About -12dB	Varies about -7dB (no relationship to AGB)
C-HV	-20 to -15dB	Varies about -14dB (no relationship to AGB)
C-VV	About -12dB	Varies about -6dB (no relationship to AGB)

**Table 6.1** General trends in radar backscatter for mangroves (Lucas et al. 2007, Proisy et al. 2002, Mougin et al. 1999). The trends are not constant and may exhibit increase for low biomass stands (shrubs) up to the standing forest where backscatter may decrease. Thus, the relationship between backscatter and biomass or height are not reliable biomass estimators.

BAND	MANGROVE FOREST PENETRATION DEPTH	TYPE OF SCATTERING CAUSED BY MANGROVE FORESTS
K	Unknown; most likely a few tens of centimeters.	Single direct bounce and volume from top of canopy.
X	Interferometric measurement indicate penetration reaches, in the mean sense, Lorey's height (~1/3 of top forest height).	Single direct bounce and volume from top of canopy, with a small surface and double bounce component. The latter increase dramatically in open forests and at low biomass.
C	Comparison of SRTM C- and X-band show it is similar to C-band. Down to the equivalent of Lorey's height (~1/3 of top forest height)	Single direct bounce and volume from upper canopy, with a small surface and double bounce component. The latter increase dramatically in open forests and at low biomass.
L	Microwave penetration into canopy is as large as half the canopy height.	Single direct bounce dominates in tall forests, with volume dominating with shorter shrub mangroves. The contribution of double bounce increases significantly at low biomass and in open forests. In large red mangrove forest, with large aerial roots, microwaves will get absorbed and volume dominates again, although diminished.
P	Similar to L-band, where microwave penetration into canopy is as large as half the canopy height.	Single direct bounce dominates in tall forests, but the contribution of double bounces increases significantly at low biomass. In large red mangrove forest, with large aerial roots, microwaves will get absorbed and volume dominates again at biomass slightly larger than L-band.

**Table 6.2** Microwave penetration depth and dominant scattering mechanisms in mangrove forests. Note this is in the mean sense as microwaves interacts with the entire canopy, all the way to the ground.

## 6.2.5 RADAR DATA FORMAT

The availability of freely available radar data has increased significantly in the last decade (refer to Table 2.5, Chapter 2). Some datasets are calibrated and georeferenced science-ready products, and others are distributed at various levels of processing. There are multiple approaches to prepare the radar data, which depends on the original format of the data that has been downloaded. Generally, data is distributed with processing levels 1.0, 1.1, 1.5, or 2.0. The 1.0 and 1.1 format refer to backscatter in the radar geometry. That is, the images are not yet projected to geographic coordi-

nates and represent the signal as seen by the radar (i.e., given by time of travel of microwaves). Level 1.5 and 2.0 have been projected into geographic coordinates, such as UTM, or geographic latitude and longitude. The level 1.X data are used for radar interferometry: two images are complex-multiplied to obtain the interferogram. The latter is typically expressed as  $A^{-\phi}$ , with amplitude  $A$  and phase  $\phi$ . Level 1.5 and 2.0 are expressed as the radar backscatter  $\sigma^0$ . The georeferenced interferogram is seldom distributed and the user is expected to perform the processing. For backscatter, the user has a wide choice of processing level, and can perform radar

processing to specifically enhance or retrieve specific image features. Processing details and analysis to obtain science-ready backscatter images are described in the tutorial.

## 6.2.6 MAPPING MANGROVE CANOPY STRUCTURE

Forest structure can be described in terms of its spatial extent, spatial heterogeneity, tree cover, canopy height profile, and AGB. However, what can be expected from radar observations of mangrove forest landscapes?

Tree cover, canopy height, and AGB are correlated. However, different radar parameters are used to estimate AGB and forest canopy height. A generic formulation relating radar backscatter to AGB is the following, where  $a$  and  $b$  are determined by the user from in situ data:

$$\sigma^0(\text{dB}) = a + b \times \log(\text{AGB}) .$$

In wet tropical forests, values are about -22.5 and 3 for  $a$  and  $b$ , respectively, at L-band. Other equations have also been published, and may require slightly more complex calculation (Yu and Saatchi 2016):

$$\sigma^0(\text{linear}) = Ax^a + (1 - e^{(-Bx)}) + C ,$$

where  $x$  is the AGB, and  $A, B, C$ , and  $a$  are coefficients that can be fitted empirically through iteration until  $x$  results in the observed  $\sigma^0$  (linear). For wet tropical forest, the coefficient values are  $a=0.013682, A=0.21116, B=0.051846, C=0.02192$ . The coefficient  $a$  should be fitted locally. In either case, the fitting parameters can change significantly for mangroves given the variation in structure for a given species and the inundation state. Several authors found radar backscatter decreased significantly with AGB, attributing increased double-bounce scattering at low AGB (e.g., Lucas et al. 2007, Cohen et al. 2013). Nonetheless, one may expect  $\sigma^0$  to increase with forest AGB up to a saturation value that depends on the radar frequency (i.e., P, L, C, X, K) before it decreases. The longer the wavelength (equivalent to lower frequency), the larger the biomass saturation point (Mougin et al. 1992, Proisy et al. 2002). While general literature is not definite on the backscatter signature of mangrove forests, the upper biomass level detectable with radar is similar to other forests, about 200, 100, 50, and 25 for P, L, C, and X, respectively. After this point, the observed  $\sigma^0$  reduces due to absorption

by the dense aerial root system found in mangroves (Lucas et al. 2007). This holds for HV; however, observation of scrub mangroves at HH and VV sometimes also display high  $\sigma^0$  due to increased penetration within the canopy and double-bounce interaction with the water surface or water-saturated ground. While these effects imply estimation of mangrove AGB from backscatter alone is generally difficult and strongly site-dependent, these polarimetric trends (i.e., volume versus double-bounce) can be used to classify mangrove type and also structure (e.g., Hong et al. 2015, Brown et al. 2016), itself related to AGB. For additional information on SAR for biomass estimation, see [Chapter 5](#).

It is difficult to map the extent of mangrove forests using radar alone, in particular when the adjacent inland landcover is another forest type. This can also be difficult with optical sensors, as 'color' (e.g., greenness) may not suffice to distinguish mangroves from other vegetation types. Therefore, it is recommended to use a combination of datasets obtained from different sensing technologies (also discussed in [Chapter 3, Sec. 3.5.4](#)). Landcover classification can generally be performed with the radar backscatter as one of the layers along with data from optical instruments such as Landsat. One can also build upon existing global maps of mangrove extent (e.g., Giri et al. 2011) to extract the area of mangrove forests from the radar data. It is more efficient to start with reliable remote sensing products and improve them rather than reinvent the wheel. Classification can be performed with supervised methods (e.g., maximum likelihood, decision trees, neural network) or unsupervised methods (e.g., ISODATA). Implementations for these algorithms can be found in all major commercial remote sensing software (e.g., ENVI) and also in open source software (QGIS or Python libraries). Training of supervised classifiers requires knowledge of mangrove forest locations that can be easily interpreted by an experienced photo-interpreter or from in situ field surveys. It is important that the training set be representative of the entire range of spectral signature observed with remote sensing. Otherwise, pixels with an untrained spectral signature in radar and optical data, will be thrown into the wrong classes, potentially classified as mangroves. To avoid these issues, an initial unsupervised classification method can be used, followed

by manual merging of 'unsupervised classes' into relevant mangroves classes.

To generate a landcover classification of mangrove forests and type, it is recommended to use radar in combination with different sensing technologies (Nascimento et al. 2013). Discrimination of mangroves against other types of forests inland with L-band data given a priori environmental information (e.g., digital elevation model and water mask) and other sources of optical data such as Landsat (Lucas et al. 2014, Bunting et al. 2018). However, classification results can vary greatly due to availability of polarimetric layers and instrument wavelength and technique. For example, Fonteh et al. (2016) found Sentinel-1 (dual-pol C-band) data did not provide significant improvement over Landsat-based landcover classification, while Zhen et al., found increased accuracy over 10% using fully polarimetric Radarsat-2 data. Aslan et al., 2016, began with a spatial segmentation of an L-band HH and HV dataset, with subsequent Landsat-8 species-specific classification refinement. There exist global maps of mangrove extent derived from optical data (e.g., Giri et al. 2011) that can be used to extract the area of mangrove forests from the radar data. These have recently been improved with L-band radar data from ALOS-1 (Bunting et al. 2018). The authors found it more efficient to start with existing but reliable remote sensing products from Giri et al. (2011) and improve the maps with radar.

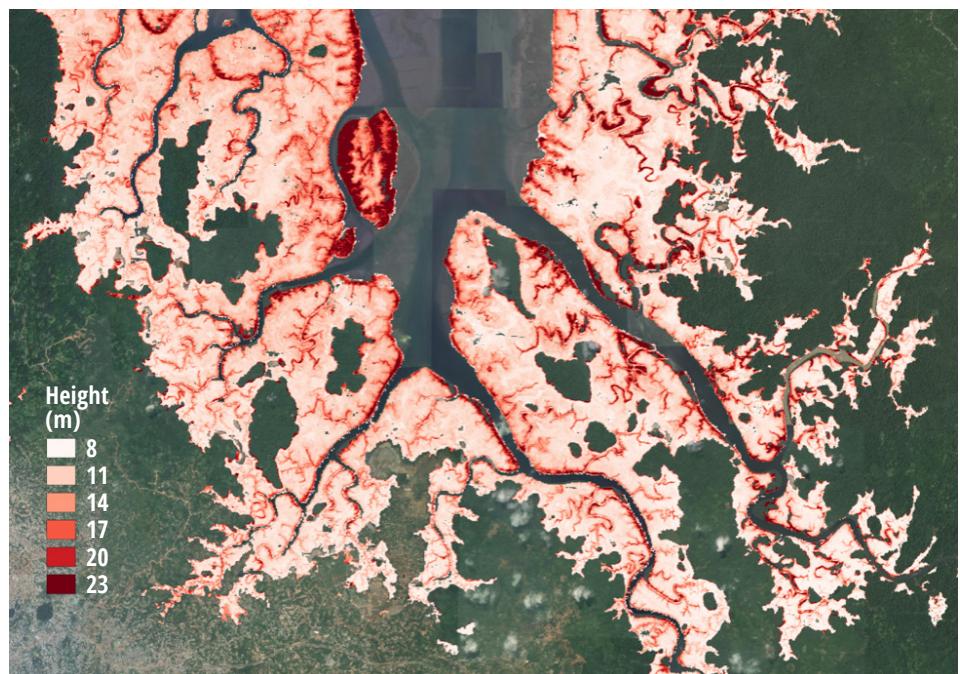
There have been significant advances in the use of Radar interferometry to map mangrove canopy height (Simard et al. 2006, Simard et al. 2008, Fatoyinbo et al. 2013, Lagomasino et al. 2015). Data from the SRTM acquired in February 2000, was the first dataset to enable measurement of mangrove canopy height. SRTM was designed to measure elevation, but due to the interaction of the radar microwave with the canopy volume, the SRTM elevation measurement is biased by forest height and density. That is, a forested hill top will appear higher than it is. Assuming mangroves are located at mean sea level with negligible topography, the SRTM elevation measurement is directly related to mangrove canopy height. While this may be a gross assumption in mangrove regions with high tides (>3m), it provides a robust method to determine patterns of mangrove height at

the landscape to global scales in the 2000 epoch. New elevation measurements derived with interferometric data was acquired with TanDEM-X in ~2015, providing a second and more recent dataset (Lee et al. 2015). The TanDEM-X elevation datasets are currently available through a proposal process or can be purchased. An example, calibrated with in situ data, is shown in **Figure 6.8**. Coarser resolution data may soon become available.

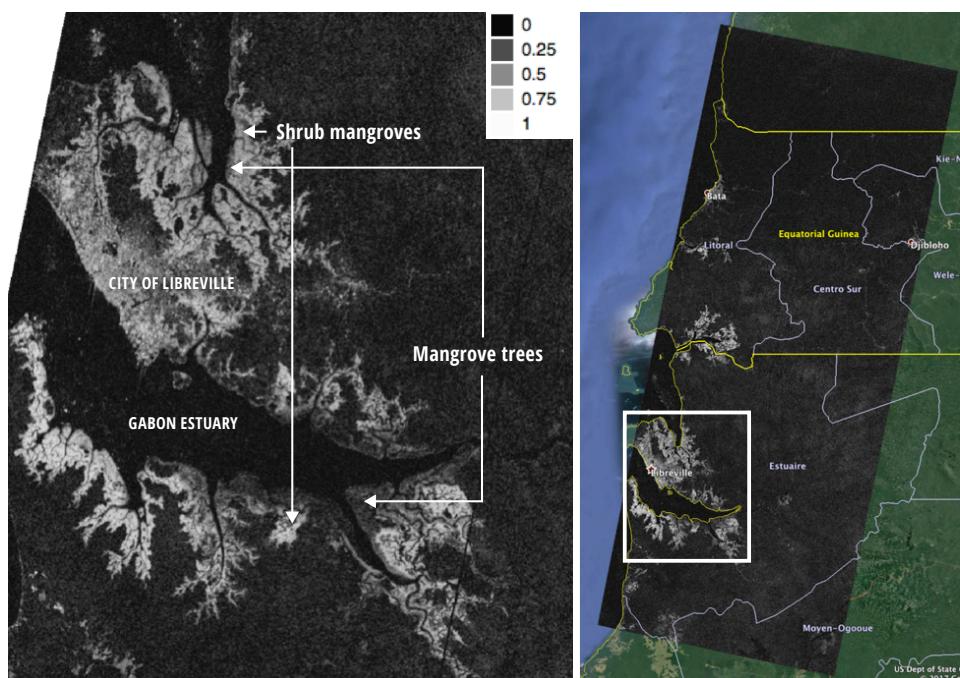
It is interesting to note that such priori remote sensing products provide a powerful tool to select the location of the in situ plots and optimize campaign logistics. Such stratified methods, based on remotely-sensed canopy height, were successfully used in mangrove forests (Trettin et al. 2016, Fatoyinbo et al. 2018). A similar stratified methodology could also be adopted to train and validate land-cover classification of mangrove structure.

### 6.3 Conclusions

Radar remote sensing is a powerful tool to monitor mangrove extent and map general structure attributes (e.g., trees versus shrubs, aerial root). It enables detection of forest cover changes regardless of cloud cover. Used in combination with optical remote sensing observations, radar can improve distinction of mangrove from other types of inland forest. However, the estimation of AGB is limited with radar backscatter alone. Instead, it is recommended to use radar interferometry to accurately map mangrove forest structural attributes, including canopy height and aboveground biomass. While fully polarimetric L-band data are freely available (e.g., UAVSAR and ALOS), such datasets at P-, C-, and X- band that span a wide range of mangrove types are rarely accessible. More research on the application of radar remote sensing to study mangrove forests is needed to understand the polarimetric and spectral signature of mangrove forest structure. Nonetheless, recent progress in global mapping and monitoring of mangrove extent (Bunting et al. 2018) and global assessment of AGB (Simard et al. 2019) clearly illustrates the power of radar remote sensing of mangrove forests.



**Figure 6.8** Map of mangrove canopy height in the Akanda National Park obtained from TanDEM-X elevation data (similar to SRTM). Same region as shown **Figure 6.7**.



**Figure 6.9** Interferometric radar coherence obtained from ALOS-2 LHH. The bright areas represent coherence close to 1, indicating strong similarity between images collected successively after 24 days. High coherence is observed in open shrub mangroves, as well as in urban areas. Tall mangrove trees exhibit low coherence due to temporal decorrelation caused by scattering in the canopy, which changes between the two radar acquisitions (i.e., 24 days in this case). Open water surface also displays low coherence due to waves constantly changing surface scattering.

## 6.4 References

- A. Aslan, A. F. Rahman, M. W. Warren, and S. M. Robeson, "Mapping spatial distribution and biomass of coastal wetland vegetation in Indonesian Papua by combining active and passive remotely sensed data," *Remote Sens. Environ.*, vol. 183, pp. 65–81, 2016.
- A. Held and C. Ticehurst, "High resolution mapping of tropical mangrove ecosystems using hyperspectral and radar remote sensing," ... *Remote Sens.*, no. October 2014, pp. 37–41, 2003.
- Alongi, D. (2002). Present state and future of the world's mangrove forests. *Environmental Conservation*, 29(3), 331-349. doi:10.1017/S0376892902000231
- Arreola-Lizárraga, José Alfredo, Francisco Javier Flores-Verdugo, and Alfredo Ortega-Rubio. "Structure and litterfall of an arid mangrove stand on the Gulf of California, Mexico." *Aquatic botany* 79.2 (2004): 137-143.
- Asbridge, Emma, et al. "Mangrove response to environmental changes predicted under varying climates: case studies from Australia." *Current Forestry Reports* 1.3 (2015): 178-194.
- Barbier, E. B. et al. e value of estuarine and coastal ecosystem services. *Ecol. Monogr.* 81, 169–193 (2011).
- Bunting, P., Rosengqvist, A., Lucas, R., Rebelo, L. M., Hilarides, L., Thomas, N., ... & Finlayson, C. (2018). The Global Mangrove Watch—A New 2010 Global Baseline of Mangrove Extent. *Remote Sensing*, 10(10), 1669. Data available here : <http://data.unep-wcmc.org/datasets/45>
- Castañeda-Moya, E., Twilley, R. R., & Rivera-Monroy, V. H. (2013). Allocation of biomass and net primary productivity of mangrove forests along environmental gradients in the Florida Coastal Everglades, USA. *Forest Ecology and Management*, 307, 226-241.
- Cheong, So-Min, et al. "Coastal adaptation with ecological engineering." *Nature climate change* 3.9 (2013): 787.
- Cohen, R., J. Kaino, J.A. Okello, J.O. Bosire, J.G. Kairo, M. Huxham, M. Mencuccini. Propagating uncertainty to estimates of above-ground biomass for Kenyan mangroves : A scaling procedure from tree to landscape level. *Forest Ecology and Management*, vol. 310, pp. 968-982, 2013.
- D. C. Donato, J. B. Kauffman, D. Murdiyarso, S. Kurnianto, M. Stidham, and M. Kanninen, "Mangroves among the most carbon-rich forests in the tropics," *Nat. Geosci.*, vol. 4, no. 5, pp. 293–297, 2011.
- Denbina, Michael, Marc Simard, and Brian Hawkins. "Forest Height Estimation Using Multi-baseline PolInSAR and Sparse Lidar Data Fusion." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 99 (2018): 1-19.
- Duke, Norman, Marilyn Ball, and Joanna Ellison. "Factors influencing biodiversity and distributional gradients in mangroves." *Global Ecology & Biogeography Letters* 7.1 (1998): 27-47.
- E. Mougin, C. Proisy, G. Marty, F. Fromard, H. Puig, J. L. Betouille, and J. P. Rudant, "Multifrequency and multipolarization radar backscattering from mangrove forests," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1 PART 1, pp. 94–102, 1999.
- Eslami-Andargoli, L., et al. "Mangrove expansion and rainfall patterns in Moreton Bay, south-east Queensland, Australia." *Estuarine, Coastal and Shelf Science* 85.2 (2009): 292-298.
- Fatoyinbo, T. E., et al., (2008). Landscape-scale extent, height, biomass, and carbon estimation of Mozambique's mangrove forests with Landsat ETM+ and Shuttle Radar Topography Mission elevation data. *Journal of Geophysical Research* 113, G02S06
- Fatoyinbo, T.E. & M. Simard (2013). "Height and biomass of mangroves in Africa from ICESat/ GLAS and SRTM", *International Journal of Remote Sensing*, Vol. 34, 2, pp. 668-681.
- Fatoyinbo, Temilola E., and Marc Simard. "Height and biomass of mangroves in Africa from ICESat/GLAS and SRTM." *International Journal of Remote Sensing* 34.2 (2013): 668-681.
- Fatoyinbo, Temilola, et al. "Estimating mangrove aboveground biomass from airborne LiDAR data: A case study from the Zambezi River delta." *Environmental Research Letters* 13.2 (2018): 025012.
- Fonteh, M. L., Theophile, F., Cornelius, M. L., Main, R., Ramoelo, A., & Cho, M. A. (2016). Assessing the Utility of Sentinel-1 C Band Synthetic Aperture Radar Imagery for Land Use Land Cover Classification in a Tropical Coastal Systems When Compared with Landsat 8. *Journal of Geographic Information System*, 8(04), 495.
- Freeman, A.; Durden, S.I. A three-component scattering model for polarimetric SAR data. *IEEE Trans. Geosci. Remote Sens.* 1998, 36, 963–973.
- Gilman, Eric L., et al. "Threats to mangroves from climate change and adaptation options: a review." *Aquatic botany* 89.2 (2008): 237-250.
- Giri, C., et al., 2011. Status and distribution of mangrove forests of the world using earth observation satellite data. *Global Ecology and Biogeography*, 20(1), 154-159
- Hansen et al., High-Resolution Global Maps of 21st-Century Forest Cover Change, *Science* 15 November 2013: 342 (6160), 850-853.
- Hashimoto, T.R. Mid-holocene development of mangrove communities featuring rhizophoraceae and geomorphic change in the Richmond River Estuary, New South Wales, Australia. *Geographical Research*, vol. 44, no. 1, pp. 63 – 76, 2006.
- Hutchison, James, et al. "Predicting global patterns in mangrove forest biomass." *Conservation Letters* 7.3 (2014): 233-240.
- I. Brown, S. Mwansasu, and L. O. Westerberg, "L-band polarimetric target decomposition of mangroves of the rufiji delta, Tanzania," *Remote Sens.*, vol. 8, no. 2, 2016.

- J. M. Kovacs, X. Jiao, F. Flores-de-Santiago, C. Zhang, and F. Flores-Verdugo, "Assessing relationships between Radarsat-2 C-band and structural parameters of a degraded mangrove forest," *Int. J. Remote Sens.*, vol. 34, no. 20, pp. 7002–7019, 2013.
- Jerath, M. An economic analysis of carbon sequestration and storage service by mangrove forests in Everglades National Park, Florida. Master's thesis, Florida International University, 2012.
- K. W. Krauss, K. L. McKee, C. E. Lovelock, D. R. Cahoon, N. Saintilan, R. Reef, and L. Chen, "How mangrove forests adjust to rising sea level," *New Phytol.*, vol. 202, no. 1, pp. 19–34, Apr. 2014.
- Kuenzer, C., A. Blumel, S. Gebhardt, T. V. Quoc, S. Dech. *Remote sensing of mangrove ecosystems: A review*. *Remote Sensing*, vol. 3, no.5, pp. 878-928, 2011.
- Lagomasino, et al, (2015). High-resolution forest canopy height estimation in an African blue carbon ecosystem. *Remote Sensing in Ecology and Conservation*, 1(1), 51-60.
- Lavalle, Marco, Marc Simard, and Scott Hensley. "A temporal decorrelation model for polarimetric radar interferometers." *IEEE Transactions on Geoscience and Remote Sensing* 50.7 (2012): 2880-2888.
- Lee, SK. et al., "TanDEM-X Pol-InSAR Inversion for Mangrove Canopy Height Estimation," in *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol.8, no.7, pp.3608-3618, July 2015.
- Locatelli, Tommaso, et al. "Turning the tide: how blue carbon and payments for ecosystem services (PES) might help save mangrove forests." *Ambio* 43.8 (2014): 981-995.
- Lucas, R. M., Mitchell, A. L., Rosenqvist, A., Proisy, C., Melius, A., & Ticehurst, C. (2007). The potential of L-band SAR for quantifying mangrove characteristics and change: case studies from the tropics. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 17(3), 245–264. <http://doi.org/10.1002/aqc.833>
- Lucas, R., Rebelo, L. M., Fatoyinbo, L., Rosenqvist, A., Itoh, T., Shimada, M., ... Hilarides, L. (2014). Contribution of L-band SAR to systematic global mangrove monitoring. *Marine and Freshwater Research*, 65(7), 589–603. <http://doi.org/10.1071/MF13177>
- Mayaux, P., P. Holmgren, F. Achard, J. Eva, H. J. Stibig, and A. Branthomme. Tropical forest cover change in the 1990s and options for future monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 360, no. 1454, 2005.
- M. F. Cougo, P. W. M. Souza-Filho, A. Q. Silva, M. E. B. Fernandes, J. R. Dos Santos, M. R. S. Abreu, W. R. Nascimento, and M. Simard, "Radarsat-2 backscattering for the modeling of biophysical parameters of regenerating mangrove forests," *Remote Sens.*, vol. 7, no. 12, 2015.
- M. L. Fonteh, F. Theophile, M. L. Cornelius, R. Main, A. Ramoelo, and M. A. Cho, "Assessing the Utility of Sentinel-1 C Band Synthetic Aperture Radar Imagery for Land Use Land Cover Classification in a Tropical Coastal Systems When Compared with Landsat 8," *J. Geogr. Inf. Syst.*, vol. 08, no. 04, pp. 495–505, 2016.
- M. Simard et al., "Mapping Height and Biomass of Mangrove Forests in the Everglades National Park with SRTM Elevation Data", *Photogrammetric Engineering & Remote Sensing*, Vol. 72, No. 3, March 2006, pp. 299–311.
- Mangrove extinction risk and geographic areas of global concern," *PLoS One*, vol. 5, no. 4, 2010.
- McKee, K. L., I. C. Feller, M. Popp, W. Wanek. *Mangrove isotopic fractionation across a nitrogen vs. Phosphorus limitation gradient*. *Ecology: Ecological Society of America*, vol. 84, no. 4, pp. 1065-1075, 2002.
- McNally, C.G., E. Uchida, and A. J. Gold. *The effect of a protected area on the tradeoffs between short-run and long-run benefits from mangrove ecosystems*. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 34, pp. 13945 – 13950, 2011.
- Mougin, E., A. Lopes, J. Dauzat, D. H. Hoekman, M.A. Karam, A. K. Fung. *Interpretation of microwave signatures of trees by a combined backscattering and plant structural model*. *IGARSS '92 International Geoscience and Remote Sensing Symposium*. 26-29 May 1992.
- Murray, B. C. *Mangroves' hidden value*. *Nature Climate Change*, 2, 773 – 774, 2012.
- Nascimento Jr, W. R., et al., 2013. *Mapping changes in the largest continuous Amazonian mangrove belt using object-based classification of multisensor satellite imagery*. *Estuarine, Coastal and Shelf Science*, 117, 83-93
- O. Hamdan, H. Khalil Aziz, and I. Mohd Hasmadi, "L-band ALOS PALSAR for biomass estimation of Matang Mangroves, Malaysia," *Remote Sens. Environ.*, vol. 155, pp. 69–78, 2014.
- Pendleton, L., D. C. Donato, B. C. Murray, S. Crooks, W. A. Jenkins, S. Sifleet, C. Craft, J. W. Fourqurean, J. B. Kauffman, N. Marba, P. Megonigal, E. Pidgeon, D. Herr, D. Gordon, A. Balderaa.. *Estimating global blue carbone from conversion and degradation of vegetated coastal ecosystems*. *Plos One*, vol. 7, 2012.
- Pinto, Naiara, Marc Simard, and Ralph Dubayah. "Using InSAR coherence to map stand age in a boreal forest." *Remote Sensing* 5.1 (2012): 42-56.
- Polidoro, B. A., K. E. Carpenter, L. Collins, N. C. Duke, A. M. Ellison, J. C. Ellison, E. J. Farnsworth, E. S. Fernando, K. Kathiresan, N. E. Koedam, S. R. Livingstone, T. Miyagi, G. E. Moore, V. N. Nam, J. E. Ong, J. H. Primavera, S. G. Salmo III, J. C. Sanciangco, S. Sukardjo, Y. Wang, J. W. H. Yong. *The loss of species : Mangrove extinction risk and geographic areas of global concern*. *Plos One*, vol. 5, no.4, 1 – 10, 2010.
- Primavera, J.H. *Development and conservation of Philippine mangroves : Institutional issues*. *Ecological Economics*, vol. 35, no. 1, pp 91 – 106, 2000.
- Proisy, C., Mougin, E., Fromard, F., Trichon, V., & Karam, M. A. (2002). *On the influence of canopy structure on the radar backscattering of mangrove forests*. *International Journal of Remote Sensing*, 23(20), 4197–4210. <http://doi.org/10.1080/01431160110107725>
- R. M. Lucas, A. L. Mitchell, A. Rosenqvist, C. Proisy, A. Melius, and C. Ticehurst, "The potential of

- L-band SAR for quantifying mangrove characteristics and change: case studies from the tropics,"* *Aquat. Conserv. Mar. Freshw. Ecosyst.*, vol. 17, no. 3, pp. 245–264, May 2007.
- Rivera-Monroy, V.H., et al., 2011a. *Saiinty and Chlorophyll a as Performance Measures to Rehabilitate a Mangrove-Dominated Deltaic Coastal Region: the Ciénaga Grande de Santa Marta-Pajarales Lagoon Complex, Colombia.* *Estuaries and Coasts* 34: 1-19.
- S. E. Hamilton and D. Casey, "Creation of a high spatio-temporal resolution global database of continuous mangrove forest cover for the 21st century (CGMFC-21)," *Glob. Ecol. Biogeogr.*, vol. 25, no. 6, pp. 729–738, 2016.
- S. H. Hong, H. O. Kim, S. Wdowinski, and E. Feliciano, "Evaluation of polarimetric SAR decomposition for classifying wetland vegetation types," *Remote Sens.*, vol. 7, no. 7, pp. 8563–8585, 2015.
- Saintilan, Neil, et al. "Mangrove expansion and salt marsh decline at mangrove poleward limits." *Global change biology* 20.1 (2014): 147-157.
- Shapiro, A.C., C.C. Trettin, H. Küchly, S. Alvinapanah and S. Bandeira. *The mangroves of the Zambezi Delta from 1994-2013: increase in extent observed via satellite.* Submitted to *Remote Sensing* (in review).
- Simard, M., L. Fatoyinbo, C. Smetanka, V. H. Rivera-Monroy, E. Castaneda-Moya, N. Thomas, T. Van der Stocken. *Mangrove canopy height globally related to precipitation, temperature, and cyclone frequency.* *Nature Geoscience*, vol. 12, pp. 40-45, 2019.
- Simard, M., et al., "A Systematic Method for 3D Mapping of Mangrove Forests Based on Shuttle Radar Topography Mission Elevation Data, ICESat/GLAS Waveforms and Field Data: Application to Ciénaga Grande de Santa Marta, Colombia," *Remote Sensing of the Environment*, Vol. 112/5, pp. 2131-2144, 2007.
- Simard, M., Fatoyinbo, T., Smetanka, C., Rivera-Monroy, V., Castaneda-Moya, E., Thomas, N., & Van der Stocken, T. (2019). *Mangrove canopy height globally related to precipitation, temperature and cyclone frequency.* *Nature Geoscience*, 12(January). <http://doi.org/10.1038/s41561-018-0279-1>
- Simard, Marc, et al. "A systematic method for 3D mapping of mangrove forests based on Shuttle Radar Topography Mission elevation data, ICESat/GLAS waveforms and field data: Application to Ciénaga Grande de Santa Marta, Colombia." *Remote Sensing of Environment* 112.5 (2008): 2131-2144.
- Simard, Marc, et al. "An empirical assessment of temporal decorrelation using the uninhabited aerial vehicle synthetic aperture radar over forested landscapes." *Remote Sensing* 4.4 (2012): 975-986.
- Simard, Marc, et al. "Mapping height and biomass of mangrove forests in Everglades National Park with SRTM elevation data." *Photogrammetric Engineering & Remote Sensing* 72.3 (2006): 299-311.
- Soares, Mário Luiz Gomes, et al. "Southern limit of the Western South Atlantic mangroves: Assessment of the potential effects of global warming from a biogeographical perspective." *Estuarine, Coastal and Shelf Science* 101 (2012): 44-53.
- Souza-Filho, P. W. M., et al. 2011. *Discrimination of coastal wetland environments in the Amazon region based on multi-polarized L-band airborne Synthetic Aperture Radar imagery.* *Estuarine, Coastal and Shelf Science*, 95(1), 88-98.
- Spalding, M., F. Blasco, C. Field. *World Mangrove Atlas.* International Society for Mangrove Ecosystems, WCMC, National Council for Scientific Research, Paris, 1997.
- Spalding, M., et al., 2010. *World atlas of mangroves.* Earthscan.
- T. C. Jennerjahn and V. Ittekot, "Relevance of mangroves for the production and deposition of organic matter along tropical continental margins," *Naturwissenschaften*, vol. 89, no. 1, pp. 23–30, 2002.
- Thomas, N., Lucas, R., Bunting, P., Hardy, A., Rosenqvist, A., & Simard, M. (2017). Distribution and drivers of global mangrove forest change, 1996–2010. *PLoS one*, 12(6), e0179302.
- Trettin, Carl C., Christina E. Stringer, and Stanley J. Zarnoch. "Composition, biomass and structure of mangroves within the Zambezi River Delta." *Wetlands ecology and management* 24.2 (2016): 173-186.
- Valiela, I., J. L. Bowen, J. K. York. *Mangrove Forests : One of the world's threatened major tropical environments : at least 35% of the area of mangrove forests has been lost in the past two decades, losses that exceed those for tropical rain forests and coral reefs, two other well-known threatened environments.* *BioScience*, vol. 51, no. 10, pp 807 – 815, 2001.
- Vo, Quoc Tuan, et al. "Review of valuation methods for mangrove ecosystem services." *Ecological indicators* 23 (2012): 431-446.
- Wilson, Nicholas C., and Neil Saintilan. "Growth of the mangrove species *Rhizophora stylosa* Griff at its southern latitudinal limit in eastern Australia." *Aquatic botany* 101 (2012): 8-17.
- Yu, Y. and S. Saatchi. (2016). *Sensitivity of L-band SAR Backscatter to aboveground biomass of global forests.* *Remote Sensing*, 8(6), 522.
- Zhen, J., Liao, J. and Shen, G., 2018. *Mapping Mangrove Forests of Dongzhraigang Nature Reserve in China Using Landsat 8 and Radarsat-2 Polarimetric SAR Data.* *Sensors*, 18(11), p.4012.

# APPENDIX E

## Mapping and Monitoring Mangrove Forests with Radar Remote Sensing – Chapter 6 Training Module

### Required open-source software and libraries:

- QGIS, SNAP, GDAL

### Freely-available sources of radar data:

- ALOS PALSAR-1 (images and mosaics)
- ALOS PALSAR-2 (mosaic only; images available via proposal)
- JERS-1 Mosaics
- Sentinel-1 a/b (images)

In this tutorial, we will process and analyze radar images, monitor land use change, and estimate aboveground biomass in mangrove forests. There are a few steps required before we can proceed, which include installing the necessary software and finding data sources.

## 1 INSTALLING STEP/SNAP SOFTWARE

First, we need to find, download and install the Sentinel Application Platform (SNAP) software, which is the main tool used for this tutorial. SNAP is distributed freely by the European Space Agency (ESA) and can be used to process various types of data including Sentinel-1 and -2, Radarsat, TerraSAR-X, ALOS/PALSAR-1 and -2, among others. Data can be processed from level 1 (i.e. Single Look Complex data) to level 2 (e.g. terrain projected) to level 3 (e.g. land cover classification) and polarimetric decomposition.

The SNAP software can be found here: <http://step.esa.int/main/download/> and it is recommended to read info as the software is regularly updated. The page should look like **Figure 1.1**. Find more information here: <http://step.esa.int/main/toolboxes/snap/> and here: STEP: <http://step.esa.int/main/>. Select the binary installer appropriate to your operating system (i.e. Windows, Mac OSX or UNIX).



**Figure 1.1** ESA browser page with access to SNAP software downloads.

## 2 FIND AND OBTAIN ALOS DATA

We will begin this tutorial with science-ready radar data which were processed by the Japanese space agency (JAXA). JAXA produced radar mosaics for each of its spaceborne radar missions: JERS-1 (1996), ALOS-1-PALSAR (2007-2010) and ALOS-2 (2014-2017). The data is science-ready and has

already been radiometrically calibrated, terrain corrected and ground projected. JAXA also produces forest/non-forest maps for each year that can be used to monitor forest disturbances. The steps below show where to find the data and register as a user. The following section explains how to select and download the data.

## 2.1 JAXA Website Registration

The radar mosaics can be obtained from the JAXA website; however, it is first required to register in order to access and download data. It is free to register. Follow the steps below:

1. Access JAXA website and register here: [http://www.eorc.jaxa.jp/ALOS/en/palsar\\_fnf/fnf\\_index.htm](http://www.eorc.jaxa.jp/ALOS/en/palsar_fnf/fnf_index.htm), then follow the instructions, including confirming your email.

2. Once registered, click on "Dataset" on the left-hand side: [http://www.eorc.jaxa.jp/ALOS/en/dataset/data-set\\_index.htm](http://www.eorc.jaxa.jp/ALOS/en/dataset/data-set_index.htm)

- Click on "Global PALSAR-2/PALSAR/JERS-1 Mosaic and Forest/Non-Forest map." Scroll down to "4. Download" and click on [http://www.eorc.jaxa.jp/ALOS/en/palsar\\_fnf/data/index.htm](http://www.eorc.jaxa.jp/ALOS/en/palsar_fnf/data/index.htm), and log in using your email and password created during the registration process.

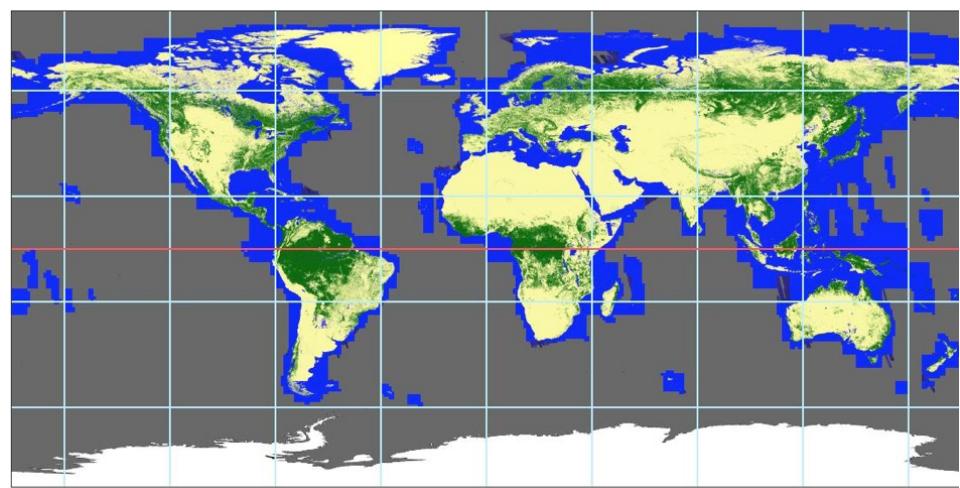
## 2.2 Downloading JAXA ALOS Mosaics

In this section, we will select and download radar mosaics from the JAXA website (**Figure 1.2**). You will find global mosaics from JERS-1 for year 1996. JERS-1 was an L-band, HH-only instrument that collected data from 1992-1998. However, in order to obtain global coverage JAXA had to use data from the full lifetime of the mission to produce a ~1996 mosaic. This dataset provides the oldest L-band radar-based global observation of the Earth, serving as a baseline for temporal analysis of land use.

For the most recent spaceborne L-band radar missions (ALOS-1 and ALOS-2), JAXA generates yearly PALSAR mosaics for 2007, 2008, 2009 and 2010 and PALSAR-2 mosaics for 2015, 2016, 2017, and soon 2018. The ALOS mosaics are distributed with dual polarization, HH and HV, which are ideal for monitoring wetlands and forests respectively. Indeed, the HH polarization exhibits a strong double-bounce effect in inundated wetlands, while HV is the best polarization configuration to map and monitor forest cover and above ground biomass. It is important to note the mosaic tiles are constructed from multiple radar images that were not acquired at the same time. As such, the mosaics will not represent, for

**Figure 1.2** Desktop version of the JAXA website with access to mosaic downloads. You must first register as a user to access the data.

**Figure 1.3** JAXA page with access to all available yearly radar mosaics.



**Figure 1.4** Large scale grid cells of the 2017 ALOS-2 mosaic are displayed. Clicking on a given cell zooms in to a finer grid. Only the  $1^{\circ} \times 1^{\circ}$  cells can be downloaded.

example, the state of inundation at its maximum annual extent. However, they provide a useful snapshot.

Click on [http://www.eorc.jaxa.jp/ALOS/en/palsar\\_fnf/data/index.htm](http://www.eorc.jaxa.jp/ALOS/en/palsar_fnf/data/index.htm) to see a page (**Figure 1.3**) showing all available mosaics.

For example, we can select the ALOS/PALSAR-1 Mosaic for 2017. Clicking on the link will lead you to another page (**Figure 1.4**) in the browser with a large-scale grid. You can reach your region of interest by clicking on

its grid cell location until the cell is  $10 \times 10$ . The mosaics are distributed by  $1^{\circ} \times 1^{\circ}$  tiles. To generate a larger mosaic, you must download several  $1^{\circ} \times 1^{\circ}$  tiles which can be merged using gdal tools or QGIS. Here are the steps:

1. Click on your region of interest. (In this workshop, this will be Kenya's coast)
2. Click until you end-up at the desired location. Note that all mosaic grid-cells are  $1^{\circ} \times 1^{\circ}$ .

### 2.3 Downloading 1-degree cells

For PALSAR-1 and PALSAR-2, there are 3 mosaics (see [Figure 1.5](#)).

**1.1 FNF:** Forest vs. Non-Forest map (See M. Shimada, T. Itoh, T. Motooka, M. Watanabe, T. Shiraishi, R. Thapa, and R. Lucas, "New global forest/non-forest maps from ALOS PALSAR data (2007-2010)," *Remote Sens. Environ.*, vol. 155, pp. 13–31, 2014.)

**1.2 HH:** Mosaic of Horizontal-transmit and Horizontal-receive polarization. Sensitive to both volume and surface. Excellent for monitoring wetlands, forest-inundations, urban infrastructure and expansion.

**1.3 HV:** Mosaic of Horizontal-transmit and Vertical-receive polarization. Sensitive to volumes such as vegetation density. Excellent to distinguish and monitor forests.

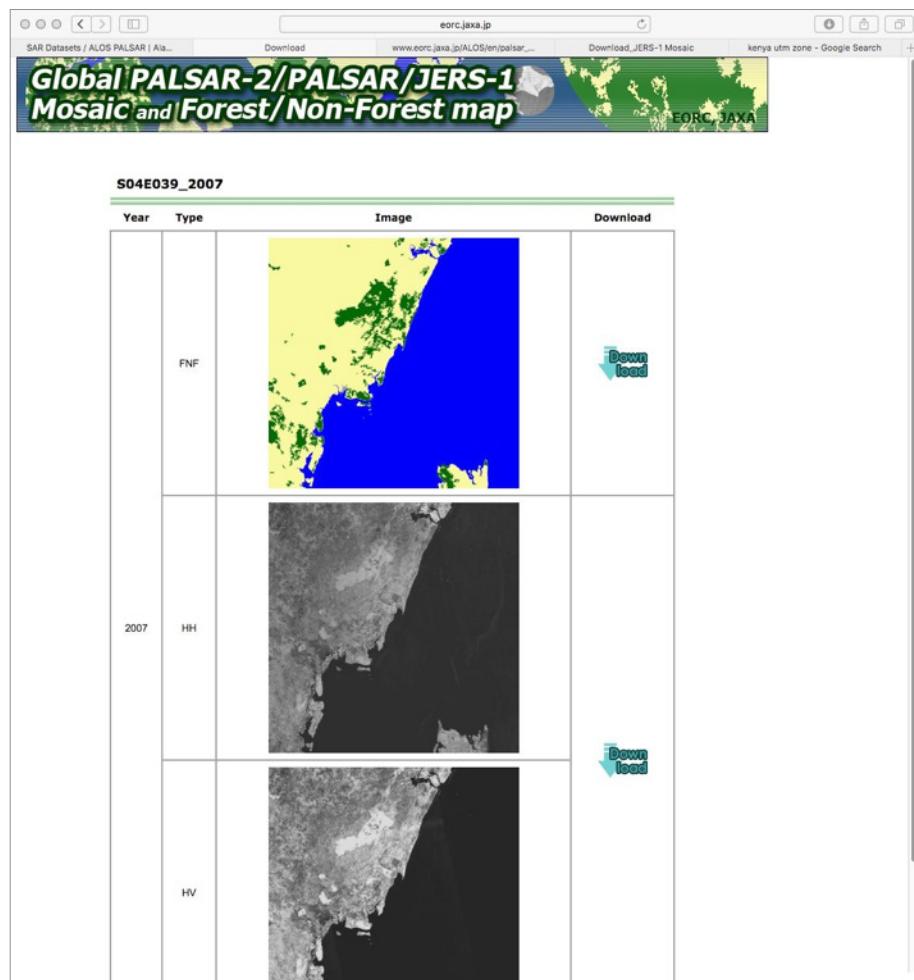
Download a file by clicking on the "Download" icon. The file will automatically download to computer as zipped file; for HH and HV this will be about 65MB. Then you can unzip the files, typically with a double-click on file. The 10x10 tiles are in an ENVI format that can be opened in ENVI or QGIS. To view in QGIS, you can simply drag-and-drop the file in the main QGIS window. As an alternative, you can also use the "gdal\_translate" command in a terminal window to convert in any format using option "-of Gtiff".

## 3 DOWNLOAD RAW ALOS/PALSAR-1 DATA

In this section, we will find and download raw images from the ALOS/PALSAR-1 instrument. To date (2018), the ALOS/PALSAR-2 data is only available via a proposal process or purchase. Other radar datasets, such as ALOS-1, are available through the ASF Vertex interface ([Figure 1.6](#)).

### 3.1 3.1 Searching for radar images in a specific location and within a time period

The steps below describe how to obtain data from the Alaska Satellite Facility (ASF) website; specifically, we can find ALOS/PALSAR-1 data acquired between 2007 and 2010. We will select a small region of interest



**Figure 1.5** JAXA's website showing selected 1<sup>o</sup> × 1<sup>o</sup> tiles over Kenya's coast.

over Kenya's coast. The list of data found by ASF's Vertex tool matching your selection criteria will appear in the right-hand side of the interface, with clickable thumbnail images to help in the selection of data. Here is how to search for data:

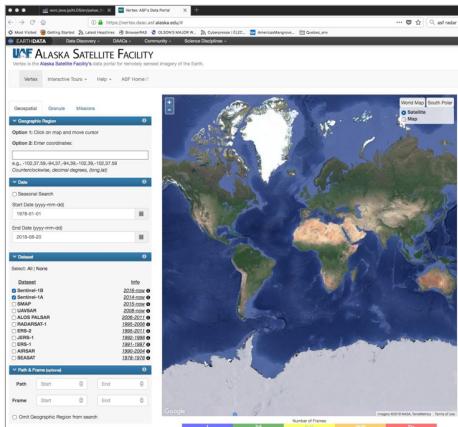
1. Go to the Alaska Satellite Facility Vertex interface: <https://vertex.daac.asf.alaska.edu/>
2. Click on "Geographic Search," and follow the instructions to register to create a (free) account.
3. Go back to the site and log in. Click and drag your cursor from the top left of your study site to the bottom right. A box will be drawn around your selected area. These coordinates will be written automatically in the "Geographic Region" tab.
4. Enter range of dates for which you seek data. For

PALSAR-1, this will be between 2007 and 2010. The "Path" and "Frame" are optional.

5. Click "Search" at the bottom.

Once we have clicked on the "Search" button, the Vertex tool will search for data matching the selected criteria. A list of files will appear on the right with low-resolution thumbnail preview images (Figure 1.7). You will notice that files with different imaging modes will be available:

- 1.1 "WB1"** is (ScanSAR) Wide-Beam #1
- 1.2 PLR** is a fully polarimetric dataset with HH, HV and VV
- 1.3 FBD** is Fine-Beam Dual polarization including HH and HV.
- 1.4 FBS** is Fine-Beam Single polarization (HH)



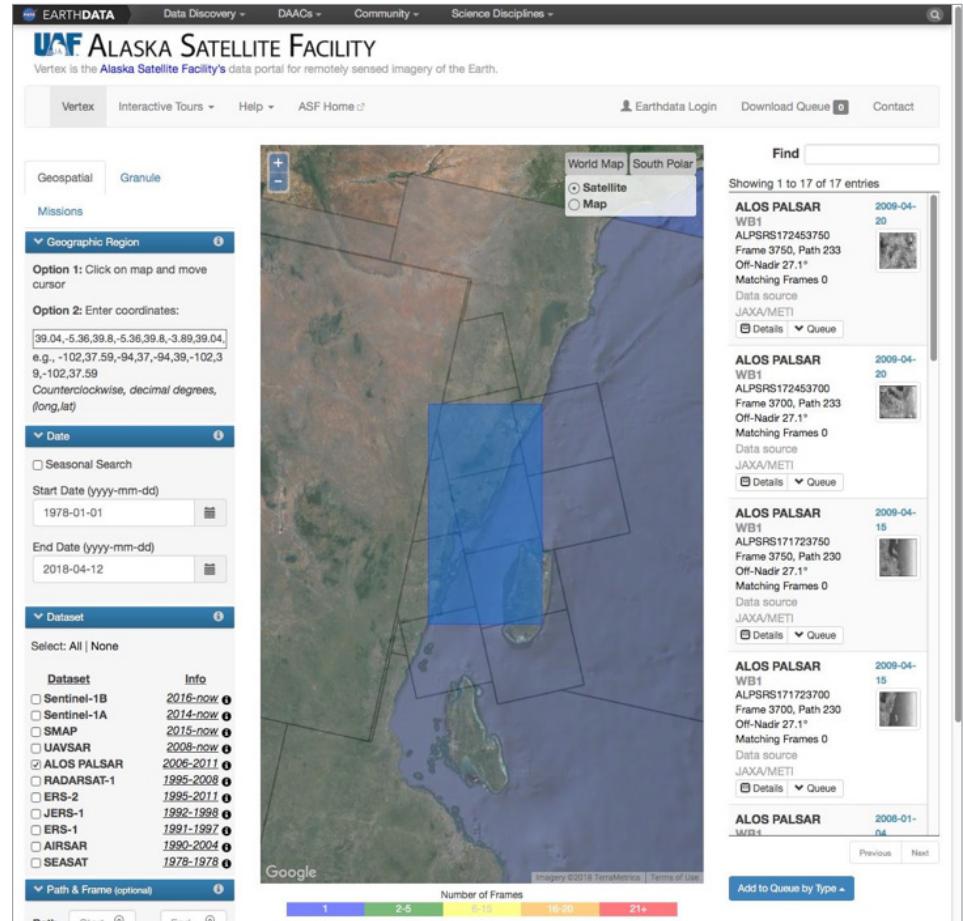
**Figure 1.6** The ASF Vertex interface.

More details on the characteristics of ALOS/PALSAR imaging modes can be found in **Table 1**. In our example, high resolution ascending-orbit images are available, along with WB1 descending-orbit images. Ascending and descending mean the targets are seen from the West and East respectively. A single-click on the thumbnail images will highlight the location of each image within the Vertex's maps.

### 3.2 Select Data

To obtain additional information for each image, click on the “Details” button below any thumbnail. To download an image, click on the “Queue” button. A dropdown menu will appear with options of Level 1.0, 1.1, 1.5, or radiometrically terrain-corrected (RTC). The available options depend on the radar imaging mode (**Table 1**). Try it out. The option to be selected (**Table 2**) depends on your objectives and your level of expertise. Typically, for most processing performed in SNAP, you will need the SLC projected images, which SNAP can convert to geographic projections. If you plan to use a radar image into a GIS system or as an additional band used in your analysis, you should consider using images with the highest level of processing (radiometrically corrected (RTC) and science-ready), similar to the JAXA mosaics but with a given date of acquisition.

It is possible to see if images processed at a given level are available for your region of interest. The ASF website allows you to see data availability as a function of processing level. Access this information



**Figure 1.7** The Vertex interface from ASF displays remote sensing data matching your search criteria. A list of images with low-resolution thumbnails appear on the right-hand side. A single-click on the latter produces the image footprint blue polygon on the map.

	WB1	FBD	FBS	PLR
<b>Swath</b> (i.e. image width)	240 km	70 km	70 km	30 km
<b>Resolution</b>	100 m	20 m	10 m	30 m
<b>Incidence angle</b> (i.e. viewing geometry)	18-43°(27.1)	34.3°	34.3°	21.5°

**Table 1** Summary of ALOS/PALSAR-1 imaging mode characteristics.

	1.0	1.1	1.5	RTC
<b>Geographic projection</b>	Radar (SLC)	Radar (SLC)	Geographic	Geographic
<b>Radiometric format</b>	Single look pixel	Equally-spaced single look pixel in radar coordinates	Multilook and calibrated but with topographic features; data equally spaced on the ground	Science-ready

**Table 2** Radar image processing levels. The radar single look complex (SLC) projection means “as the radar” sees it (i.e., time of return of radar echoes), as opposed to a geographic coordinate projection.

by clicking on <https://www.asf.alaska.edu/sar-data/palsar/alos-acquisition-maps/>. You will be led to the page shown in **Figure 1.8**. If you are interested in ALOS/PALSAR-1 images that have been corrected for the effect of terrain (radiometric terrain correction), click on the last map icon to reach the page shown in **Figure 1.9**.

## 4 RADAR DATA PROCESSING WITH SNAP

In this section, we will use the SNAP software to open and display radar images, perform radiometric calibration, and filter the images to reduce speckle noise (for more information on speckle, refer to Chapter 2 of this Handbook, section 2.1.5). We will then manipulate data with GDAL and load, display, and perform analyses on the images in QGIS.

### 4.1 Read and Calibrate Data

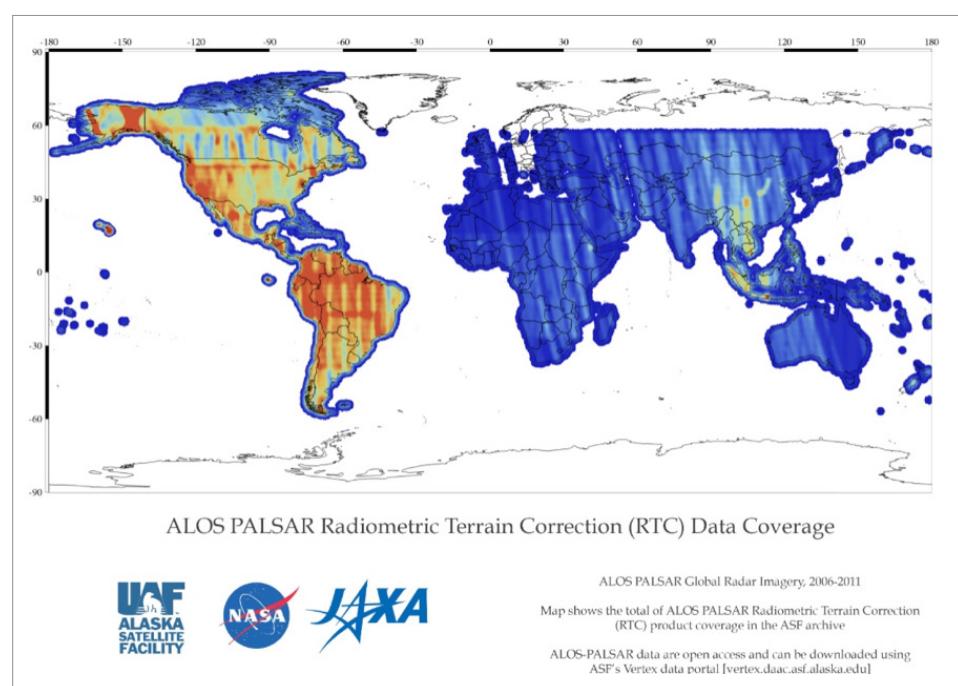
To read in the images, let's begin processing with Level 1.1 data. The 1.5 and RTC product data are already processed. Can you do better using SNAP, bringing Level 1.1 to RTC?

To display the image, first, drag and drop the Level 1 file (either downloaded using the above steps or use the included images in the data.zip file for this training: ALPSRP073337100-L1.1-->VOL-ALPSR-PO73337100-H1.1\_\_A) into the “Product Explorer” window. Double click the file in “Product Explorer”, then double click “Bands”, then click any “Intensity” band. At this point, you should see your image within the SNAP window as shown in **Figure 1.10**.

To calibrate the images:

- Go to the “Radar” menu at the top of the screen. Click “Radiometric,” and then “Calibrate.”
- A window will pop up. Here, specify the input band and output file location.
- In the same window, click the “Processing Parameters” tab. Select any “intensity” band. Select beta0 band in order to obtain radar backscatter as seen by the radar (i.e. in its geometry).

**Figure 1.8** ASF coverage maps indicating the availability of ALOS-1 imagery for different processing levels.



**Figure 1.9** Global availability of ALOS-1 radiometrically terrain corrected (RTC) imagery. The color ramp indicates the number of available processed scenes.

## 4.2 Apply Terrain Correction

The previous step calibrated the radiometry of the images, given the instrument parameters. Now, we seek to remove the impact of topography from the radar image. This is often called “radiometric terrain flattening”.

- Go to the “Radar” top menu and select “Radiometric” > “Radiometric Terrain Flattening.”

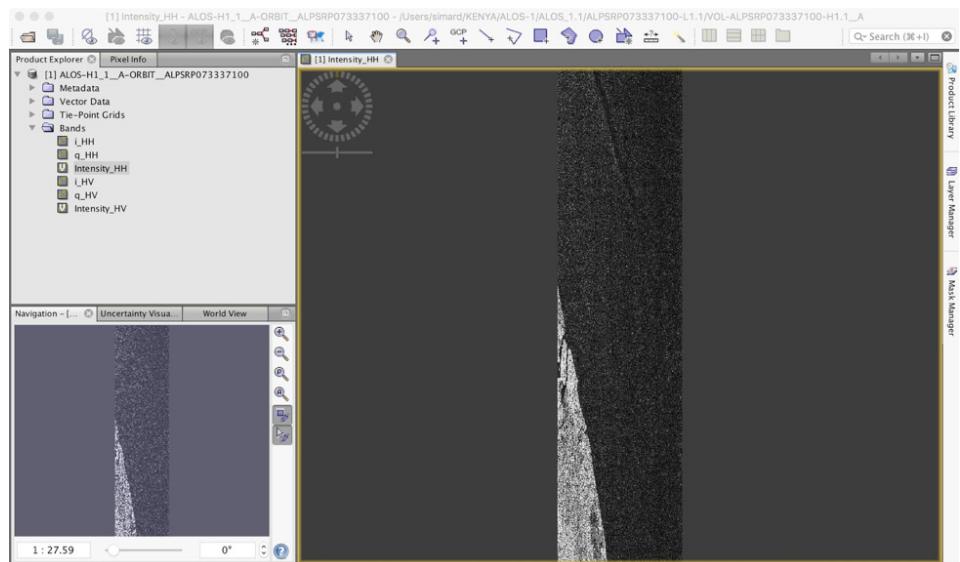
In the “Processing Parameters” tab, select “Beta0” (See [Figure 1.11](#)). Click “Run”

- Note this brings the image from Beta0 to Gamma0 by removing topographic effects on the backscatter.

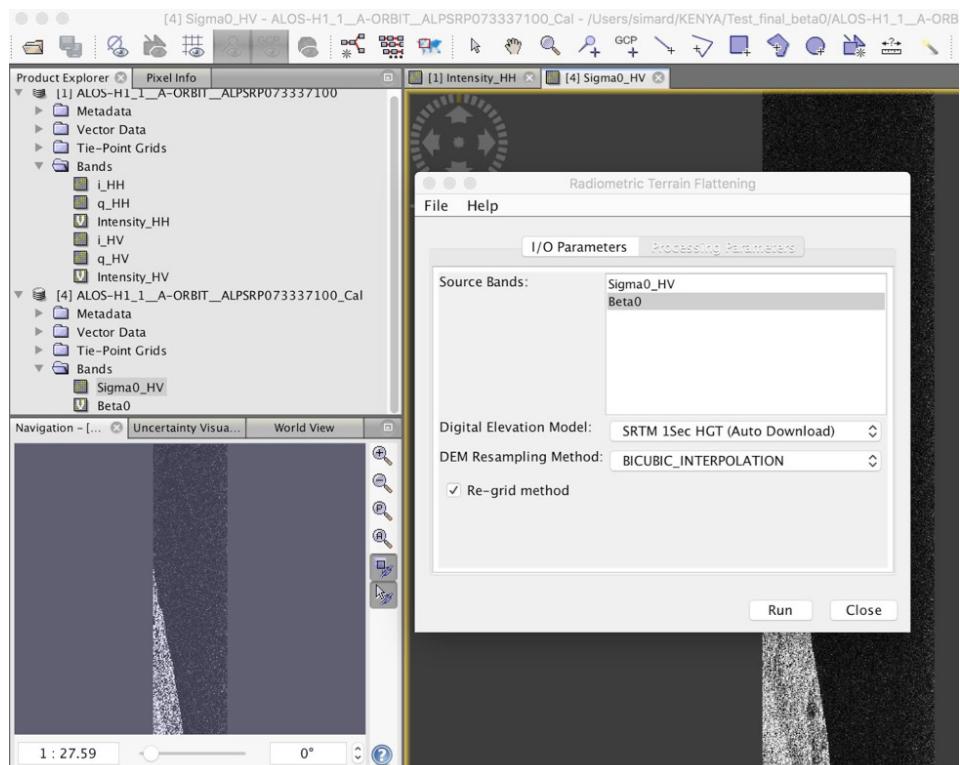
## 4.3 Project Image to Ground Range

Up to this section, the images have been in radar viewing geometry (often called “slant range”). We need to convert this geometry to a geographic projection in order to use the images for scientific analysis and other applications. To perform the geographic projection:

- Click “Radar” > “Geometric” > “Terrain Correction” > “Range-Doppler Terrain Correction”. A window will pop up.
- Select the input filtered band (the image you want to reproject) and specify the output.
- In “Processing parameters”, select Gamma0\_HH or Beta0. If you want to get Gamma0 at this step, it is time efficient and provides a terrain flattened image.
- Select the spatial resolution of output file (the default is radar sampling), and save useful bands, e.g., “Selected source band”, “DEM”, “projected local incidence angle”
- If after visual inspection, you realize that ground projection is incorrect, use the “Radar” > “Geometric” > “Terrain Correction” > “SAR-Simulation Terrain Correction” instead of the “Range-Doppler Terrain Correction”. This method uses a digital elevation model (DEM) to simulate a radar image that is subtracted from the observed image. The DEM is automatically downloaded.



**Figure 1.10** Viewing an intensity image in SNAP



**Figure 1.11** Radiometric Terrain Flattening (sometimes called “Radiometric Terrain Correction”) in SNAP.

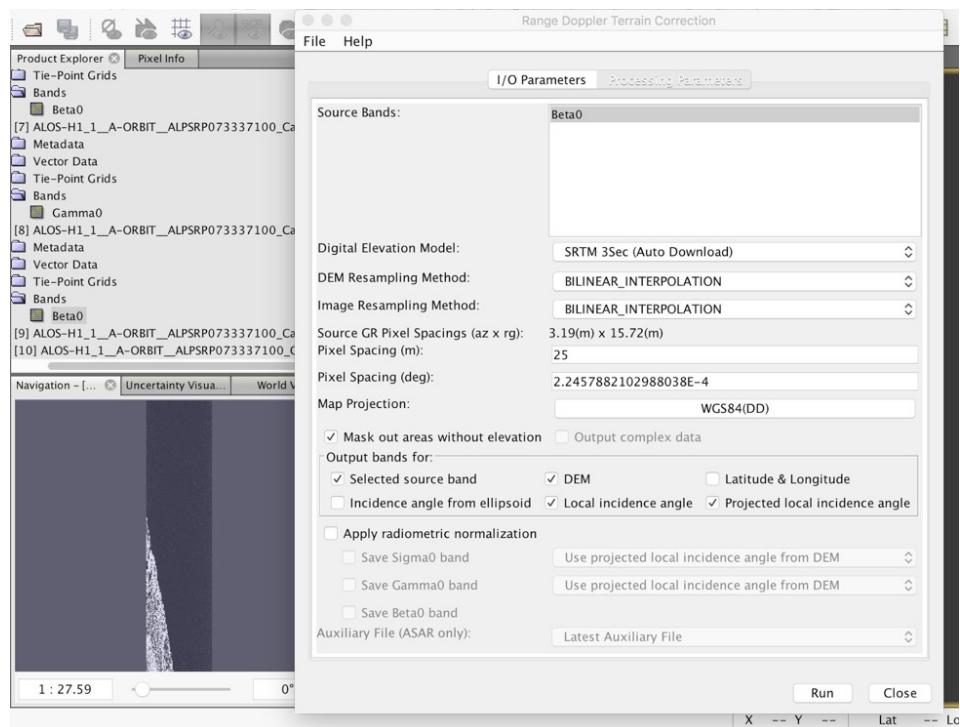
## 4.4 Converting from Linear Amplitude to Decibels (dB)

Radar images consist of a wide range of values, with more pixels at lower values. To compress this range, improve image display, and reduce the impact of multiplicative noise, it is often preferred to represent image values in decibels. To convert from linear (amplitude or intensity) values to decibels (dB), click in the top menu: "Raster" > "Data Conversion" > "Converts bands to/from dB" and select the input band and output files. To write the output, click "File" > "Export" > "Geotiff".

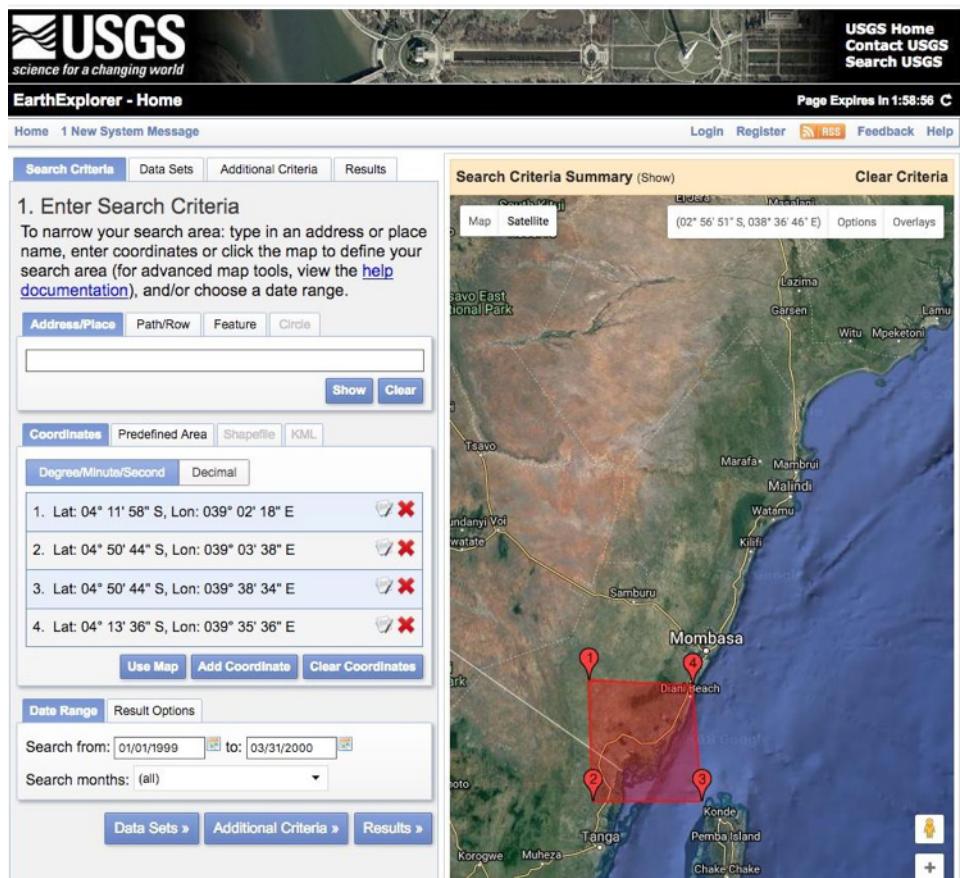
## 5 USING QGIS TO ESTIMATE MANGROVE BIOMASS

We will discuss two methods to estimate mangrove biomass. The first will use canopy height derived from a Digital Elevation Model (DEM). The second method uses the radar backscatter images generated in the previous exercise. As discussed in Chapter 6 of the SAR Handbook, radar backscatter is not a reliable indicator of biomass. Nonetheless, it is worth learning about the technique. Mangrove canopy height is well-estimated with digital elevation models derived from radar interferometry (e.g. SRTM and TanDEM-X). We can obtain the Shuttle Radar Topography Mission (SRTM) DEM from the USGS Explorer website ([Figure 1.13](#)):

- Go to: <https://earthexplorer.usgs.gov/> and create an account to download data
- In "Search Criteria", click "Use Map"
- Click "Clear Coordinates", and generate a polygon for an area of interest (see Figure 13)
- Go to the "Data Sets" tab, and select the type of data needed. The SRTM DEM will be found in "Digital Elevation" > "SRTM". Use the "SRTM 1 Arc-Second Global". Note the upcoming NASA-DEM will not work for this purpose.
- Click "Results" and select files for download. Alternatively, instead of downloading the DEM from earthexplorer.usgs.gov, you can use the DEM\_SRTM file in the data.zip folder provided for this training.



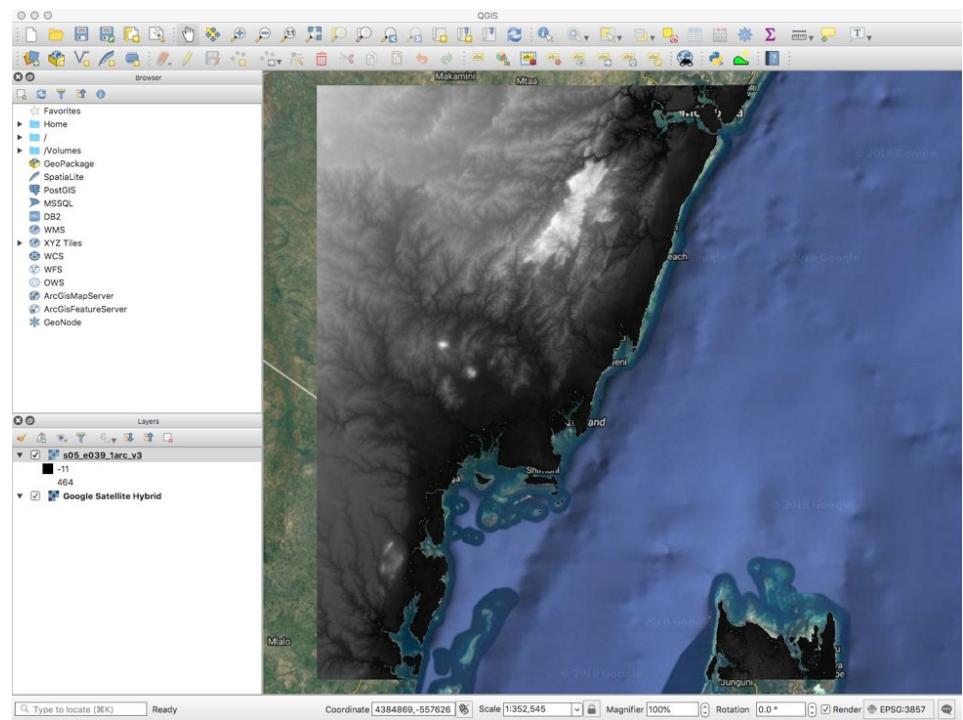
**Figure 1.12** Projecting a radar image from slant range to ground range in SNAP.



**Figure 1.13** The USGS EarthExplorer interface

## 5.1 Reading in and Displaying SRTM DEM Data in QGIS

To visualize the DEM in QGIS, simply drag and drop the DEM file (by default in .tif format) into QGIS (**Figure 1.14**). To adjust the colors, go to “properties” by right clicking on DEM file name in QGIS under “Transparency” > “Additional value” = 0. You can set the background to the Google Satellite image layer using the “XYZ Tiles” setup, which is a QGIS plugin. Drag and drop available bands on the left-hand side in QGIS to set the layer order, with the basemap below the band information layers. Finally, to select an area with mangroves, open the mangrove shapefile that can be found in the data.zip file (Mangroves > africa\_mangrove.shp).



## 5.2 Loading in polygons to clip a raster in QGIS

This section explains how to use existing polygons (e.g. shapefiles) to extract values from a raster image. Follow the steps below:

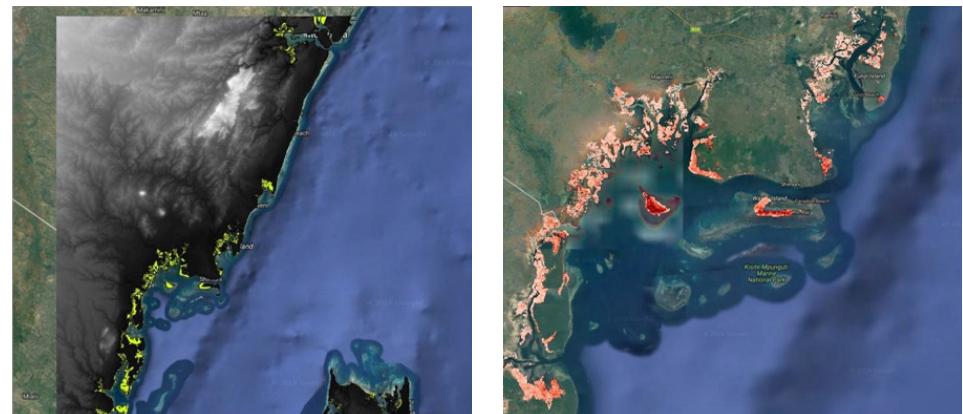
- Drag and drop the mangrove shapefile into QGIS (Figure 1.15). You can change the polygon display color via a right click > “Properties”
- Cut the SRTM DEM with a polygon: Via the top menu: “Raster” > “extraction” > “Clip Raster by Mask Layer”

If QGIS fails to clip the SRTM DEM, you may need to use the command line. Copy and paste this text into a terminal window:

```
gdalwarp -ot Float32 -of GTiff -tr
0.00027778 -0.00027778 -tap -cut-
line /YourDirectoryWithShape-
File/africa_mangrove_withindem-
shp -crop_to_cutline -dstno-
data 0.0 / YourDirectoryWithS-
RTM /s05_e039_1arc_v3.tif /
YourDirectoryWithOutputs /
s05_e039_1arc_v3_mangroves.tif
```

You now have a canopy height map. With SRTM, elevation corresponds to basal area weighted height (also called Lorey's height). The maximum height is generally 1.6X this value. (Simard et al., 2019). See the results in **Figure 1.15**.

**Figure 1.14** View of the SRTM DEM overlaid on the Google Satellite baselayer in QGIS.



**Figure 1.15** View of mangrove extents from an existing polygon (left) and via corresponding canopy height (right).

### 5.3 Applying the Generic Height-Based Allometric Equation

In this section, we will estimate mangrove biomass using canopy height estimates from SRTM DEM. This can be easily achieved applying the allometric equation directly to the DEM. There are several generic equations relating SRTM to canopy height and aboveground biomass (Simard et al., 2019):

- Basal area weighted height:  $H_{ba} \sim 1.08^*SRTM$
- Maximum canopy height:  $H_{max} \sim 0.93^*1.7^*SRTM$
- Aboveground Biomass:  $B \sim 3.25^*H_{ba}^{1.53}$

To apply the allometric equation, follow these steps:

In QGIS, click on “Raster” in the top menu, then “Raster Calculator” (see **Figure 1.16**)

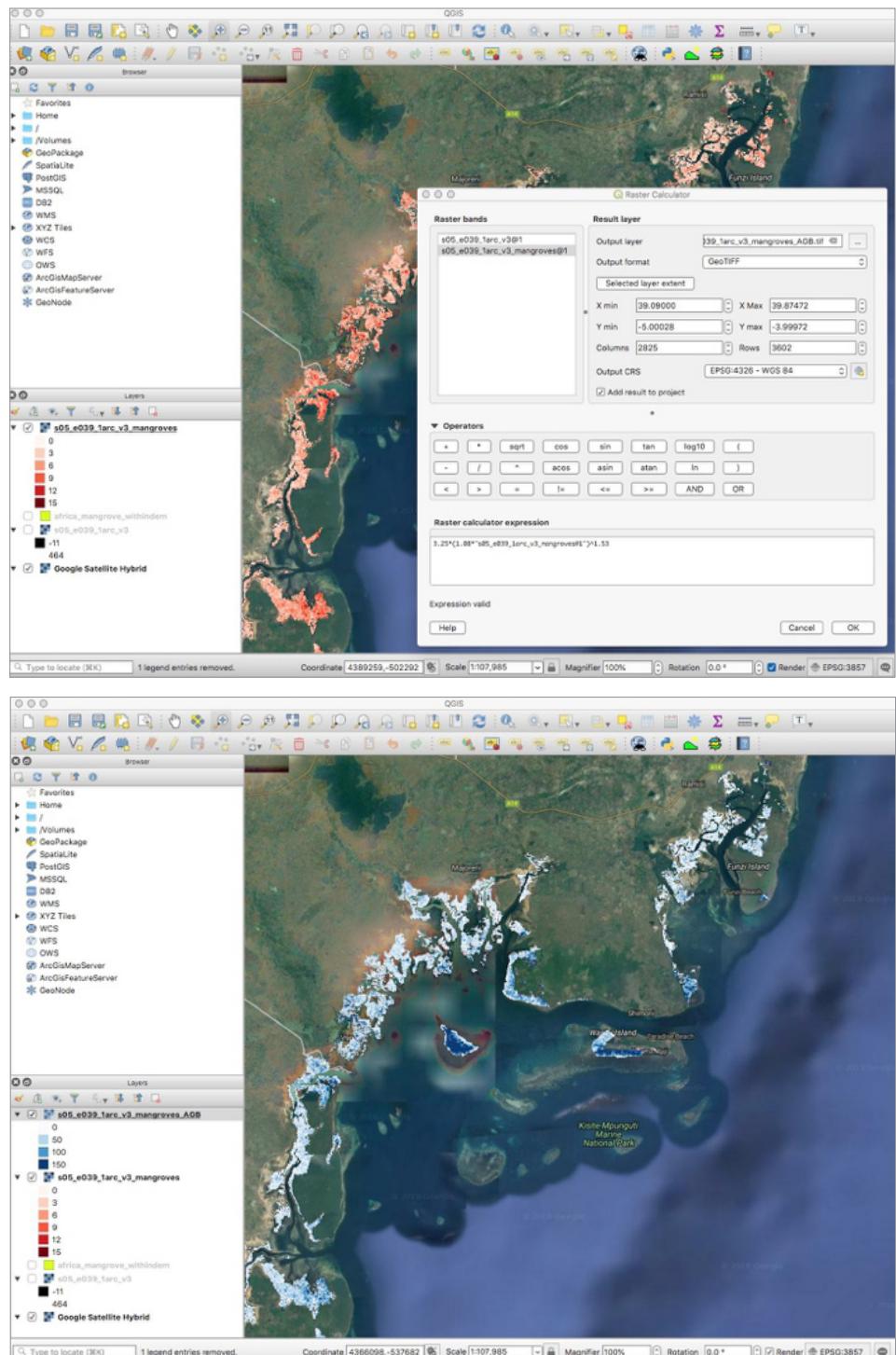
- Define the output layer filename, location and format (Geotiff is a favorite)
- Enter the allometric equation for SRTM:  $B=3.25^*H_{ba}^{1.53}$ , where  $H_{ba} = 1.08^*SRTM$ , and the equation becomes  $3.25^*(1.08^*SRTM)^{1.53}$

### 5.4 Applying the Generic Backscatter-Based Allometric Equation

In this section, we will estimate mangrove biomass from radar backscatter. Keep in mind this method does not work very well for mangrove forests, as site-specific allometry may be required. Here are generic (i.e. tailored to tropical forests) equations relating aboveground biomass B in tons per hectare (t/ha) to backscatter in decibels (dB):

- $S^0_{HV}(dB) \sim -22.5 + 3.0 * \ln(B)$
- $S^0_{HV}(\text{linear}) \sim a \cdot B \cdot 1^{\prime} (1 - e^{-B \cdot b}) + c$ , where  
 $a=0.013682$ ,  $b=0.21116$ ,  $b=0.051846$ ,  
 $c=0.02192$  (Yu & Saatchi, 2016). To simplify, use  $B(t/ha) = 0.5 \times 10^{(40.2 \times S^0_{HV})}$   
with  $S^0_{HV}$  in linear units ( $m^2/m^2$ )

Note: There is no definitive relationship for mangroves because of their intermittent inundation and increased root absorption at high biomass levels. However, the logarithmic geometry remains. To apply these equations to the science-ready radar images:



**Figure 1.16** The results of performing band math within QGIS to obtain estimates of mangrove biomass.

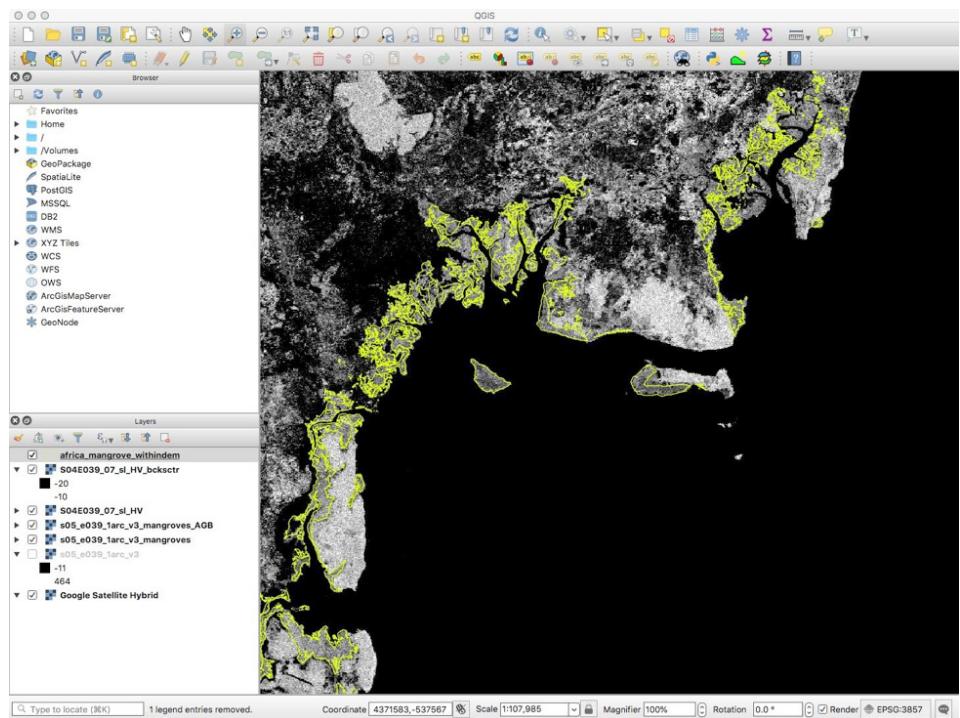
- Upload either your SNAP-processed radar images (in dB), or the ALOS-PALSAR mosaic (found in the data, zip folder provided for this training: ALOS\_Mosaic-20190325T215536Z-001) into QGIS (**Figure 1.17**).

Note: For simplicity and efficiency, let's use the JAXA ALOS mosaics for this exercise. Remember to convert the mosaics' digital number (DN) value to dB using the following JAXA-provided formula:

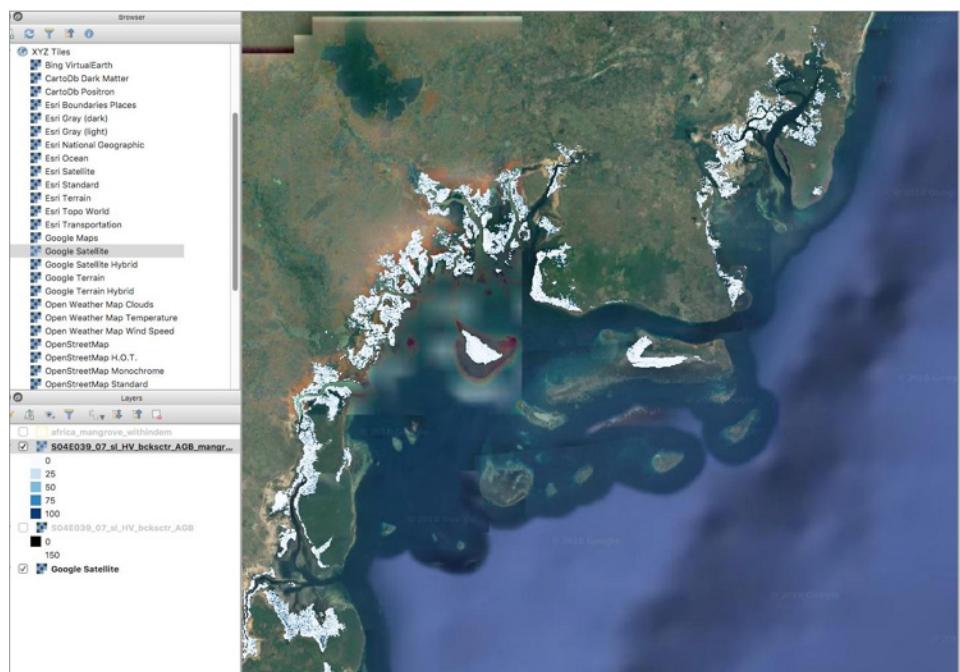
$$S^o_{HV}(\text{dB}) = 20 * \log_{10}(\text{DN}) - 83$$

- In the top menu, click "Raster", then "Raster Calculator".
- Enter the formula in "Raster Calculator Expression". Double-click on the band name to insert an expression.
- The JAXA formula to convert DN to dB results in images that are in dB. However, the allometric equation that we used is in linear units. Always be aware of units used, as biomass allometry could also be using dB.
- In top menu, click on "Raster" and then "Raster Calculator". Enter the allometric equation via the "Raster Calculator Expression", i.e.  $0.5 * 10(40.2 * \text{RasterImage})$  with "RasterImage" in linear units ( $\text{m}^2/\text{m}^2$ ). That means that if using a dB image, the equations become:  $0.5 * 10 * (40.2 * 10^{(\text{RasterImage}/10)})$
- Use the same mangrove shapefile you used in the SRTM DEM portion of the exercise to extract the mangrove extent from the image. Use an older version of QGIS (currently not working in 3.0 and 3.2) or copy/paste the following into your terminal window to obtain **Figure 1.18**:

```
gdalwarp -ot Float32 -of GTiff -tr
0.0002222 -0.0002222 -tap -cut-
line /DirectoryWherePolygonLocat-
ed/africa_mangrove_withinDEM.shp
-crop_to_cutline /DirectoryWhere-
BiomassTransformedImageIsLocated/
SO4E039_07_sl_HV_bcksctr_AGB.tif /
YourOutputDirectory/SO4E039_07_
sl_HV_bcksctr_AGB_mangroves.tif
```



**Figure 1.17** ALOS-1 JAXA mosaic using HV polarization.



**Figure 1.18** Aboveground biomass obtained from the backscatter curve. Clearly, the retrieved biomass is much lower than that obtained with SRTM. Backscatter is known to be limited in mangrove forest, changing with water level and mangrove types, sometimes decreasing at biomass above 100t/ha in red mangroves. You may have to derive your own site-specific allometric equation relating biomass to either height or backscatter (See the next section for further information).

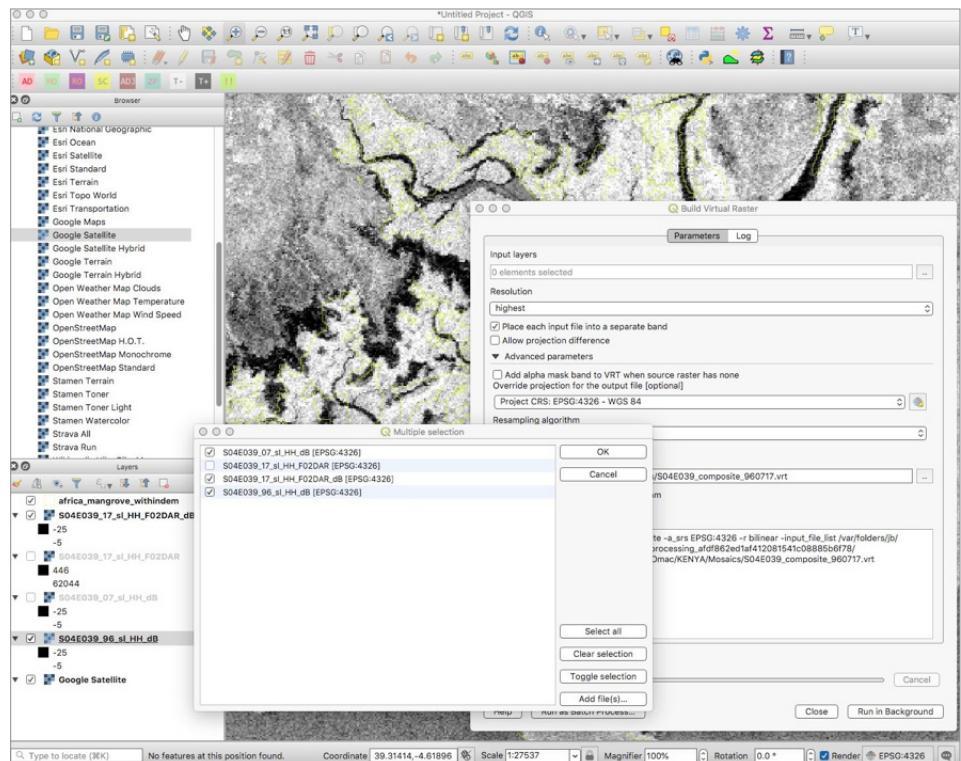
## 5.5 Deriving Your Own Allometric Equations

To derive site-specific allometric equations relating canopy height or radar backscatter to in situ biomass, we must obtain field data with latitude, longitude and biomass values, which can be formatted into columns of a .csv (comma-separated values) delimited text file. Once the data is formatted as a .csv, it can be taken into QGIS to extract radar image or interferometric height values:

1. Save the data as a .csv file.
2. Read into QGIS: In the top menu: click “Layer” > “Add Layer” > “Add Delimited Text Layer”. Then navigate to your CSV file and follow the instructions. You may need to specify the projection in “Geometry”.
3. If you have installed the QGIS “Point Sampling Tool” plugin, click on its icon and select your CSV layer.
4. Select the attributes you want to preserve in the output file. These must include, for example, biomass and backscatter.
5. Define your “Output point vector layer”. Via the pop-up window, select “Comma Separated Values (\*.csv)” as the output format.
6. Open your new CSV file in Excel and fit an equation.
7. In QGIS, apply the new equation to the radar backscatter image using the “Raster Calculator”, as explained in section 5.4.
8. At this point, you can independently generate your own biomass map.
9. Using this same method, you can also validate your biomass maps using available field data.

## 5.6 Monitoring Mangrove Forest Loss and Gain

While radar backscatter is not the best variable for estimating biomass in mangrove forests, one can use it to monitor mangrove forest loss and gain over time. To do this, you need to open several images into QGIS and use either band math to subtract images and detect change, or a 3-image color composite to visualize changes. Let’s try the latter first:



**Figure 1.19** Generating a “virtual raster” from 3 bands to make a color composite image representing change over time.

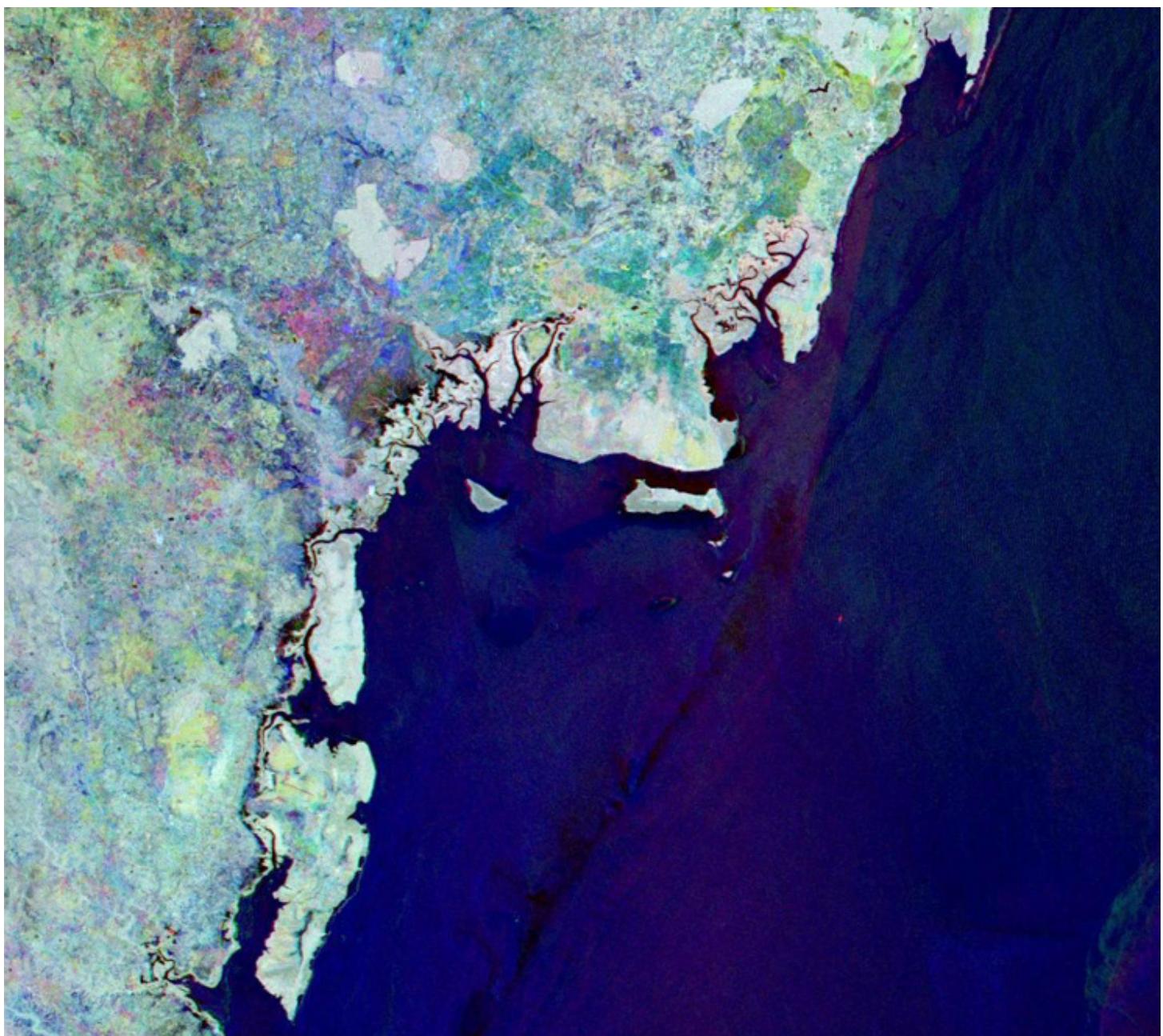
1. Load 3 images from different times into QGIS, e.g. JERS-1 (1996), ALOS-1(2007) and ALOS-2 (2017) mosaic tiles.
2. From the top menu, click “Raster” > “Miscellaneous” > “Build Virtual Raster”. A window pops-up (See **Figure 1.19**)
3. Select the output parameters, for example, “resolution”=highest. In the “Advanced parameters” tab, select “Bilinear” (note that this may decrease resolution slightly). Select bands by clicking on “Input layers”.

Note the default band order. To change the order, use the temporary file called “buildvrtInputFiles.txt” generated by QGIS, which is shown in the “GDAL/OGR console call” frame. Copy and edit that file. You can relaunch the virtual raster from the command line using your newly-edited file.

4. Be sure to select the “Place each input file into a separate band” box, and click “Run”.

You can load the virtual raster into QGIS and change

colors using the “Properties” menu by right clicking on the virtual raster. Then, set all bands to the same range of values. In dB, this should be around -20 and -5, and -25 and -10 for HH and HV respectively. You now have a picture where colors represent change between images (**Figure 1.20**). If red, change occurred between the first input band and the others. If green, change occurred only for the 2nd band. (i.e. change occurred between band 1 and 2; however, change again may be returning to original land cover between image 2 and 3.) If blue, there may have been some regeneration. Note: because we used HH (1996 mosaic only has HH), the radar signal is strongly impacted by soil moisture and inundations. Overall, the color changes seen here are in agricultural areas and observed changes may be due to crop maturity at the time of imaging. To accurately monitor changes, generate your time-series from either ALOS or Sentinel-1 time-series data via SNAP. You can choose acquisition dates corresponding to plant phenology and climatic trends. We found no significant changes in mangrove forests for the study region. Did you?



**Figure 1.20** RGB image composite of 1997, 2007 and 2017.

## 5.7 Determining change areas with simple thresholds

While a more thorough statistical analysis will provide additional insight on gain and loss uncertainty, the extent of change (e.g. loss) can be determined by a simple threshold. In QGIS, use "Raster" > "Raster Calculator" to compute the difference image between two datasets from different years. We will use ALOS 2007 minus 1996 as a test of a small area of change in a forest. To determine a threshold that will indicate whether an area has changed or not, we can begin with a visual analysis. If you find an area that has changed, you can draw a polygon around it and used the observed statistics of the difference image to determine the appropriate threshold.

To draw the polygon of observed change, follow these steps:

- In the top menu, click "Layer" > "Create Layer" > "New Shapefile Layer"
- Define the "Geometry Type" as "polygon".
- Define the output filename in "File Name" and in "New Field", use whole numbers. We'll use 1 for change.
- In the "Layers" window, click on this new layer.
- In the top menu, click "Layer" > "Toggle Editing". You can now add features within the file by clicking in the top menu: "Edit" > "Add polygon feature".
- Draw a polygon around observed changes in the difference map with a mouse click on the vertices of the polygon. To finish the polygon geometry, right click. Don't forget to save your changes by clicking "Toggle Editing" again.

To populate the polygons with attribute values containing statistics from the difference image, use "SAGA" in the toolbox. To see the toolbox, click "Processing" in the top menu and then "Toolbox". Navigate to "SAGA" > "Vector-Raster" > "Raster Statistics for polygons".

- Select the difference image for "Grids", then your change polygon for "Polygons", and select your favorite statistical parameters (include mean and standard deviation). You should use these values to select the threshold that deter-

mines if a change has occurred or not.

You can also train areas that have not changed to improve performance and prepare for an accuracy assessment.

- A new polygon band is created with the stats as an attribute. Right click on the new band and select view "Attribute table".

Then, use "Raster" > "Raster Calculator" to make a change map. The equation can, for example, have the following form: ("S04E039\_07minus96@1" < -7.0) and ("S04E039\_07\_sl\_HH\_dB@1" > -15). The logical operators (<, >, and, etc) select all values smaller than -7 (i.e. change) that are above -15 in 2007. The latter condition removes some water surfaces when using HH. Optionally, you can build your own water mask and use it instead. To clean up the change map, use the "Majority Filter" found in "SAGA" > "Raster Filter". A filter of radius 1 is sufficient to remove false changes that may result from radar speckle noise. Try again, but filter the radar images with the "Multidirectional Lee Filter" found in "SAGA" > "Raster Filter".

**Figure 1.19** shows the resulting change map overlaid on Google Earth imagery. In this case, a road seems to have been constructed between 1996 and 2007, which may have led to further forest disturbance after 2007.

## 6 TRAINING MODULE SUMMARY

After completing this training module, you have successfully:

1. Processed raw level 1.1 ALOS images using the ESA Sentinel toolbox (SNAP):

- Opened, displayed, and calibrated images
- Radiometric terrain flattening
- Ground registration
- Performed image filtering

2. Identified archives of radar datasets:

• ASF (ALOS and Sentinel), JAXA (ALOS mosaics), NASA/USGS (SRTM DEM, etc)

3. Performed data processing and analysis in QGIS:

- Cut images to mangrove polygons



**Figure 1.19** Land use change (disturbance map) produced with threshold calculations on time-series imagery.

- Manually generated polygons delineating features of interest
- Derived statistics to populate polygon attributes
- Converted radar images to dB
- Filtered raster images
- Performed computations with raster layers
- Displayed resultant color-composite images

### 4. Estimated biomass

- Used SRTM DEM to estimate canopy height and biomass of mangrove forests (Note: You can also use DEMs from other sensors, such as TanDEM-X)
- Used ALOS backscatter to estimate mangrove biomass (with limited accuracy)

### 5. Used radar backscatter from different time periods (e.g. 1996, 2007 and 2017) to map land cover change

- We did not find significant loss in the mangrove forests examined, but did test the methodology on inland forests





**DR. HANS-ERIK ANDERSEN** received the Ph.D. degree in quantitative resource management from the University of Washington, Seattle, WA, USA. He was a Research Scientist with the University of Washington Precision Forestry Cooperative from 2003 to 2006, where he developed applications of airborne lidar and interferometric SAR for forest inventory and wild-fire fuels assessment. He joined the USDA Forest Service Pacific Northwest (PNW) Research Station as a Research Forester in 2006, based in Anchorage, AK, USA, where he worked on the development of multi-level forest inventory designs for remote regions utilizing both field and remote sensing data. Since 2011, he has been a Team Leader of the Vegetation Monitoring and Remote Sensing (VMaRS) team within the PNW Resource Monitoring and Assessment Program, based in Seattle, WA, USA

*Funding for this project was provided by the NASA-SERVIR program. The author would also like to thank reviewers for the help in improving this chapter.*

# CHAPTER 7

## Sampling Designs for SAR-Assisted Forest Biomass Surveys

Hans-Erik Andersen, Research Forester/Vegetation Monitoring and Remote Sensing (VMaRS) Team Leader, USDA Forest Service

### ABSTRACT

Sampling designs that efficiently integrate information from plot data and a variety of remote sensing systems, including spaceborne SAR, are required to support cost-effective monitoring of forest biomass/carbon at regional and global scales. In particular, sampling designs and statistical modelling/estimation frameworks are desired that provide sound, statistically-rigorous assessments of uncertainty and make efficient use of expensive field plot data and more extensive use of less-expensive remotely-sensed information. In addition, these designs should also provide the flexibility to accommodate a variety of field plot configurations and remote sensing data acquisition strategies/resolutions. This chapter discusses several important considerations in quantifying uncertainty in multi-level sampling designs, including both the model-assisted and model-based inferential frameworks, and use simulation to illustrate the statistical properties of the estimators associated with these designs, with the goal of informing the design of forest inventory and monitoring programs in remote regions.

### 7.1 Background

International efforts to reduce carbon emissions from the forest sector have created increased demands on the capabilities of national and regional forest monitoring systems to provide timely, accurate information on forest carbon stocks and changes due to deforestation and degradation (GFOI 2016). At the same time, it is recognized that traditional forest inventory sampling designs, which typically rely heavily on large numbers of field plot measurements distributed over a region, are difficult or impossible to implement in many remote, underdeveloped regions of the world (e.g., high latitudes, tropics) due to logistical complexity and/or high costs. For this reason, there is increasing interest in the development of new sampling designs for the monitoring of forest biomass/carbon that can efficiently utilize the low-cost mapped information on forest structure (biomass/carbon), at the global scale, that is increasingly available with the recent and future launches of several satellite SAR missions, such as Advanced Land Observation Satellite (ALOS) Phased Array type L-band Synthetic Aperture Radar (PALSAR) (Hoekman et al. 2010) and ALOS-2 PALSAR-2 (JAXA 2014). For this reason, sampling designs and statistical modelling/estimation frameworks are increasingly sought with the following properties:

- (1) Provide the basis for sound, statistically-rigorous assessment of uncertainty (e.g., Gregoire et al. 2016)
- (2) Use a fewer number of expensive field plots and more extensive, efficient use of less-expensive remotely-sensed information (including airborne light detection and radar (lidar), satellite-based L-band SAR)
- (3) Provide flexibility to accommodate a variety of field plot configurations and remote sensing data acquisition strategies/resolutions

This chapter discusses several important considerations in the assessment of uncertainty in forest biomass surveys and how these considerations should factor into the design and implementation of a sampling design for biomass inventory and monitoring using L-band spaceborne SAR in remote regions.

#### 7.1.1 SOURCES OF UNCERTAINTY IN A CARBON INVENTORY AND MONITORING PROGRAM

There are three primary sources of variability in the context of a forest carbon inventory and monitoring system: (1) measurement error, (2) modelling error, and (3) sampling error. In making the choice of a field measurement protocol, sampling design, and inferential framework, all three types of errors should be considered. In the context of carbon monitoring pro-

grams, measurement error—or discrepancy between a recorded field measurement and the expected value of the measurement as defined by documented protocol—is often introduced through inadequate training or lack of adherence to protocol. In practice, measurement error is usually assessed and mitigated (if possible) through quality assurance/quality control (QA/QC) procedures (that can be quite costly to implement), and otherwise is assumed to be minimal in comparison to the measurement itself (Gregoire & Valentine 2008).

In the context of forest carbon monitoring using remote sensing, modelling error is introduced in two ways: (1) the use of allometric models to estimate tree-level biomass/carbon using various tree measurements (diameter at breast height, height, etc.), and (2) the use of models relating the remotely-sensed measurement (SAR backscatter, air photo-derived canopy height and cover, etc.) to the plot-level biomass/carbon.

#### 7.1.2 ALLOMETRIC MODELS FOR BIOMASS

Given the difficulty of measuring aboveground tree biomass directly, virtually all carbon monitoring programs rely upon allometric models to convert tree measurements obtained in a forest inventory (e.g., height, stem diameter) to aboveground biomass (or carbon) estimates. Due to the relatively small samples

used to develop these models, and the wide range of variability in wood density and height/diameter relationships across the geographic range of trees, it is widely acknowledged that lack-of-fit in the allometric models used to estimate biomass can contribute significantly to the true overall error budget for carbon monitoring—although national forest inventory programs often do not explicitly account for this error in official reports. While several recent efforts have made progress in improving the quality of allometric models used in national- or regional-scale carbon monitoring programs (Chojnacky et al. 2014, Chave et al. 2014) and the emergence of new technologies, such as terrestrial laser scanning (Calders et al. 2015) hold promise for improving the efficiency of field measurements, uncertainty due to allometric modelling remains the most difficult source of error to account for in large scale carbon monitoring programs (Duncanson et al. 2017).

### 7.1.3 ESTIMATION OF BIOMASS USING SAR

Due to its sensitivity to forest biomass, global coverage, and capability to penetrate cloud cover, L-band satellite radar has been used extensively as an auxiliary source of data to support forest monitoring programs across a range of biomes (Ryan et al. 2011, Hoekman et al. 2010). L-band dual-polarization (HH, HV) backscatter has been shown to be well-correlated with forest biomass up to approximately 150 Mg/ha, lending it particular utility in assessing forest biomass levels in low-biomass forests characteristic of high-latitude boreal forest biome as well as semi-arid, savanna forests of the tropics (Atwood et al. 2014, Tanase et al. 2014). However, it has been noted that generalizing relationships between L-band radar backscatter and biomass across forest types is inadvisable since radar backscatter from a forest scene is a function of numerous forest structural characteristics (stem density, height, stem diameter), as well as other scene properties (soil moisture, slope, etc.) with varying correlation to tree biomass (Woodhouse et al. 2012). Although the L-band backscatter signal saturates at higher biomass levels ( $>150$  Mg/ha), limiting its usefulness as a stand-alone correlate for biomass in high-biomass forests, there is evidence to suggest that including additional forest structure information, perhaps obtained from lidar

or repeat-pass interferometry (Treuhaft and Siqueira 2000), can help to decouple the complex relationships between backscatter and forest structural attributes that can obscure the biomass-backscatter signal at higher biomass levels (Joshi et al. 2017).

Once the measurement protocols and modelling frameworks have been established in a forest carbon monitoring system, the next step is determining the proper sampling design to obtain the required precision for carbon estimators within the limitations of the available resources. Although it is typical to only be able to directly measure trees (and estimate biomass via allometry) on a very limited portion of the landscape—leading the third source of variability in carbon estimates, sampling error—the use of remote sensing provides a means of obtaining a much more comprehensive picture of forest structure across an area of interest. This chapter explores sampling approaches that utilize a combination of field data and auxiliary information—including wall-to-wall satellite SAR imagery and sampled high-resolution (e.g., lidar) in multilevel inventory designs—to estimate support forest monitoring programs.

## 7.2 Use of Remote Sensing to Support Carbon Surveys

### 7.2.1 MODES OF INFERENCE

Traditionally, forest inventory and monitoring programs have been based on the principles of *design-based* inference, where field plots were distributed as a probability sample, and each unit in the population of interest has a positive probability of being selected in a sample. In design-based sampling, the population is considered fixed, and all uncertainty in the estimation of a population parameter (total biomass, volume, etc.) is due to variability between randomly drawn samples from the population. Depending on the objectives of the study or inventory, probabilities of selection can vary across the population to reduce costs or increase the statistical precision of the estimates. For example, in stratified sampling, the units of the population can be grouped into homogeneous strata and the population-level estimate is calculated as a weighted average of the stratum-level estimates with the weights based on

stratum sizes. *Model-assisted* inference is a means of using lower-cost auxiliary data (e.g., maps, imagery, photo plots) and a model describing the relationship between auxiliary measurements and inventory parameters to improve precision of estimates within the design-based inferential paradigm. In model-assisted approaches, data at every level are still collected as probability samples, but the number of field plots required to achieve a given level of precision can be reduced significantly (compared to designs using only field plots) if there is a strong correlation between auxiliary data and inventory parameters.

In contrast, *model-based* inference is usually based on the so-called *superpopulation* model, where each value from an element in the population is considered a realization of a random variable with a specific probability distribution. Therefore, all population-level values (e.g., total or mean biomass, etc.) are also considered random variables. In the model-based inferential paradigm, the uncertainty in the estimation of a population parameter is due to randomness in the values observed for each population element. Because the validity of inferences in the model-based paradigm are not dependent upon a random (probability) sample, it can be applied in situations where collecting a sufficiently large probability sample of field plots is either too expensive or logistically difficult, such as estimation within small areas or remote regions lacking transportation infrastructure.

Due to very different underlying assumptions, the results from model-based and model-assisted approaches are difficult to compare directly. The advantage of model-based approaches is that there is no requirement that the field plots be a probability sample, while this is a requirement of model-assisted approaches. However, inferences in the model-based context are conditional on the model and may produce severely biased estimators in cases where the model is developed using an unrepresentative sample. In contrast, design-based (including model-assisted) estimators can, from a practical standpoint, be considered unbiased (for reasonably large sample sizes) regardless of the model that is used. Obviously, in a regional or national forest inventory and monitoring context—where estimates are often used to

INTERFERENCE TYPE	DESCRIPTION	STRENGTHS	WEAKNESSES	APPLICATION
<b>Design-based</b>	All data collected as a probability sample; Population is considered fixed; uncertainty is due to variability between randomly-drawn samples	Simple, well-documented designs and formulae for point and variance estimators; design-unbiased estimation; reliable confidence intervals	Requirement of probability sample may be logistically infeasible to cost-prohibitive in some cases; Less efficient if strongly-correlated auxiliary data is available	National forest inventories, National Greenhouse Gas inventories
<b>Model-assisted</b>	Uses lower-cost auxiliary data and models to improve precision of estimates within the design-based paradigm; data at every level are still collected as probability samples	Increased efficiency (fewer field plots for given level of precision) and lower cost if there is a strong correlation between auxiliary data and inventory parameters	Probability samples required at every level of the design; Form of estimators potentially very complex; Only design-unbiased for large samples; confidence intervals less reliable for small samples	REDD+ applications, NFI in remote regions
<b>Model-based</b>	Population values and parameters are random variables. Uncertainty due to randomness in the values observed for each population element.	Probability sampling not required; potentially much less expensive to implement than design-based approaches	Not design-unbiased; Estimators based on models developed with unrepresentative samples can be severely biased	Small area estimation; Tactical forest management; Inventory over large, remote regions lacking transportation infrastructure

**Table 7.1** Description of strengths, weaknesses and main applications of the models addressed in this chapter.

support forest policy decisions and fulfill Reducing Emissions from Deforestation and forest Degradation (REDD+) and Net Green House Gas (NGHG) monitoring and reporting requirements—the quality of unbiased data is critical and the model-assisted approach may be more appropriate. Model-based approaches may be more appropriate for assessment of remote, or small, inadequately sampled areas, or to support tactical-level forest management decisions.

## 7.3 Exercise 1: Simulating an Artificial Population

Simulation can be a useful approach to gain insight into the statistical properties of various survey estimators, especially in the case of somewhat complex, multi-level sampling designs (Ene et al. 2016, Saarela et al. 2017). Here, simulation implemented in the R statistical software package is used to demonstrate the implementation of several SAR-assisted, multi-level sampling designs. Proficiency in R programming is not required to carry out the exercises, since the scripts can be run by simply copying and pasting the code at the R command line.

When generating a simulated population, it is desirable to include realistic correlations between the response variable (e.g., biomass) and the predictor variables used in the inventory. While a multivariate normal distribution can be used to model correlation between several variables, it may also be important for the purposes of gaining insight into the properties of the point and variance estimators, as well as implications for sample size and modelling effort, that

these variables have more realistic marginal distributions (gamma, exponential, etc.). A copula function is a useful mathematical tool to simulate a population with specified multivariate correlation structure and marginal distributions (Ene et al. 2012, Nelsen 2006). While an in-depth discussion of copula models is outside the scope of this chapter, they essentially allow for expressing multivariate distributions in terms of their corresponding univariate marginal distributions and a copula function. In this exercise, a copula function is used to simulate a large population where each element has a value for forest/nonforest classification, biomass (Mg/ha), a lidar-based measurement (function of lidar-derived height and cover), and a SAR-based measurement (function of HH and HV backscatter). Realistic marginal distributions and correlation structure between remote sensing measurements and field-based biomass were developed based on an analysis of airborne lidar, SAR, and field biomass data from a site in interior Alaska (Andersen et al. 2013). In order to introduce realistic spatial heterogeneity across the simulated area, a binary random field ( $150 \times 150$  grid cells) generated an image with a realistic simulated spatial distribution of “forest” and “nonforest” areas. The grid cells within the simulated forest/nonforest image were then populated with elements from the simulated population generated using the copula function. In this way, each element in the image had a value for forest/nonforest, biomass, lidar, and SAR, and the simulated population had realistic marginal distributions, correlation structure, and long-range spatial heterogeneity (Figs. 7.1 and 7.2).

### 7.3.1 DESIGN-BASED ESTIMATION

#### 7.3.1.1 Simple Random Sampling

Simple random sampling (SRS) represents the most fundamental type of design-based sampling and is often used as the basis of comparison for more complex sampling designs. Given a probability sample of elements of size  $n$  from a population of size  $N$ , where a forest attribute of interest ( $\gamma$ ) is obtained for each element  $i$ , the SRS estimator of the population mean is given by the sample mean:

$$\hat{\mu}_{SRS} = \hat{\mu}_{ma,1} = \bar{\gamma} = \frac{1}{n} \sum_{j=1}^n \gamma_j \quad (7.1)$$

and the variance estimator is given by

$$\hat{V}(\hat{\mu}_{SRS}) = \frac{1}{n(n-1)} \sum_{i=1}^n (\gamma_i - \bar{\gamma})^2 \quad (7.2)$$

#### 7.3.3 POST-STRATIFICATION

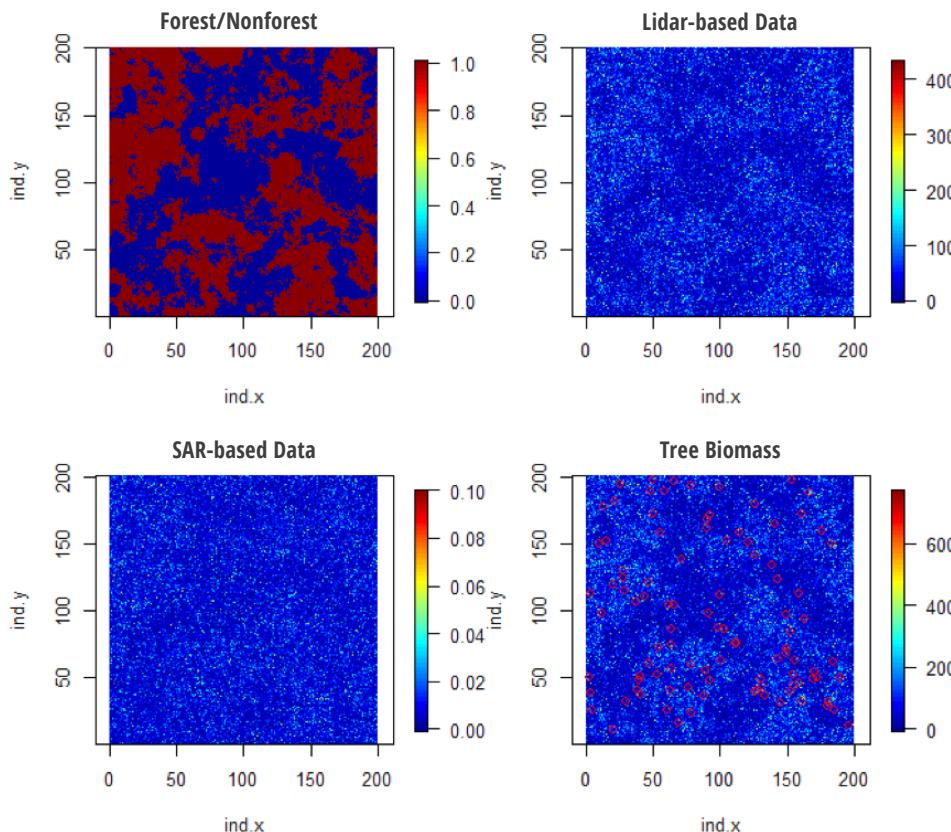
The precision of an SRS estimator can be increased at the estimation stage if the population can be stratified in such a way that plots with similar values for an inventory parameter are grouped together in the same class or stratum, a technique called post-stratification. In post-stratification, the estimator of the population mean is given by

$$\hat{\mu}_{PS} = \sum_h W_h \bar{\gamma}_h \quad (7.3)$$

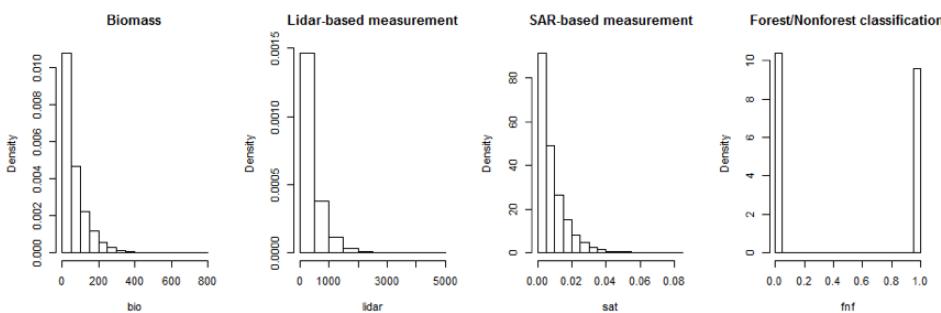
with a variance estimator:

$$\hat{V}(\hat{\mu}_{PS}) = \frac{1}{n} \left( \sum_h W_h n_h V(\bar{\gamma}_h) + \frac{1}{n^2} \sum_h (1-W_h) n_h V(\bar{\gamma}_h) \right), \quad (7.4)$$

where  $W_h$  is the proportion of the population in stratum  $h$  (i.e.,  $W_h = \frac{n_h}{N}$ ) and  $V(\bar{\gamma}_h)$  is the variance of the mean of plots in stratum  $h$ .



**Figure 7.1** Simulated population with biomass, forest/nonforest, lidar-based measurements, and SAR-based measurements. Simulated plots (red) are shown overlaid on tree biomass image.



**Figure 7.2** Simulated marginal distributions of biomass, lidar-based measurements, SAR-based measurements, and forest/nonforest classification. Exponential distributions used to model biomass, lidar, and SAR variables; Bernoulli distribution used to model forest/non-forest class.

	Biomass	Lidar-based	SAR-based	Forest/Nonforest
Biomass	1.00	0.88	0.66	0.36
Lidar-based	0.88	1.00	0.56	0.30
SAR-based	0.66	0.56	1.00	0.15
Forest/Nonforest	0.36	0.30	0.15	1.00

**Table 7.2** Correlation matrix for a simulated population

## 7.4 Exercise 2: Properties of Estimators via Simulation

The statistical properties of the various estimators can be assessed using the simulated population developed previously. At each iteration, a simple random sample of  $n$  elements is drawn from the population, and the point estimator  $\hat{\mu}$  and the variance estimator  $V(\hat{\mu})$  are calculated.

Given that we know the actual population mean  $\mu$ , we can then calculate the mean percent bias of the point estimator

$$\left( \hat{\mu}_{\text{iterations}} - \mu \right) / \mu \times 100\% , \quad (7.5)$$

the relative standard error of the point estimator

$$SD(\hat{\mu}_{\text{iterations}}) / \hat{\mu}_{\text{iterations}} \times 100\% , \quad (7.6)$$

and the empirical coverage probability of the 95% confidence interval for the point estimator:

$$\text{Prob}\left( \hat{\mu} - t_{(0.025, n_2-2\rho)} \sqrt{V(\hat{\mu})} < \mu < \hat{\mu} + t_{(0.975, n_2-2\rho)} \sqrt{V(\hat{\mu})} \right) \times 100\% . \quad (7.7)$$

The empirical coverage probability provides an indication of how reliable (i.e., unbiased) the variance estimator is for a parameter. An empirical coverage probability (95% CP) near 95% is an indicator that the 95% confidence intervals (CIs) calculated using this estimator are reliable. Empirical coverage probabilities of less than 95% indicate that the calculated 95% CIs are giving a falsely precise estimate of uncertainty, while coverage probabilities greater than 95% indicate that the 95% CIs obtained from this estimator are overly conservative.

When the SRS estimator is assessed via simulation, the results indicate the increase in precision due to increasing sample size, as well as the improvement in 95% coverage probability with increasing sample size (it is well-documented that variance estimators can be biased for small samples drawn from highly-skewed populations).

When using a forest/non-forest layer for post-stratification of the SRS sample, the precision is increased a small amount.

### 7.4.1 REGRESSION ESTIMATORS

Model-assisted estimators essentially provide a means to use models based on auxiliary data (e.g., remote sensing) to improve inferences within the

		Bias (%)	SE (%)	95% CP
n	25	0.4%	20.1%	92.1%
	50	-0.3%	15.1%	92.4%
	100	-0.3%	10.6%	93.2%
	200	0.4%	7.0%	95.2%

**Table 7.3** Statistical properties (bias, relative standard error, and 95% coverage probability) for SRS estimator (based on 1,000 iterations).

		Bias (%)	SE (%)	95% CP
n	25	0.0%	19.2%	91.5%
	50	-0.5%	13.3%	94.7%
	100	0.2%	9.8%	93.1%
	200	0.3%	6.7%	94.2%

**Table 7.4** Statistical properties (bias, relative standard error, and 95% coverage probability) for post-stratified estimator (based on 1,000 iterations).

design-based inferential framework (McRoberts et al. 2014). In other words, random (probability) sampling at all levels in the design is the basis for all inference. Model-assisted regression estimators are based on a model of the relationship between the forest attribute of interest (e.g., biomass/carbon),  $\gamma$ , and a vector  $X$ , of auxiliary variables, formulated as,

$$\hat{\gamma}_i = f(X_i; \beta) + \varepsilon_i, \quad (7.8)$$

where  $f(X_i; \beta)$  expresses the mean of  $\gamma$  given observation of  $X$ ,  $\beta$  are the parameters to be estimated, and  $\varepsilon_i$  is a random residual term. In practice, the entire population is not observed, but the parameters of the regression relationship  $\hat{\beta}$  based on a sample of the population is estimated. Then this regression model and observed vector of auxiliary variables are used to predict the inventory attribute for a particular unit of the population:

$$\hat{\gamma}_i = f(X_i; \hat{\beta}). \quad (7.9)$$

A regression estimator for the population mean,  $\hat{\mu}_{ma,1}$  when a single source of wall-to-wall auxiliary information (e.g., satellite SAR or spectral imagery) is given by the following expression:

$$\hat{\mu}_{ma,1} = \frac{1}{N} \sum_{j=1}^N \hat{\gamma}_j + \frac{1}{n} \sum_{i=1}^n (\gamma_i - \hat{\gamma}_i), \quad (7.10)$$

where  $N$  is the population size,  $n$  is the sample size, and  $\hat{\gamma}_j$  is obtained from Eq. 7.9 (Särndal et al. 1992). The first right-hand term in this equation is the sum of the model predictions for the entire population, and the second right-hand term is a correction term which, when added to the first term, compensates for model bias. The regression estimator can be expressed in different forms, but the above formulation is the easiest form to interpret in our context, since the model predictions  $\hat{\gamma}_i$  are based on remotely sensed imagery or measurements and the second term is the mean of the residuals observed at the field plots. For  $n$  much smaller than  $N$ , an approximately unbiased estimator of the corresponding variance is formulated as:

$$\hat{V}(\hat{\mu}_{ma,1}) = \frac{1}{n(n-1)} \sum_{i=1}^n (\gamma_i - \hat{\gamma}_i)^2. \quad (7.11)$$

The advantage of the regression estimator over the SRS estimator is that the variance estimator is based on residuals,  $\gamma_i - \hat{\gamma}_i$  rather than differences,  $\gamma_i - \bar{\gamma}$ , between observations and their mean. Therefore, it can be seen that the degree to which the relationship with  $X$  explains variability in  $\gamma$  will determine the gain in precision from using the regression estimator as opposed to the SRS estimator. It should be noted that post-stratification—where population-level strata proportions are used to improve precision of an estimate in the estimation (rather than the design) stage—is a special case of regression estimation where the predictors are categorical variables (for example, satellite image-based landcover classes).

## 7.5 Exercise 3: Simulation-Based Assessment of Model-Assisted Estimator With a Single Source Of Auxiliary Data

The statistical properties of the model-assisted estimator with one source of auxiliary data (assumed to be collected wall-to-wall, such as SAR imagery) and various sample sizes for field plots (**Table 7.5**). It is evident from these results that there is a small reduction in the standard error (in comparison to the SRS estimator) through including a single auxiliary

that is moderately correlated with biomass (Mandalaz et al. 2013). It is noted that the sampling distribution of this variance estimator is bell-shaped, but with heavier tails than a normal distribution. Therefore, this approach was followed and confidence intervals calculated using a student's t-distribution with  $n_2 - 2p$  degrees of freedom.

		Bias (%)	SE (%)	95% CP
n <sub>2</sub>	25	0.3%	16.1%	90.6%
	50	0.4%	11.5%	92.7%
	100	-0.1%	7.8%	93.5%
	200	0.1%	5.5%	94.4%

**Table 7.5** Statistical properties of a model-assisted regression estimator with single-auxiliary (bias, relative standard error, 95% coverage probability) for four different phase-1 sample sizes (250, 500, 1,000, 2,000) and four different phase-2 sample sizes (25, 50, 100, 200), based on 1,000 iterations.

### 7.5.1 MODEL-BASED APPROACHES

Following McRoberts et al. (2010) and Saarela et al. (2016) if  $\gamma$  is the random variable (Above Ground Biomass (AGB)) with a mean  $\mu$  and standard deviation  $\sigma$ , the observed AGB value at the  $i^{\text{th}}$  pixel ( $y_i$ ) can be represented as

$$y_i = \mu_i + \varepsilon_i, \quad (7.12)$$

where  $\varepsilon_i \sim N(0, \sigma^2)$ . The mean AGB at the  $i^{\text{th}}$  pixel is then given by

$$\mu_i = f(X_i; \beta), \quad (7.13)$$

which is estimated by

$$\hat{\mu}_i = f(X_i; \hat{\beta}), \quad (7.14)$$

where  $X_i$  is the lidar-based predictor variable at the  $i^{\text{th}}$  pixel, and  $\hat{\beta}$  is the vector of  $p$  predicted regression coefficients. The model-based estimate of mean AGB over the entire areas is:

$$\mu_U = \mathbf{1}'_U \mathbf{X}_U \hat{\beta}, \quad (7.15)$$

where  $\mathbf{1}'_U$  is an  $N$ -length column vector where every element equals  $1/N$ ,  $\mathbf{X}_U$  is an  $N \times (p+1)$  matrix of satellite auxiliary variables available for each element in the population  $U$ . The variance of the model-based mean

AGB estimate is given by

$$\widehat{\mathbf{V}}(\mu_y) = \mathbf{U}' \mathbf{X} \mathbf{U} \mathbf{V}_\beta \mathbf{X}' \mathbf{U}' \mathbf{U}, \quad (7.16)$$

where  $\mathbf{V}_\beta$  is the variance-covariance matrix for the regression model parameter estimates  $\hat{\beta}$ . For example, in the case of  $p=2$ ,  $\mathbf{V}_\beta$  is given by:

$$\begin{bmatrix} \hat{v}(\hat{\beta}_0) & \widehat{\text{cov}}(\hat{\beta}_0, \hat{\beta}_1) \\ \widehat{\text{cov}}(\hat{\beta}_1, \hat{\beta}_0) & \hat{v}(\hat{\beta}_1) \end{bmatrix}. \quad (7.17)$$

It should be noted that when using internal models developed from an SRS sample at all levels of the sampling design, the model-based estimator will yield virtually the same point estimate and variance estimator as the model-assisted estimator. However, as noted above, the assumptions behind these estimators differ and provide more flexibility in the application of the model-based estimator (e.g., application to nonprobability samples). Care must be taken to ensure that models are based on a representative (if not random) sample to reduce bias in the point and variance estimators (see Exercise 4).

## 7.6 Exercise 4: Simulation-Based Assessment of Model-Based Estimator with One Source Of Auxiliary Data

In order to illustrate the perils of an incorrectly specified model in the context of model-based estimation, in this exercise, the model is developed from a sample selected only from the forested plots within the population, and then used to estimate biomass—using both

model-assisted and model-based estimators—over the entire population. **Table 7.6** indicates that use of an incorrectly specific model (based on an unrepresentative sample) can lead to significant bias in the point estimates (28% in this case), while the model-assisted estimator remains virtually unbiased (0.5%).

		Bias (%)	SE (%)	95% CP
$n_2 = 50$	M-A, 1-aux	0.5%	11.3%	96.8%
	M-B, 1-aux	28.2%	9.2%	39.1%

**Table 7.6** Statistical properties of model-assisted and model-based regression estimators with single-auxiliary (bias, relative standard error, 95% coverage probability) using a mid-range second-phase sample size of 50 (based on 1,000 iterations).

### 7.6.1 SAMPLING DESIGNS WITH MULTIPLE SOURCES OF AUXILIARY DATA

In some cases, two types of auxiliary information are available, where one (e.g., satellite SAR imagery) is collected wall-to-wall and another type of (more expensive and higher resolution) remotely-sensed data is collected in a sampling mode. For example, multi-level sampling design may consist of: (1) a large sample of relatively inexpensive photo-interpreted plots distributed over an area of interest, with (2) detailed, relatively expensive, field measurements of the attribute of interest (e.g., tree biomass/carbon) collected on a subsample of these photo plots, and (3) free, or very inexpensive, satellite image data (SAR) available over the entire area. Depending on application and how the data were collected, this type of multi-level sampling design can

be approached from a model-based or model-assisted inferential standpoint.

#### 7.6.2 MODEL-ASSISTED

Following Mandallaz et al. (2013), a model-assisted estimator of mean aboveground tree biomass can be developed using field plot data and two sources of auxiliary data in the following manner: as in the previous example, (1) a vector  $X_{ij}$  of remote sensing-derived variables that are known for all  $N$  elements in the population ( $U$ ), and (2) a vector  $X_{ij}$  of remote sensing-derived variables that are known only for the elements in the first phase sample of  $n_1$  units, and the inventory attribute of interest,  $Y_{ij}$ , is only measured on a relatively small second-phase subsample of  $n_2$  photo plots. As a specific example, the  $X_{ij}$  variables may represent satellite image data (e.g., SAR HV/HH backscatter, Landsat tasseled cap bands) available wall-to-wall over the entire study area, and the  $X_{ij}$  variables represent photo plot measurements (average tree height, cover, forest type) that are only available at a sample of locations distributed over the area of interest. Regression analysis is used to develop a linear model for predicting biomass from photo-based measurements:

$$Y_{2j} = f(X_{1j}; \beta_1) + \epsilon_{1j}, \quad (7.18)$$

while satellite-derived predictor variables are used to predict biomass using satellite-based measurements:

$$Y_{2j} = f(X_{uj}; \beta_0) + \epsilon_{uj}. \quad (7.19)$$

Again, following Mandallaz et al. (2013), this design yields the following estimator of mean biomass for the study area:

$$\hat{Y}_{ma,2} = \frac{1}{N} \sum_{j=1}^N \hat{Y}_{uj} + \frac{1}{n_1} \sum_{i=1}^{n_1} (\hat{Y}_{ij} - \hat{Y}_{uj}) + \frac{1}{n_2} \sum_{k=1}^{n_2} (\hat{Y}_{kj} - \hat{Y}_{uj}) \quad (7.20)$$

$n_2$	25	$n_1$											
		250			500			1000			2000		
		Bias (%)	SE (%)	95% CP	Bias (%)	SE (%)	95% CP	Bias (%)	SE (%)	95% CP	Bias (%)	SE (%)	95% CP
25	0.5%	11.0%	91.3%	0.0%	10.8%	89.0%	0.5%	10.8%	89.0%	0.4%	10.3%	89.6%	
50	-0.2%	8.3%	91.5%	-0.1%	7.6%	92.8%	0.5%	7.4%	92.0%	0.0%	7.7%	89.9%	
100	0.2%	6.5%	93.3%	0.0%	5.7%	94.3%	0.4%	5.3%	94.0%	0.0%	5.1%	93.5%	
200	0.0%	5.4%	93.7%	0.0%	4.4%	95.2%	0.1%	4.0%	94.4%	-0.1%	3.6%	95.2%	

**Table 7.7** Statistical properties of model-assisted estimator with two auxiliaries (bias, relative standard error, 95% coverage probability) for four different phase 1 sample sizes (250, 500, 1,000, 2,000) and four different phase 2 sample sizes (25, 50, 100, 200). Satellite image-derived measurements were assumed to be available for every unit in the population (based on 1,000 iterations).

		$n_1$											
		250			500			1000			2000		
		Bias (%)	SE (%)	95% CP	Bias (%)	SE (%)	95% CP	Bias (%)	SE (%)	95% CP	Bias (%)	SE (%)	95% CP
$n_2 = 50$	M-A, 2-aux	-0.4%	8.2%	95.6%	-0.2%	7.8%	96.0%	-0.1%	7.6%	95.6%	0.1%	7.3%	95.7%
	M-B, 2-aux	10.8%	7.8%	85.7%	11.2%	7.1%	82.9%	10.7%	7.1%	82.8%	11.5%	6.5%	81.1%

**Table 7.8** Statistical properties of model-assisted and model-based regression estimators with two auxiliaries (bias, relative standard error, 95% coverage probability) using a mid-range second-phase sample size of 50 (based on 1,000 iterations).

with variance estimator:

$$\hat{v}(\hat{\mu}_{ma,2}) = \frac{1}{n_1 n_2} \sum_{j=1}^{n_2} (\hat{Y}_{2j} - \hat{Y}_{1j})^2 + \frac{1}{n^2} \left( 1 - \frac{n_2}{n_1} \right) \sum_{j=1}^{n_2} (\hat{Y}_{2j} - \hat{Y}_v)^2 \quad (7.21)$$

$\hat{Y}_{1k} = \mathbf{x}'_{1k} \hat{\mathbf{B}}_{1s}$  (for  $k \in U$ ) are the satellite image-based predictions at each satellite image pixel (obtained via linear regression).

## 7.7 Exercise 5: Statistical Properties of Model-Assisted Estimators with Two Sources of Auxiliary Data

### 7.7.1 MODEL-BASED

A model-based approach to utilizing auxiliary data collected at multiple levels was developed by Saarela et al. (2016). As in the previous example of model-based estimator, the relationship between the inventory attribute  $Y$ , which is the random variable (AGB) with a mean  $\mu$  and standard deviation  $\sigma$ , the observed mean AGB value at the  $i^{\text{th}}$  pixel ( $y_i$ ) can be represented as:

$$\mu_i = f(X_i; \beta) + \epsilon_i, \quad (7.22)$$

where  $\epsilon_i \sim N(0, \sigma^2)$ . The mean AGB at the  $i^{\text{th}}$  pixel is given by

$$\hat{\mu}_i = f(\mathbf{x}_i; \beta) \quad (7.23)$$

which is estimated by

$$\hat{\mu}_i = f(\mathbf{x}_i; \hat{\beta}) \quad (7.24)$$

where  $X_i$  is the set of lidar-based predictor variables available for the second phase sample  $n_2$  of the population and  $\hat{\beta}$  is the vector of  $p$  predicted regression coefficients. This linear model is used to estimate the mean AGB at every pixel in the first phase sample  $n_1$ :

$$\hat{\mu}_{1i} = f(\mathbf{x}_{1i}; \hat{\beta}). \quad (7.25)$$

In this hierarchical modelling framework, a second model is developed relating the satellite-based predictor variables  $Z_u$  available over the entire population to the  $\hat{\mu}_1$  predictions available within the first phase sample:

$$\hat{\mu}_{1i} = f(Z_{1i}; \alpha_1) + \omega_{1i} \quad (7.26)$$

where  $\omega_{1i} \sim N(0, \sigma^2)$  and  $\alpha_1$  is the vector of model coefficients linking lidar-estimated AGB values and the satellite predictor variables estimated by  $\hat{\alpha}_1$ . The model-based estimate of mean AGB over the entire area is

$$\hat{\mu}_U = \mathbf{z}'_U \mathbf{Z}_U \hat{\alpha}_1. \quad (7.27)$$

The variance of the model-based mean AGB estimate is given by:

$$\mathbf{V}(\hat{\mu}_U) = \mathbf{z}'_U \mathbf{Z}_U \mathbf{V}_{\alpha_1}^{-1} \mathbf{Z}'_U \mathbf{z}, \quad (7.28)$$

where  $\mathbf{V}_{\alpha_1}$  is the variance-covariance matrix for the regression model parameter estimates given by:

$$\mathbf{V}_{\hat{\alpha}} = \frac{\hat{\omega}'_1 \hat{\omega}_1}{\mathbf{M} - \mathbf{q} - 1} (\mathbf{Z}'_1 \mathbf{Z}_1)^{-1} + (\mathbf{Z}'_1 \mathbf{Z}_1)^{-1} \mathbf{z}'_1 [\mathbf{x}_1 \mathbf{V}_{\hat{\beta}} \mathbf{x}'_1] \mathbf{z}_1 (\mathbf{Z}'_1 \mathbf{Z}_1)^{-1}, \quad (7.29)$$

where  $\mathbf{V}_{\hat{\beta}}$  is the variance-covariance matrix for the regression model parameter estimates  $\hat{\beta}$  and  $\hat{\omega}_1 = \mathbf{x}'_1 \hat{\beta} - \mathbf{z}'_1 \hat{\alpha}_1$  is an  $n_1$  length vector of model residuals.

## 7.8 Exercise 6: Statistical Properties of Model-Based Estimators with 2-Auxiliaries with Biased Model

Refer to **Table 7.8**.

## 7.9 Exercise 7: Estimation of Tree Biomass Using Field Plots, Lidar Plots, and SAR

In this exercise, the model-based estimators are applied with two sources of auxiliary data using an actual dataset collected for a region of interior Alaska (USA). The data consist of: (1) estimates of aboveground tree biomass (Mg/ha) collected over relatively sparse sample of field plots ( $n_2 = 30$  1/30<sup>th</sup> ha circular plots), (2) height-based metrics collected over a denser (systematic) sample of lidar plots ( $n_1 = 325$  1/30<sup>th</sup> ha circular plots), and (3) wall-to-wall L-band satellite SAR-derived imagery (see **Fig. 7.3**). Tree height, tree diameter, and species were collected for each tree on the plots, and allometric models were applied to these measurements to estimate tree—and aggregated plot-level biomass (Yarie et al. 2007).

Here, the model-based estimator with one source of auxiliary data (SAR imagery) developed in Exercise 4, and the estimator for two sources of auxiliary data (wall-to-wall L-band SAR backscatter, large sample of lidar plots) developed in Exercise 6 are compared in the estimation of total aboveground tree biomass (**Table 7.9**).

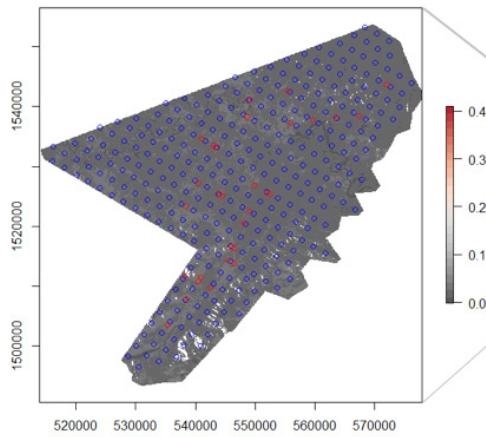
	Mean (Mg/ha)	SE (Mg/ha)
Model-based estimator: SAR only	49.35	9.43
Model-based estimator: SAR and Lidar plots	50.83	7.07

**Table 7.9** Mean biomass estimate for Tok study area (interior Alaska) using field data, lidar plots, and SAR imagery.

## 7.10 Application in Monitoring, Reporting, and Verification (MRV) systems

### 7.10.1 REQUIREMENTS OF REDD+ MRV PROGRAMS

The Intergovernmental Panel on Climate Change (IPCCC) has specified good practice as it pertains to the concept of REDD+ forest monitoring as inventory design that “neither over- nor under-estimates so far as can be judged, and in which uncertainties are reduced as far as is practicable” (GFOI 2016). This guidance essentially promotes the implementation of monitoring programs that maximize precision of estimates, while minimizing bias, within the constraints of available resources. The multi-level estimators for forest biomass present-



**Figure 7.3** L-band radar imagery, lidar plots (blue), and field plots (red) for Tok study area, interior Alaska, U.S.



ed in this chapter provide a range of options for design of carbon monitoring programs, including model-assisted approaches requiring probability samples for all levels of the design that provide design-unbiased estimators and model-based ap-

proaches that may be less expensive to implement due to the lack of requirements for a probability sample, but at the cost of a possibly biased estimator if the model is incorrectly specified.

## 7.11 References

- Andersen, H., J. Strunk, H. Temesgen, D. Atwood, and K. Winterberger. 2012. Using multilevel remote sensing and ground data to estimate forest biomass resources in remote regions: A case study in the boreal forests of interior Alaska. *Canadian Journal of Remote Sensing* 37: 596-611.
- Atwood, D.K., H-E Andersen, B Matthiss, F Holecz. 2014. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing Vol. 7(8): 3262 – 3273.
- Calders, K. G Newnham, A. Burt, S. Murphy, P. Raumonen, M. Herold, M. Kaasalainen. Nondestructive estimates of above-ground biomass using terrestrial laser scanning. *Methods Ecological Evolution* 6(2):198-208.
- Chave, J. et al. 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology* 20(10):3177-90
- Chojnacky, DC, L. Heath, J. Jenkins. 2014. Updated generalized regression biomass equations for North American tree species, *Forestry* 87(1):129-51..
- Duncanson, L. W. Huang, K. Johnson, A. Swatantran, R. McRoberts, and R. Dubayah. 2017. Implications of allometric model selection for county-level biomass mapping. *Carbon Balance and Management* 12:18.
- Ene, L. T., E. Naesset, T. Gobakken, T. G. Gregoire, G. Stahl, R. Nelson. 2012. Assessing the accuracy of regional lidar-based biomass estimation using a simulation approach. *Remote Sensing of Environment* volume 123: 579-592.
- Ene, L., E. Naesset, and T. Gobakken. 2016. Simulation-based assessment of sampling strategies for large-area biomass estimation using wall-to-wall and partial coverage airborne laser scanning surveys. *Remote Sensing of Environment* 176:328-340.
- GFOI 2016, *Integration of remote-sensing and ground-based observations for estimation of emissions and removals of greenhouse gases in forests: Methods and Guidance from the Global Forest Observations Initiative, Edition 2.0*, Food and Agriculture Organization, Rome
- Gregoire, T.G., Næsset, E., McRoberts, R. E.; Ståhl, G. Andersen, H.-E.; Gobakken, T., Ene, L. Nelson, R. 2016. Statistical rigor in LiDAR-assisted estimation of aboveground forest biomass. *Remote Sensing of Environment*. 173: 98-108.
- Gregoire, T. and H. Valentine. 2008. *Sampling strategies for national resources and the environment*. Chapman and Hall/CRC, Boca Raton.
- Hoekman, D. 2010. PALSAR wide-area mapping of Borneo: methodology and map validation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 3, No. 4.
- JAXA. 2014. *ALOS-2/Calibration Result of JAXA Standard Products; Japan Aerospace Exploration Agency, Earth Observation Research Center: Tsukuba, Japan*.
- Joshi, N., E.T.A. Mitchard, M. Brolly, J. Schumacher, A. Fernández-Landa, V. K. Johannsen, M. Marchamalo and R. Fensholt. 2017. Understanding 'saturation' of radar signals over forests. *Scientific Reports* 7:3505.
- McRoberts, R. 2010. Probability and model-based approaches to inference for proportion forest using satellite imagery as ancillary data. *Remote Sensing of Environment* 114:1017-1025.
- McRoberts, R., H.-E. Andersen, and E. Naesset. 2014. Using airborne laser scanning data to support forest sample surveys. In Maltamo, M., E. Naesset, and J. Vauhkonen. 2014. *Forest applications of airborne laser scanning*. Springer: Dordrecht.
- Mandalaz, D., J. Breschan, A. Hill. 2013. New regression estimators in forest inventories with two-phase sampling and partially exhaustive information: a design-based Monte Carlo approach with applications to small-area estimation. *Canadian Journal of Forest Research*, 2013, 43:1023-1031.
- Nelsen, R. B. 2006. *An Introduction to Copulas*. Springer-Verlag New York.
- Ryan, C., T. Hill, E. Woollen, C. Ghee, E. Mitchard, G. Cassells, J. Grace, I. Woodhouse, M. Williams. 2012. Quantifying small-scale deforestation and forest degradation in African woodlands using radar imagery. *Global Change Biology* 18:243-257.
- Saarela, S., Holm, S., Grafström, A., S. Schnell, E. Naesset, T. Gregoire, R. Nelson, and G. Ståhl. 2016. Hierarchical model-based inference for forest inventory utilizing three sources of information. *Annals of Forest Science* 73: 895
- Särndal, C.E, Swensson, B., Wretman, J. 1992. *Model-Assisted Survey Sampling*. Springer-Verlag, New York.
- Tanase, M., R. Panciera, K. Lowell, S. Tian, J. M. Hacker, J. P. Walker. 2014. Airborne multi-temporal L-band polarimetric SAR data for biomass estimation in semi-arid forests *Remote Sensing of Environment* 145:93-104.
- Treuhhaft, R., and P. Siqueira. 2000. Vertical structure of vegetated land surfaces from interferometric and polarimetric radar. *Radio Science* 35(1):141-177.
- Woodhouse, I., E. Mitchard, M. Brolly, D. Maniatis, C. Ryan. 2012. Radar backscatter is not a "direct measure" of biomass. *Nature Climate Change* 2:556-557.
- Yarie, J., E. Kane, and M. Mack. 2007. Aboveground biomass equations for the trees of interior Alaska. *AFES Bulletin* 115, University of Alaska-Fairbanks, Fairbanks, Alaska, USA (<https://www.uaf.edu/files/snre/B115.pdf>) last accessed 12/20/2018.

# **APPENDIX F**

## **Sampling Design for Forest Biomass Surveys – Chapter 7 Training Module**

## INTRODUCTION

Please refer to Chapter 7 of the SAR Handbook for further background for the following exercises and related scripts. Scripts and data used for this tutorial can be found here: <https://bit.ly/2HNZXWu> as well as on the SERVIR global website training page.

Step 1: Download and install R. The download link is available here: <https://ftp.osuosl.org/pub/cran/>

Optionally, install R Studio (free license) as a GUI to develop your scripts from here: <https://www.rstudio.com/products/rstudio/download/#download>

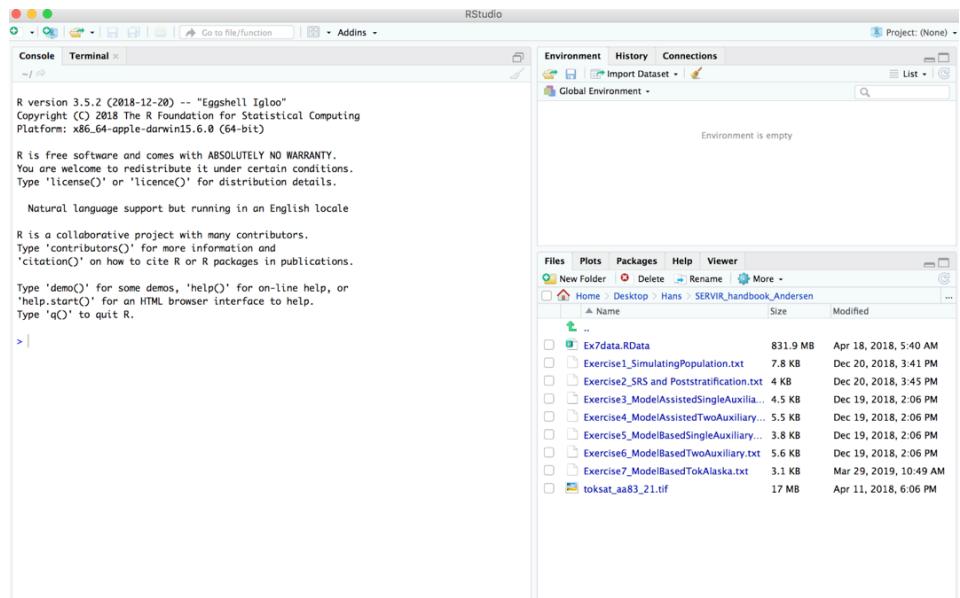
After installation, you will have a platform as shown in **Figure 1.1**. Please note these scripts have been created to run on a Windows operating system. When running them in a Linux or Mac environment, they will need slight modifications to function properly. This document provides guidelines to do those changes.

## EXERCISE 1: Simulating a population

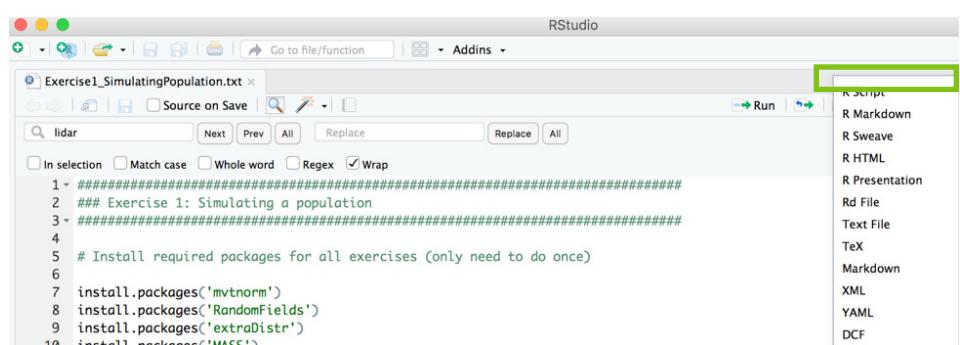
Here we use simulation, implemented in the R statistical software package, to demonstrate the implementation of several SAR-assisted, multi-level sampling designs. Proficiency in R programming is not required to carry out the exercises, since the scripts can be run by simply copying and pasting the code at the R command line. Simulation can be a useful approach to gain insight into the statistical properties of various survey estimators, especially in the case of somewhat complex, multi-level sampling designs (Ene et al., 2016; Saarela et al., 2017).

## Open Script #1:

Double click on the script so it appears in the upper left window in RStudio. Make sure in the editor this file is selected as a R Script (**Figure 1.2**).



**Figure 1.1** The R Studio interface. The section in the lower right allows navigation to the folder containing the scripts for the exercises in this module.



**Figure 1.2** A window displaying options for selecting the file type in R Studio, with "R Script" highlighted.

## Install Packages:

In RStudio, install the following packages by copying and pasting the following script lines:

```
install.packages('mvtnorm')
install.packages('RandomFields')
install.packages('extraDistr')
install.packages('MASS')
install.packages('psych')
install.packages('fields')
install.packages('CompRandFld')
install.packages('raster')
install.packages('RColorBrewer')
```

Note these lines are also included in the Exercise 1 script, and are commented with (#). Make sure all these packages are installed correctly before proceeding.

Modify script appropriately:

In line 249, starting with "save.image", modify where the final file will be saved (the path) and check the file name to which it will be saved; in this case, "servirsim-data.RData" (highlighted section below):

```
# save the simulated population to a RData file - this might take a minute or two...
# this is the file that will be loaded at the beginning of all subsequent exercises
save.image("C:\\\\Users\\\\hander-sen\\\\Desktop\\\\SERVIR _ handbook\\\\servirsimdata.RData")
```

### Run Script:

Then select the rest of the script, starting with “library..” and click on Run (**Figure 1.3**). The purpose of this script is to create and save the simulated data file: “servirsimdata.RData”. This file will be used in subsequent exercises. As final step, check that such file was created. Navigate to the folder where it should be and make sure it exists.

Additional note: If you are working on a Windows computer, the single backslash will work when writing paths in the script, but if you work on a Mac, the double backslash is preferred.

### **EXERCISE 2: Assessment of simple random sampling estimator via simulation**

The statistical properties of the various estimators can be assessed using the simulated population developed previously. In this exercise, at each iteration a simple random sample (SRS) of  $n$  first-phase elements is drawn from the population. The point estimator  $\hat{U}$  and the variance estimator  $\hat{V}(\hat{U})$  are calculated, as well as the coverage probability of the 95% confidence interval of the point estimator. These statistics indicate the bias and precision of the point estimator and the variance estimator.

**Step 1:** Modify this line with the appropriate folder and path for the file created in first script:

```
load("C:\\\\Users\\\\h Andersen\\\\Desktop\\\\Handbook\\\\servirsimdata.RData")
```

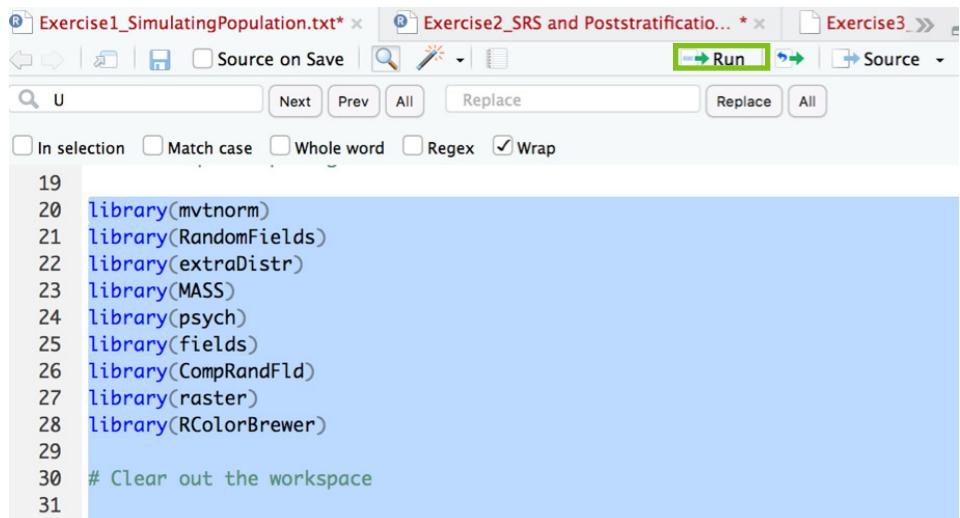
**Step 2:** Make sure the “R Script” is selected on the low right corner (**Figure 1.2**).

**Step 3:** After loading the file, make sure all objects were created. You can enter:

```
<ls()
```

to see all the objects contained in the file.

**Step 4:** If you’re running the script on a Mac, the Progress Bar may not work, as it is a Windows-specific function. Therefore, comment out lines 98, 99 and 129 in this



A screenshot of the R Studio interface. The top menu bar shows three tabs: 'Exercise1\_SimulatingPopulation.txt\*', 'Exercise2\_SRS and Poststratificatio...', and 'Exercise3...'. Below the tabs is a toolbar with icons for 'Run', 'Source', and other options. The main area is a code editor with the following R script:

```

19
20 library(mvtnorm)
21 library(RandomFields)
22 library(extraDistr)
23 library(MASS)
24 library(psych)
25 library(fields)
26 library(CompRandFld)
27 library(raster)
28 library(RColorBrewer)
29
30 # Clear out the workspace
31

```

**Figure 1.3** Screenshot of the R Studio interface with the “Run” option highlighted.

script. Alternately, do not run those lines, as seen below:

```

# update progress bar
#Sys.sleep(0.1)
#setWinProgressBar(pb, t, title=paste(round(t/iterations*100, 0), "% done"))
# close the progress bar
#close(pb)

```

**Step 5:** You should get values and figures, and your console should look like this:

```

> # output results
> print(paste("n: ", n))
[1] "n: 15"
> print(paste("Relative bias for SRS estimator: ", (mean(Muhat_vec.srs)-Mu.U)/Mu.U ))
[1] "Relative bias for SRS estimator: 0.00807804934249307"
> print(paste("Relative SE for SRS estimator: ", sd(Muhat_vec.srs)/mean(Muhat_vec.srs)))
[1] "Relative SE for SRS estimator: 0.273858702223831"
> print(paste("95% coverage prob for SRS estimator: ", sum(cov_vec.srs)/iterations))
[1] "95% coverage prob for SRS estimator: 0.883"
>
> # output results
> print(paste("Relative bias for PS estimator: ", (mean(Muhat_vec.ps)-Mu.U)/Mu.U ))
[1] "Relative bias for PS estimator: 0.00983612308208642"
> print(paste("Relative SE for PS estimator: ", sd(Muhat_vec.ps)/mean(Muhat_vec.ps)))
[1] "Relative SE for PS estimator: 0.279487533308954"
> print(paste("95% coverage prob for PS estimator: ", sum(cov_vec.ps)/iterations))
[1] "95% coverage prob for PS estimator: 0.877"

```

You can use the <ls() command to list all the objects that are now available. See below for a partial list:

```

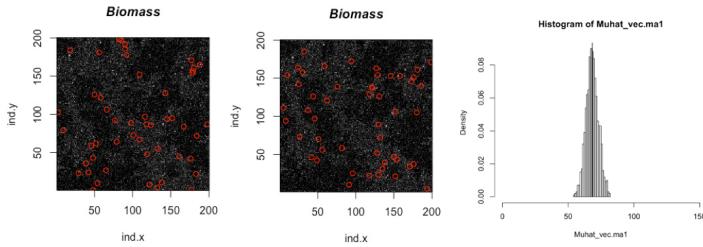
> ls()
[1] "AB"           "bio"          "CI.95perc"    "cov_vec.ps"
[5] "cov_vec.srs"  "fnf"          "hx"           "ind.x"
[9] "ind.y"        "iterations"   "lidar"        "model"
[13] "Mu.U"         "Muhat_vec.ps" "Muhat_vec.srs" "Muhat.ps"
[17] "Muhat.srs"    "n"            "N"            "n.f"
[21] "n.nf"         "p"            "randsamp"    "s"

```

### EXERCISE 3: Assessment of model-assisted estimator with single auxiliary via simulation

In this exercise, we assess the statistical properties of the model-assisted estimator with one source of auxiliary data (assumed to be collected wall-to-wall, such as SAR imagery) and various sample sizes of field plots.

Repeat steps 1 through 4 of Exercise 2. If running on a Mac computer, comment out the script lines for the Progress Bar as shown in the previous section. This will result in the creation of the following figures, shown below:

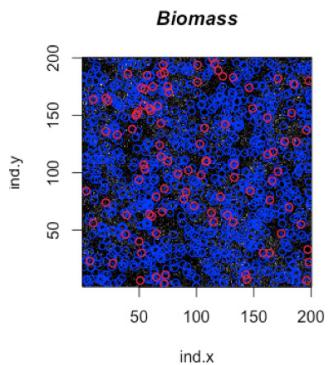


### EXERCISE 4: Assessment of model-assisted estimator with two auxiliaries via simulation

In this exercise, we assess the statistical properties of a model-based estimator of mean biomass. In order to illustrate the perils of an incorrectly specified model in the context of model-based estimation, for this exercise the model is developed from a sample selected only from the forested plots within the population. This is then used to estimate biomass—using both model-assisted and model-based estimators—over the entire population.

Repeat steps 1 and 2 from Exercise 2. If running on a Mac computer, comment out the script lines for the Progress Bar as shown in Exercise 2, step 4.

This will result in the creation of the following figure, shown below:



In addition, you should get the following output results:

```
> # output results
> print(paste("n: ", n))
[1] "n: 50"
> print(paste("Relative bias for SRS estimator: ", (mean(Muhat_vec.srs)-Mu.U)/Mu.U ))
[1] "Relative bias for SRS estimator: 0.00566173451784223"
> print(paste("Relative SE for SRS estimator: ", sd(Muhat_vec.srs)/mean(Muhat_vec.srs)))
[1] "Relative SE for SRS estimator: 0.143681227009878"
> print(paste("95% coverage prob for SRS estimator: ", sum(cov_vec.srs)/iterations))
[1] "95% coverage prob for SRS estimator: 0.946"
>
> print(paste("Relative bias for MA estimator: ", (mean(Muhat_vec.ma1)-Mu.U)/Mu.U ))
[1] "Relative bias for MA estimator: 0.00172107039170955"
> print(paste("Relative SE for MA estimator: ", sd(Muhat_vec.ma1)/mean(Muhat_vec.ma1)))
[1] "Relative SE for MA estimator: 0.0682497752244089"
> print(paste("95% coverage prob for MA estimator: ", sum(cov_vec.ma1)/iterations))
[1] "95% coverage prob for MA estimator: 0.918"
>
```

You should also get the following output results:

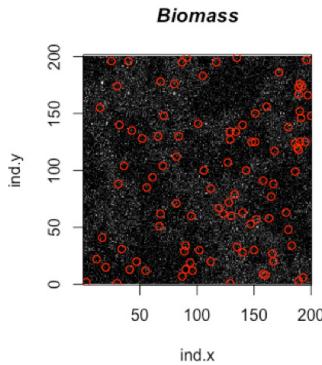
```
> print(paste("n.1: ", n.1))
[1] "n.1: 1000"
> print(paste("n.2: ", n.2))
[1] "n.2: 100"
> print(paste("Relative bias for SRS estimator: ", (mean(Muhat_vec.srs)-Mu.U)/Mu.U ))
[1] "Relative bias for SRS estimator: 0.00772093714715049"
> print(paste("Relative SE for SRS estimator: ", sd(Muhat_vec.srs)/mean(Muhat_vec.srs)))
[1] "Relative SE for SRS estimator: 0.0897102448658417"
> print(paste("95% coverage prob for SRS estimator: ", sum(cov_vec.srs)/iterations))
[1] "95% coverage prob for SRS estimator: 0.99"
>
> print(paste("Relative bias for MA estimator (1 aux): ", (mean(Muhat_vec.ma1)-Mu.U)/Mu.U ))
[1] "Relative bias for MA estimator (1 aux): 0.00573356140916007"
> print(paste("Relative SE for MA estimator (1 aux): ", sd(Muhat_vec.ma1)/mean(Muhat_vec.ma1)))
[1] "Relative SE for MA estimator (1 aux): 0.0770069221725778"
> print(paste("95% coverage prob for MA estimator (1 aux): ", sum(cov_vec.ma1)/iterations))
[1] "95% coverage prob for MA estimator (1 aux): 0.95"
>
> print(paste("Relative bias for MA estimator (2 aux): ", (mean(Muhat_vec.ma2)-Mu.U)/Mu.U ))
[1] "Relative bias for MA estimator (2 aux): 0.00587626584655108"
> print(paste("Relative SE for MA estimator (2 aux): ", sd(Muhat_vec.ma2)/mean(Muhat_vec.ma2)))
[1] "Relative SE for MA estimator (2 aux): 0.048642993404468"
> print(paste("95% coverage prob for MA estimator (2 aux): ", sum(cov_vec.ma2)/iterations))
[1] "95% coverage prob for MA estimator (2 aux): 0.93"
```

## EXERCISE 5: Assessment of model-based estimator with single auxiliary via simulation

In this exercise, we assess the statistical properties of the model-assisted estimator with two sources of auxiliary data (e.g. field plots, photo plots, and wall-to-wall SAR imagery). This estimator does not assume that the model is correct, but does assume that the underlying sample of field and/or remote sensing data is collected via a probability sample.

Repeat steps 1 and 2 of Exercise 2. If running on a Mac computer, comment out the script lines for the Progress Bar as shown in Exercise 2, step 4.

The following plot should be created:

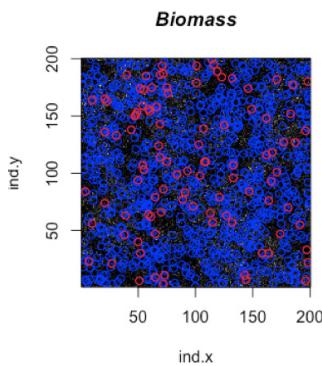


## EXERCISE 6: Model-based estimation with two auxiliaries

In this exercise, we assess the statistical properties of the model-based estimator with two sources of auxiliary data (e.g. field plots, photo plots, and wall-to-wall SAR imagery). This estimator assumes that the model is correct (unbiased), but does not assume that the underlying sample of field and/or remote sensing data is collected via a probability sample.

Repeat steps 1 and 2 of Exercise 2. If running on a Mac computer, comment out the script lines for the Progress Bar as shown in Exercise 2, step 4.

This will result in the creation of the following figure, shown below:



And the console should show output results as follows:

```
> print(paste("n: ", n))
[1] "n: 100"
> print(paste("Relative bias for SRS estimator: ", (mean(Muhat_vec.srs)-Mu.U)/Mu.U ))
[1] "Relative bias for SRS estimator: -0.00384690540648969"
> print(paste("Relative SE for SRS estimator: ", sd(Muhat_vec.srs)/mean(Muhat_vec.srs)))
[1] "Relative SE for SRS estimator: 0.10153260962062"
> print(paste("95% coverage prob for SRS estimator: ", sum(cov_vec.srs)/iterations))
[1] "95% coverage prob for SRS estimator: 0.946"
>
> print(paste("Relative bias for MB estimator (1 aux): ", (mean(Muhat_vec.mb1)-Mu.U)/Mu.U ))
[1] "Relative bias for MB estimator (1 aux): 0.00280324872056235"
> print(paste("Relative SE for MB estimator (1 aux): ", sd(Muhat_vec.mb1)/mean(Muhat_vec.mb1)))
[1] "Relative SE for MB estimator (1 aux): 0.0779438850708484"
> print(paste("95% coverage prob for MB estimator (1 aux): ", sum(cov_vec.mb1)/iterations))
[1] "95% coverage prob for MB estimator (1 aux): 0.943"
```

You should also get the following output results:

```
> print(paste("n.1: ", n.1))
[1] "n.1: 1000"
> print(paste("n.2: ", n.2))
[1] "n.2: 100"
> print(paste("relative bias srs: ", (mean(Muhat_vec.srs)-Mu.U)/Mu.U ))
[1] "relative bias srs: 0.00772093714715049"
> print(paste("relative se srs: ", sd(Muhat_vec.srs)/mean(Muhat_vec.srs)))
[1] "relative se srs: 0.0897102448658417"
> print(paste("coverage prob srs: ", sum(cov_vec.srs)/iterations))
[1] "coverage prob srs: 0.99"
>
> print(paste("relative bias mb1: ", (mean(Muhat_vec.mb1)-Mu.U)/Mu.U ))
[1] "relative bias mb1: 0.00573356140916007"
> print(paste("relative se mb1: ", sd(Muhat_vec.mb1)/mean(Muhat_vec.mb1)))
[1] "relative se mb1: 0.0770069221725777"
> print(paste("coverage prob mb1: ", sum(cov_vec.mb1)/iterations))
[1] "coverage prob mb1: 0.95"
>
> print(paste("relative bias mb2: ", (mean(Muhat_vec.mb2)-Mu.U)/Mu.U ))
[1] "relative bias mb2: 0.00494724740208265"
> print(paste("relative se mb2: ", sd(Muhat_vec.mb2)/mean(Muhat_vec.mb2)))
[1] "relative se mb2: 0.0502598157306141"
> print(paste("coverage prob mb2: ", sum(cov_vec.mb2)/iterations))
[1] "coverage prob mb2: 0.96"
```

## EXERCISE 7: Model-based estimation with field, lidar plots, and SAR data from Tok, Alaska, USA

In this exercise, we apply the model-based estimators with 2 sources of auxiliary data using an actual dataset collected for a region of interior Alaska (USA). The data consist of: 1) estimates of aboveground tree biomass (Mg/ha) collected over relatively sparse sample of field plots ( $n_2= 30$  1/30th ha circular plots), height-based metrics collected over a denser (systematic) sample of lidar plots ( $n_1= 325$  1/30th ha circular plots) and 3) wall-to-wall L-band satellite SAR derived imagery. Tree height, tree diameter and species were collected for each tree on the plots, and allometric models were applied to these measurements to estimate tree-level and aggregated plot-level biomass (Yarie et al., 2007).

Please note that for this exercise, we will use different input datasets provided along with the scripts: "Ex7data.RData" and L-band ALOS PALSAR backscatter image, "toksat\_aa83\_21.tif"

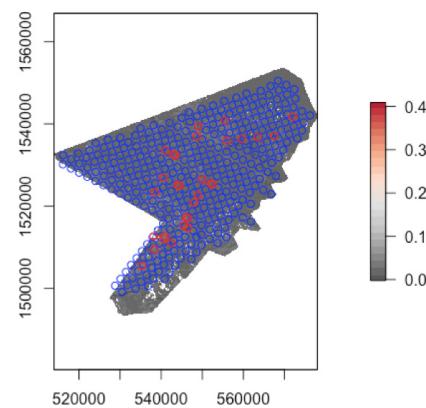
**Step 1:** Ensure the right path is included in the script for these input files:

```
# Load the data into the workspace
load("C:\\\\Users\\\\handersen\\\\Desktop\\\\SER-
VIR _ handbook\\\\Ex7data.RData")
```

**Step 2:** Select the script and run it. In case you run into an error stating "perhaps you need to install rgdal first", install "rgdal" with the command:

```
> install.packages('rgdal')
```

Then, run the script again. The following plot should be generated:



And the console should show output results as follows:

```
[1] "n.1: 325"
> print(paste("n.2: ", n.2))
[1] "n.2: 30"
>
> print(paste("Mean biomass estimate - Model-based w/ SAR only: ", Muhat.mb1))
[1] "Mean biomass estimate - Model-based w/ SAR only: 49.3490722208938"
> print(paste("Relative SE - Model-based w/ SAR only: ", SE_Muhat.mb1/Muhat.mb1))
[1] "Relative SE - Model-based w/ SAR only: 0.191143807433193"
>
> print(paste("Mean biomass estimate - Model-based w/ SAR and lidar plots: ", Muhat.mb2))
[1] "Mean biomass estimate - Model-based w/ SAR and lidar plots: 50.827765915292"
> print(paste("Relative SE - Model-based w/ SAR and lidar plots: ", SE_Muhat.mb2/Muhat.mb2))
[1] "Relative SE - Model-based w/ SAR and lidar plots: 0.139095094263627"
```

## REFERENCES

Ene, L., E. Naesset, and T. Gobakken. 2016. Simulation-based assessment of sampling strategies for large-area biomass estimation using wall-to-wall and partial coverage airborne laser scanning surveys. *Remote Sensing of Environment* 176:328-340.

Saarela, S., Holm, S., Grafström, A., S. Schnell, E. Naesset, T. Gregoire, R. Nelson, and G. Ståhl. 2016. Hierarchical model-based inference for forest inventory utilizing three sources of information. *Annals of Forest Science* 73:895

Yarie, J., E. Kane, and M. Mack. 2007. Aboveground biomass equations for the trees of interior Alaska. AFES Bulletin 115, University of Alaska-Fairbanks, Fairbanks, Alaska, USA (<https://www.uaf.edu/files/snre/B115.pdf>) last accessed 12/20/2018.





*We would like to acknowledge the support of various individuals and institutions without whom this work would not have been completed. We acknowledge Dr. Nancy Searby, Program Manager for NASA's Applied Sciences Capacity Building program, who was part of the inspiration for this work. We would also like to thank the U.S. Agency for International Development (USAID) for their support, particularly through the SilvaCarbon partnership. We would like to thank Dr. Julian Aukema for her stewardship of SilvaCarbon activities on behalf of USAID, as well as Jenny Frankel-Reed, SERVIR Chief of Party at USAID, and Drs. Kevin Coffey and Pete Epanchin. Extreme gratitude is also owed to Sylvia Wilson, the technical lead for SilvaCarbon at the U.S. Geological Survey (USGS), who has worked to ensure that forest monitoring capacity is developed across the globe. We also extend our gratitude to the respective Chiefs-of-Party of the four SERVIR Hubs involved in this effort, namely Dr. Robinson Mugo of the Regional Centre for Mapping of Resources for Development (RCMRD), Birendra Bajracharya of the International Centre for Integrated Mountain Development (ICIMOD), Dr. Peeranan Towashiraporn of the Asian Disaster Preparedness Center (ADPC), Marc Dawson of TetraTech/SERVIR-West Africa, and Issifou Alfari of the Agro-meteorology, Hydrology, and Meteorology Center (AGRHYMET) regional center/SERVIR-West Africa. The respective Agreement Officer Representatives and Contracts Officer Representatives for the various USAID regional offices, namely Chihenyo Kangara, Karl Wurster, Dr. Gary Jahn, and Margaret McMorrow, should also be acknowledged.*

# CHAPTER 8

## Perspectives on the Future Application of SAR in Forest and Environmental Monitoring

Emil Cherrington<sup>1,2</sup>, Africa Flores-Anderson<sup>1,2</sup>, Rajesh Bahadur Thapa<sup>3</sup>, Kelsey E. Herndon<sup>1,2</sup>, Anastasia Wahome<sup>4</sup>, Phoebe Oduor<sup>4</sup>, Kenneth Mubea<sup>4</sup>, Edward Ouko<sup>4</sup>, Nguyen Hanh Quyen<sup>5</sup>, David Saah<sup>6,10</sup>, Mamane Bako<sup>7</sup>, Paul Bartel<sup>8</sup>, Foster Mensah<sup>9</sup>, Stella Ofori-Ampofo<sup>9</sup>, Eric Anderson<sup>1,2</sup>, Emily Adams<sup>1,2</sup>, Rebekke Muensch<sup>1,2</sup>, Amanda Markert<sup>1,2</sup>, Andrea Nicolau<sup>1,2</sup>, W. Lee Ellenburg<sup>1,2</sup>, Kel Markert<sup>1,2</sup>, Betsy Hernandez Sandoval<sup>1,2</sup>, Ashutosh Limaye<sup>1</sup>, Robert Griffin<sup>1,2</sup>, Daniel Irwin<sup>1</sup>

### ABSTRACT

To consider how the use of SAR in forest and environmental monitoring is likely to change in the future, it is useful to reflect on how SAR is currently used across a range of applications. As documented in the preceding chapters, earlier research has transitioned to applications ranging from the use of SAR for tracking deforestation to its use in estimating forest height and biomass. The applications landscape is expected to change due to the growing public availability of SAR data to feed applications, with important missions like the NISAR (NASA-ISRO Synthetic Aperture Radar) and the European Space Agency's (ESA's) Biomass mission planned for launch within a five-year horizon from the publication of this handbook. Limitations notwithstanding, other factors expected to further accelerate the application of SAR data include the growing availability of desktop and online platforms with which to process and analyze radar data, and—as evidenced by this handbook—a growing set of resources to build the user community's capacity to apply SAR data. In support of Global Forest Observation Initiative (GFOI) and SilvaCarbon, the global network of hubs of the SERVIR program are also poised as important resource centers to help strengthen capacities within their respective regions to apply SAR for forest monitoring, as well as in other environmental monitoring contexts.

### 8.1 SAR in the Present Forest Monitoring Context

At the close of this handbook, the reader should have acquired a series of skills that, applied, will put one well on the way to practical use of SAR data for forest monitoring. Toward that practical use, the previous chapters have addressed the following topics:

- Preprocessing of raw SAR data to radiometrically terrain corrected products
- Monitoring deforestation and forest degradation
- Mangrove monitoring
- Forest structure estimation
- Biomass estimation
- Sampling design

Taking a synoptic view, this handbook has thus responded to fundamental questions that need to be addressed not only for use of radar remote sensing, but also for other techniques in general, such as:

- What are the characteristics of the data?

- Where can the data be obtained?
- What can be done with the data?
- What tools can be used to process this data and where are they found?
- How can specific products be generated from the data?

—Toward that end, and for the benefit of a general public who might likely not have access to commercially licensed software, this handbook's practical exercises have generally shown how freeware tools such as Alaska Satellite Facility (ASF) MapReady, PolSARPro, SNAP, and Python scripts can be used for generating different types of derived products.

This chapter follows those questions by addressing one last, related query:

- What can be expected in terms of future developments in SAR regarding the forest and broader environmental monitoring contexts?

A major takeaway from the previous chapters was

that SAR provides unique capabilities for forest monitoring, complementing the capabilities provided by optical remote sensing systems. The preceding chapters have likewise reflected on how far spaceborne imaging radar has come since NASA's SeaSat mission in 1978, the development of subsequent research missions through the 1990s and the 2000s, and the launch of the operational Sentinel-1 satellites of the European Commission's Copernicus Programme in 2014. Nevertheless, as Meyer termed it in this book and elsewhere, today represents a "Golden Age of SAR," and the next few years will see the launch of satellite missions that may have significant implications for the science, practices, and policies of monitoring forests. (At the time of the writing of this handbook, [UNAVCO maintains a register of SAR missions](#).) Those implications concern not only the new wavelengths covered by future missions (e.g., P-band SAR missions like Biomass), but also newer imaging techniques and capabilities like SAR tomography (Tomo-SAR).

<sup>1</sup> NASA Marshall Space Flight Center / SERVIR Science Coordination Office, <sup>2</sup> University of Alabama in Huntsville, <sup>3</sup> International Centre for Integrated Mountain Development (ICIMOD) / SERVIR-Hindu Kush Himalaya, <sup>4</sup> Regional Centre for Mapping of Resources for Development (RCMRD) / SERVIR-Eastern & Southern Africa, <sup>5</sup> Asian Disaster Preparedness Center (ADPC) / SERVIR-Mekong, <sup>6</sup> Spatial Informatics Group (SIG) / SERVIR-Mekong, <sup>7</sup> Agro-meteorology, Hydrology, and Meteorology regional center (AGRHYMET) / SERVIR-West Africa, <sup>8</sup> Tetra Tech, Inc. / SERVIR-West Africa, <sup>9</sup> Centre for Remote Sensing and Geographic Information Services (CERSGIS) / SERVIR-West Africa, <sup>10</sup> University of San Francisco

They are also complementary to spaceborne lidar missions like Ice, Cloud, and land Elevation Satellite-2 (ICESat2) and Global Ecosystem Dynamics Investigation (GEDI)—both launched in 2018. This chapter focuses on potential future applications of SAR, and particularly in relation to ongoing work across SERVIR’s global network. To a lesser extent, this handbook addresses research aspects of SAR, but to the extent possible, the focus is on proven applications.

## 8.2 Future SAR Developments

It is expected that future SAR missions—including the continuation of current missions under the European Union’s Copernicus Programme—will impact how forests are monitored. One barrier to countries’ regular use of SAR data for forest monitoring stems from a lack of archived SAR data that coincide with reference periods suggested for international agreements (e.g., 1990, 2000). Moreover, historical SAR imagery tend to have variable acquisition parameters, making comparison difficult. Nevertheless, if future SAR missions are able to expand the observations made under programs such as Copernicus (e.g., same polarizations, acquisition angles, beam modes; the influences of these factors on the SAR remote sensing of forests was also addressed in the second chapter of this handbook), that would greatly facilitate the continued use of those data. (Copernicus’ Sentinel-1A and Sentinel-1B satellites provide for systematic observation of the Earth, by which the places imaged by the twin C-band SAR instruments are acquired every 12 days, with the same beam modes, and with the same imaging angles, facilitating comparison of imagery acquired on different days.) Taking that and other factors into context, the following subsections explore how the future SAR application landscape is likely to be impacted by:

- (1) Expanded **availability of data** (from existing and upcoming missions)
- (2) Desktop and cloud computing **tools** with which to process radar data
- (3) Capacity building efforts

### 8.2.1 AVAILABILITY OF SAR DATA

Some previously restricted SAR datasets are becoming more available (e.g., the open availability of Advanced Land Observation Satellite (ALOS) Phased Array type L-band Synthetic Aperture Radar (PALSAR) data), and more SAR satellite missions have been launched (e.g., the Sentinel-1 satellites). In terms of future SAR missions, [Chapter 1](#) has already catalogued the near-term radar satellites relevant to forest monitoring (i.e., SAOCOM-1B, Radar Doppler Multifunction (RCM), Biomass, NISAR, TanDEM-L). In fact, of the satellite missions previously listed, since the writing of this handbook commenced, the PAZ, NovaSAR-S, and SAOCOM-1A missions have already launched (in February 2018, September 2018, and October 2018, respectively). Beyond noting that there are upcoming G-, L-, S-, and P-band missions, what is salient is that with various missions having “free and open” data policies, these will add to the already growing public archive of radar imagery of the Earth. As noted previously, this provides a significant opportunity for SAR-based and SAR-enhanced monitoring of the Earth’s resources—something particularly unprecedented prior to Sentinel-1. That opportunity, however, also presents practical challenges to image analysts and their geospatial and/or data departments, specifically with regard to processing and storing large volumes of SAR data and derived products. Nevertheless, as was already covered earlier, it bears repeating that resources such as the [Copernicus Open Access Hub](#) and the [ASF’s Vertex portal](#) are crucial resources that provide access to SAR data. It is anticipated that the significance of these portals will only grow over time as “one-stop shops” for acquiring data from historical and ongoing missions. The forthcoming NISAR mission is likewise poised to generate more data than any NASA Earth observing mission. Faced with significant file size and data volume challenges, an initiative “Getting Ready for NISAR” (GRFN) is part of efforts to move NASA Earth observing data and services to the commercial cloud (Blumenfield 2018).

### 8.2.2 PROCESSING OF SAR DATA: DESKTOP COMPUTING

There has been another welcome development relating to software platforms for processing SAR data.

Prior to ten years ago, radar data processing was partly the domain of commercially licensed software packages. In 2008, ASF released its MapReady software platform, which allowed for the conversion of SAR data from particular Committee on Earth Observation Satellites (CEOS) data formats to GeoTIFF format. This lowered the barrier to visualizing SAR imagery by allowing more users to easily visualize SAR products in GIS platforms, something that was already possible for optical datasets like Landsat imagery. Furthermore, around 2009, in parallel to the development of MapReady, ESA commissioned the development of the Next ESA SAR Toolbox (NEST), which was later folded into the Sentinel-1 Toolbox, and which itself later became a part of what is currently the Sentinel Application Platform (SNAP). As it stands, SNAP allows for preprocessing, processing, and visualizing data from a wide range of SAR missions (as well as data from non-SAR missions). In a similar vein to MapReady, SNAP allows for converting radar imagery to GIS-readable formats such as GeoTIFF. Overall, these tools have allowed a larger pool of GIS analysts and optical remote sensing specialists to begin to bridge over to the study of SAR data.

While MapReady and SNAP serve as key examples of what is currently available in terms of openly available desktop software platforms, the evolution of such platforms may also herald what is to come in terms of additional functionalities. The shift from NEST to the Sentinel-1 Toolbox to SNAP, for instance, saw the addition of application-oriented functions. Responding to the potential for wide applications of SAR data, SNAP currently allows for height estimation, change detection, ocean object detection, and oil spill detection. As this chapter is being written, the InSAR Scientific Computing Environment (ISCE) processing package is undergoing further development in preparation for the forthcoming NISAR mission (Rosen et al. 2018).

### 8.2.3 PROCESSING OF SAR DATA: CLOUD COMPUTING

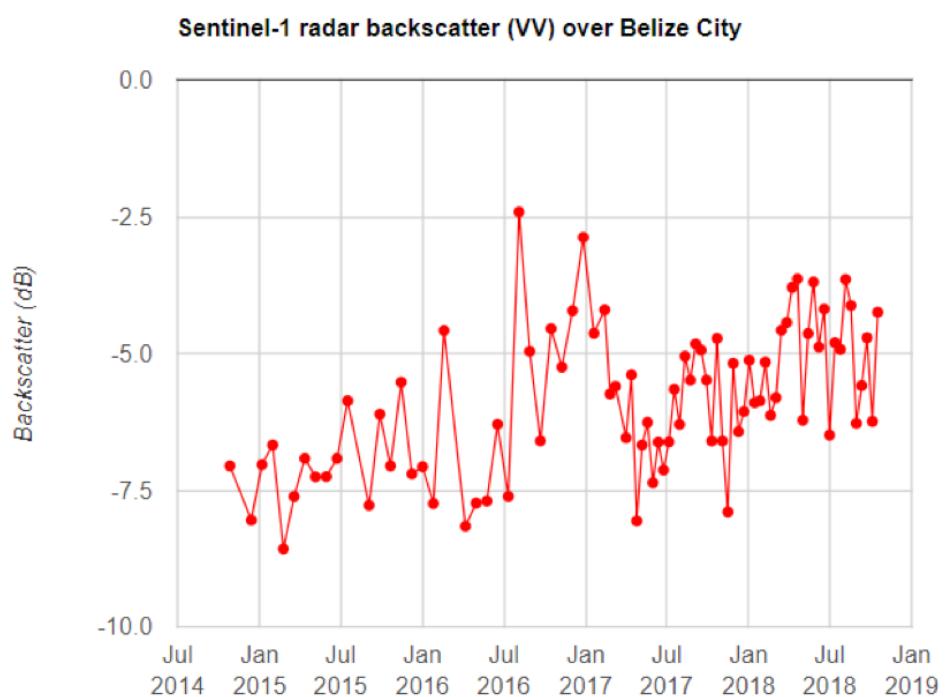
Beyond changes in how desktop applications process and manage SAR data, there has also been a significant shift in how SAR data are processed. Much of the community of users is moving away from downloading large SAR datasets from the Internet and processing them on desktop computer systems,

and is moving toward bulk processing online, “in the cloud.” In only the last few years, a number of web-based platforms are beginning to offer SAR processing and analysis capabilities. These include, but are not limited to:

- ASF’s Hybrid Pluggable Processing Pipeline (HyP3): <http://hyp3.asf.alaska.edu>
- ESA’s Thematic Exploitation Platforms (TEPs): <https://tep.eo.esa.int>
  - There are currently seven individual TEPs, focusing on: coastal areas, food security, forestry, geo-hazards, hydrology, polar areas, and urban areas.
- Google Earth Engine (GEE): <https://earthengine.google.com/>

In the case of GEE, registered users are able to access and analyze the full and continually updated archive of Copernicus Sentinel-1 data, as well as the Japan Aerospace Exploration Agency’s (JAXA’s) Japanese Earth Resources Satellite 1 (JERS-1) and ALOS PALSAR global mosaics. In comparison, ASF HyP3 (pronounced “hype”) allows for more comprehensive processing of SAR data, including pre-processing and the automatic generation of value-added products such as stacked and calibrated image time series, as well as Interferometric SAR (InSAR) products. ASF HyP3 contrasts with GEE, which only provides radiometrically terrain corrected (RTC) amplitude or backscatter datasets, but currently does not support interferometric analyses needed for some forestry applications. Currently, the HyP3 service permits the generation of a limited number of data products per month. Users are encouraged to provide feedback on the functionality of the service and the quality of the provided products. Furthermore, with regard to the upcoming Biomass and NISAR missions, ESA and NASA are collaborating on the development of a Multi-Mission Analysis Platform (MMAP) that will also allow for online processing and analysis of SAR data (Albinet et al. 2018).

ASF HyP3, GEE, and MMAP also highlight the growing significance of the Analysis Ready Data (ARD) concept. With ARD, users are unburdened from some preprocessing and are less likely to waste resources processing extraneous data outside of the specific geographic area of interest. GEE offers, for instance,

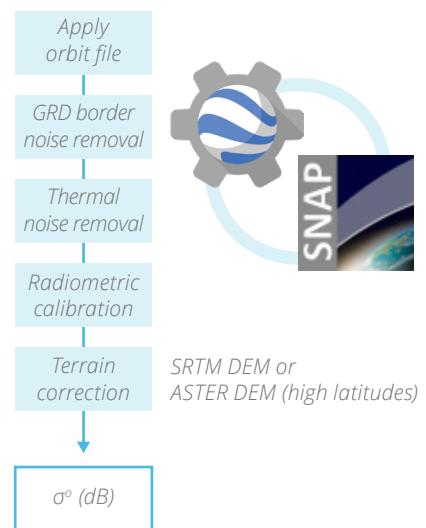


**Figure 8.1** Sentinel-1 backscatter time-series over a part of Belize City, extracted using GEE.

easy extraction and processing of ALOS PALSAR or Sentinel-1 imagery for specific areas of interest, without forcing users to download one or multiple 1-GB scenes to preprocess. It also offers simple functions to calculate time series statistics from the full 4-year archive of Sentinel-1 (S1) imagery in seconds (see Fig. 8.1). In contrast, the workflow to conduct the same process on a desktop or on-premise system (by downloading S1 data from the Copernicus Open Access Hub or from ASF’s Vertex and performing preprocessing, processing, and postprocessing on a software mentioned above), would take significantly more time and bandwidth.

Something to mention are the pre-processing steps GEE uses for the Sentinel-1 data currently available in their platform. At the time of writing this handbook, GEE uses SNAP and, as explained in Chapter 3, there are certain tradeoffs about using SNAP for SAR processing, particularly for the displacement issue observed in Sentinel-1 data and discussed in Chapter 3. This has particular implications in time series analyses, where it can affect change detection. The following is a flowchart of the preprocessing steps GEE uses for their Sentinel-1 data.

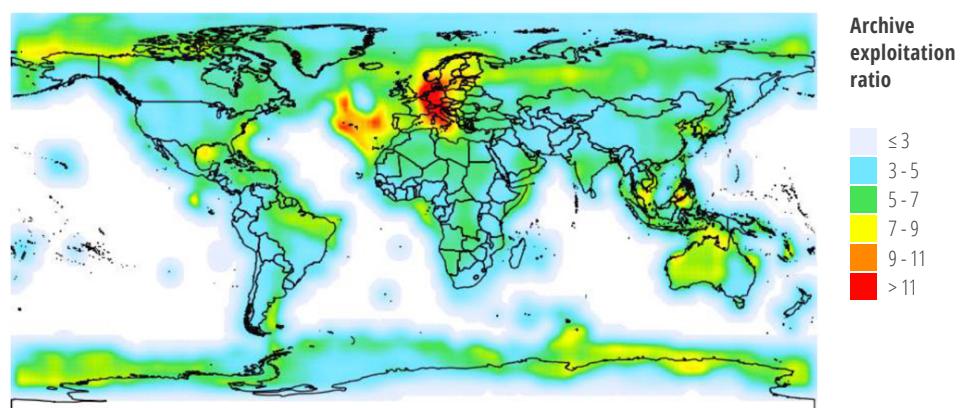
Sentinel-1 imagery in GEE consists of Ground



**Figure 8.2** Sentinel-1 preprocessing steps used by GEE following *Sentinel-1 Toolbox* to derive the backscatter coefficient,  $\sigma^0$ , in each pixel. Last updated August 15, 2018 (source: <https://developers.google.com/earth-engine/sentinel1>)

Range Detected (GRD) imagery processed to backscatter coefficient,  $\sigma^0$ , in decibels. As implemented by the Sentinel-1 Toolbox (<https://sentinel.esa.int/web/sentinel/toolboxes/sentinel-1>), the steps are described as: (1) orbit metadata are updated with an orbit file, (2) low-intensity noise and invalid data on scene edges are removed, (3) additive noise in sub-swaths for scenes in multi-swath acquisition modes are removed, (4) backscatter intensity is computed using sensor calibration parameters in the GRD metadata, and (5) takes into account terrain factor by converting to  $\sigma^0$ , representing target backscattering area (radar cross-section) per unit ground area in decibels. GEE is in the process of reprocessing the Sentinel-1 data to keep all floats unchanged, since the current preprocessing steps convert float32 values to 2-byte unsigned integers, keeping only the 99<sup>th</sup> percentile values.

Application-specific tradeoffs must be considered when deciding to perform analyses on-premise or on-cloud. For instance, how critical is it to control each pre-processing step, or is it acceptable to yield some choice to a free, on-cloud platform? Accuracy, precision, data volume (and bandwidth), and data latency are all considerations. Here, one scenario is discussed that considers data volume and time-series analysis. For instance, if one wanted to look at radar backscatter patterns over Belize City, Belize, between October 26, 2014, and October 17, 2018, 84 descending mode images were acquired in relative orbit no. 128. At roughly 1 GB per GRD scene, analyzing data only over Belize City would still require downloading and subsequently analyzing about 84 GB of raw data. That would certainly be a challenge using a desktop computer in a bandwidth-limited environment. Even if downloading such a volume of data were not a challenge, one can imagine how long it would take to process the 84 scenes from GRD to RTC products, and how much space would be required for the various intermediate steps. Furthermore, the hypothetical case provided pertains only to one scene in one particular mode, only over a 4-year period. If one were to include both ascending and descending data, and choose a study area (e.g., an entire country) that covered multiple scenes, and if the study period were extended (e.g., imagine processing S1 imagery in 2022), the data would become even more difficult to manage and process using desktop computing. This hypothetical



**Figure 8.3** Map of density of Sentinel-1 data downloads (source: [https://scihub.copernicus.eu/twiki/pub/SciHubWebPortal/AnnualReport2017/COPE-SERCO-RP-17-0186\\_-\\_Sentinel\\_Data\\_Access\\_Annual\\_Report\\_2017-Final\\_v1.4.1.pdf](https://scihub.copernicus.eu/twiki/pub/SciHubWebPortal/AnnualReport2017/COPE-SERCO-RP-17-0186_-_Sentinel_Data_Access_Annual_Report_2017-Final_v1.4.1.pdf)).

case merely highlights the quandary that many application-oriented users seeking to exploit SAR data will and have encountered, and how cloud computing can help circumvent that challenge. There is a growing interest in machine learning (ML) in this field, as demonstrated in many recent examples that use on-premise computing (e.g., Chen et al. 2018, Ndkumana et al. 2018, Belgui et al. 2016). Given challenges discussed in this section, the barrier to enter into ML applications in remote sensing and environmental monitoring applications may be lower in cloud environments (e.g., Hird et al. 2017, Shelestov et al. 2017).

### 8.2.3.1 System for Earth Observation Data Access, Processing and Analysis for Land Monitoring (SEPAL) and Food and Agricultural Organization (FAO) Collaboration

SEPAL is a cloud computing platform for autonomous land monitoring using remotely sensed data. It is a combination of GEE and open source software such as ORFEOToolbox, GDAL, R, R Studio Server, R Shiny Server, SNAP Toolkit, and OpenForis Geospatial Toolkit. It allows users to access powerful cloud-computing resources to query, access, and process satellite data quickly and efficiently for creating advanced analyses. SERVIR is working with the Food and Agriculture Organization (FAO) of the United Nations, who are the developers of SEPAL, to ingest the SAR processing scripts generated as part of this handbook and used during training events offered by the SAR experts authoring the main six chapters of this handbook at SERVIR hubs in Africa and Asia.

### 8.2.4 CAPACITY BUILDING EFFORTS

It is difficult to fully characterize the current state of various countries' capacity to use SAR data for forestry and biomass applications, especially at a granularity relevant for creating partnerships to raise awareness, promote literacy, and develop expertise. Without dedicated efforts to establish baselines, that task might similarly be daunting in the future. There are, nevertheless, proxies for understanding the current and evolving status of SAR capacity. One notable development is that the "club" of countries launching SAR satellites has been growing. Whereas, roughly a decade ago, the principal SAR missions (airborne and spaceborne) were mainly the fruits of Canadian, European, Japanese, and U.S. manufacture, recent and upcoming SAR missions include contributions from other nations. (Note that besides ESA's SAR satellites, other European efforts have included the TerraSAR-X and TanDEM-X satellites of Germany, and the CosmoSkyMed satellites of Italy.) These include Argentina (the SAOCOM satellites) and India (NISAR, jointly with the U.S.), and additional entries from European countries that previously did not have their own dedicated SAR missions, including the United Kingdom (the NovaSAR-S satellite), and Spain (the PAZ satellite). While a few additional countries launching SAR missions may not necessarily say much about SAR capacity, the use of SAR data and the "consumption" of capacity building opportunities may give another perspective on capacity. For instance, in its implementation of Copernicus, the European Commission publishes annual assessments of the usage of Copernicus data,

including from the Sentinel-1 satellites. As shown in **Figure 8.3**, certain regions and countries have much less data downloaded than others. Although Sentinel-1 is only one of a number of available sources of radar data, this is likely an indicator of the low uptake of SAR capacity.

Regarding online opportunities to acquire skills relating to the processing of SAR data, there are a number of notable efforts open to the public, among others:

- [Copernicus' Research & User Support \(RUS\) webinars](#)
- [EO College's Echoes in Space: Introduction to Radar Remote Sensing course](#)
- [NASA ARSET's Introduction to SAR and Advanced SAR courses](#)

In addition, the University of Alaska-Fairbanks, in collaboration with the University of Alabama in Huntsville, and the NASA Jet Propulsion Laboratory (JPL) is coordinating a three-year NASA-funded effort to develop a virtual capacity building center (CBC) for SAR (SAR-CBC) (Meyer et al. 2018). That effort is being piloted in three Spanish-speaking countries, and the Spanish- and English-language curricula and materials developed will be made

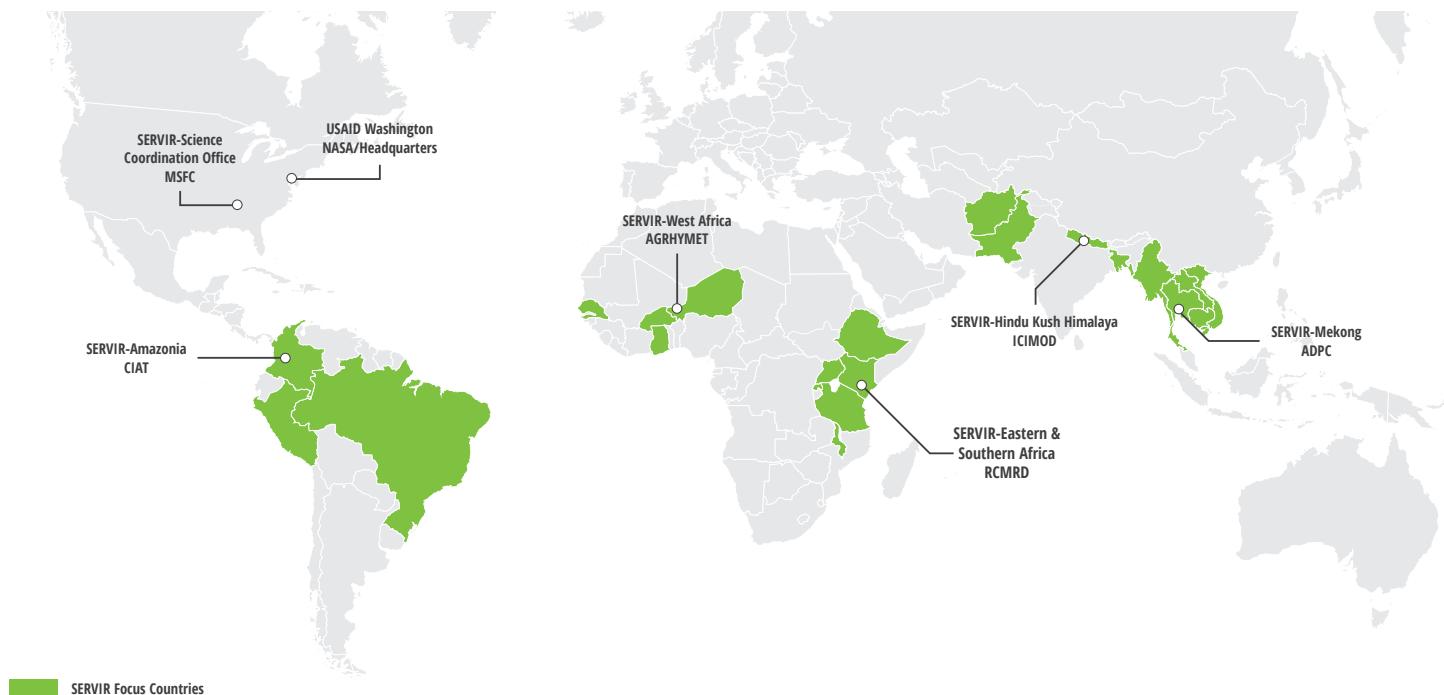
open to the public. The commonality among these efforts is a growing recognition that the internet has provided an opportunity for distance-based learning, open to the public. *Echoes in Space*, for instance, which launched in late 2017, is based on the premise that while there have been a number of online courses focused on various aspects of remote sensing, few focused on building capacity to use SAR data. *Echoes in Space*, in fact, built off an earlier platform, *SAR-EDU*, which launched in 2015. Another key resource to mention is NASA's Applied Remote Sensing Training (ARSET) that has provided introductory and advance trainings in the use of radar remote sensing. The high demand for SAR trainings has pushed ARSET to update their license to perform webinars and host more than 200 participants at a time. This highlights not only the high demand but also the critical mass of SAR users that is starting to form. GEE also provides an Internet-based platform that facilitates access and processing capacity of SAR data, especially in developing countries where access to proprietary software and hardware for data archiving has been a limiting factor to uptake of EO data processing for environmental and natural resource monitoring.

## 8.3 SAR Use in SERVIR Hubs

Given the potential of SAR for forest and broader environmental monitoring, SERVIR hub organizations are particularly interested in assisting with strengthening SAR capabilities in their respective focus regions (**Fig. 8.4**). Based on their current work plans and strategies, the following subsections provide details on SERVIR hubs likely future applications of SAR data. The SERVIR hubs were also engaged in the production of this handbook by hosting the hands-on trainings, which were converted into the hands-on sections of the respective chapters. Within the SERVIR network, the value of a practical handbook on SAR usage is that it can help prepare these regional centers of excellence for the data coming out of future missions, as it is expected that more data will be both free and available.

### 8.3.1 SERVIR-EASTERN AND SOUTHERN AFRICA

The Regional Centre for Mapping of Resources for Development (RCMRD, see: <http://rcmrd.org>) is an inter-governmental organization, which became the hub for the SERVIR-Eastern and Southern Africa (SER-



**Figure 8.4** Map of SERVIR focus regions.

VIR-E&SA, see <http://servir.rcmrd.org>) program in 2008. In 2018, RCMRD, hosted the trainings done by Marc Simard and Hans Andersen, focusing on SAR for mangrove monitoring and sampling design, respectively. Representatives of RCMRD also participated in the full range of SAR trainings under the SERVIR-SilvaCarbon collaboration. That included participation in training workshops on SAR data processing and the application of SAR for forest monitoring, respectively, led by Franz Meyer and Josef Kellndorfer, as well as trainings on the forest height estimation led by Paul Siqueira, and on biomass estimation, led by Sassan Saatchi. RCMRD is especially interested in using SAR for monitoring mangrove forests on the coasts of East Africa. In that context, RCMRD has already started giving training workshops to its member nations in the use of SAR. RCMRD is collaborating with the Kenya Forest Service (KFS) and the Kenya Forestry Research Institute (KEFRI) to develop their capacities in the use of SAR specifically for forest-related applications. In addition, RCMRD plans to support its member States as part of its new Strategic Plan 2019–2022 in the capacity-building activities related to the use of SAR for biomass estimation (e.g., Tanzania, Uganda, and Rwanda). Furthermore, RCMRD will engage with universities as part of its Summer School program in order to build capacity in the use of SAR in research.

### 8.3.2 SERVIR-WEST AFRICA

Since 2016, the Agro-meteorology, Hydrology, and Meteorology regional center (AGRHYMET, see <http://agrhymet.cilss.int>), a technical arm of the Permanent Interstate Committee for Drought Control in the Sahel (CILSS), has been the lead organization in the SERVIR-West Africa (see <http://servir.cilss.int/en/>) six-member consortium. In implementing SERVIR-West Africa, AGRHYMET has partnered with the African Center of Meteorological Application for Development (ACMAD), the African Regional Institute for Geospatial Science and Technology (AFRIGIST), the Centre for Remote Sensing and Geographic Information Services (CERSGIS), the Centre de Suivi Ecologique (CSE), and the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT). Even outside of the context of SERVIR, AGRHYMET has had prior experience working with SAR data.

As part of its involvement in the ESA TIGER initiative focused on monitoring water resources in Africa, AGRHYMET has collaborated with ESA researchers for using ESA SAR imagery for monitoring water bodies. Within the SERVIR context, in 2018, AGRHYMET hosted a joint training workshop by Franz Meyer and Josef Kellndorfer on the topics of radar data processing and using SAR to monitor deforestation and forest degradation. SERVIR-West Africa consortium members AFRIGIST, CERSGIS, CSE, and ICRISAT also participated in that training. In terms of future activities, and building off the earlier trainings, the SERVIR-West Africa consortium plans to integrate SAR data into its services, particularly those related to landcover change. The forested parts of West Africa are particularly susceptible to high cloud cover. Therefore, SAR presents an opportunity to more frequently monitor land-use and landcover change. Additionally, the SERVIR-West Africa consortium plans to apply the various methods of change detection to various land degradation services, such as the charcoal production monitoring service in Ghana, as well as one that focuses on understanding the community level forest changes.

### 8.3.3 SERVIR-HINDU KUSH HIMALAYA

The International Centre for Integrated Mountain Development (ICIMOD, see: <http://www.icimod.org>) is an intergovernmental organization, which, since 2010, has been the hub for the SERVIR-Hindu Kush-Himalaya (SERVIR-HKH, see: <http://servir.icimod.org>) program. In 2018, ICIMOD hosted two SERVIR/SilvaCarbon-sponsored regional trainings: (1) Franz Meyer and Josef Kellndorfer on SAR for mapping of forest degradation and deforestation, and (2) Sassan Saatchi on SAR for monitoring of forest carbon stocks and biomass estimation. ICIMOD's SAR capacity, strengthened through these workshops and trainings, is increasing across several thematic areas. Where ICIMOD collaborates with a number of government agencies of its member countries (Afghanistan, Bangladesh, Bhutan, China, India, Myanmar, Nepal, and Pakistan) in forest monitoring—including the recent establishment of a Regional Land Cover Monitoring System (RLCMS), leveraged from SERVIR-Mekong—ICIMOD plans to exploit SAR data

for national forest monitoring. While not specifically in the forest domain, ICIMOD has conducted wheat mapping activities in Afghanistan using a combined SAR/Optical approach. Going forward, ICIMOD is collaborating with Afghanistan's National Statistics and Information Authority (NSIA)/Central Statistics Organization (CSO) for capacity building regarding application of SAR data for seasonal wheat area estimation and plan to expand these methodologies for other crops in Bangladesh through a collaboration with the Bangladesh Agriculture Research Council (BARC). Additionally, building off the successful SERVIR trainings highlighted in this handbook, ICIMOD has committed to facilitate a follow-on workshop (2019) in the region, with global partners, to scope out the next steps in SAR capacity building in the HKH region related to SAR fundamentals, forest monitoring, and biomass estimation.

### 8.3.4 SERVIR-MEKONG

The Asian Disaster Preparedness Center (ADPC, see <http://adpc.net>) is an international organization, which, in 2015, became the hub for the SERVIR-Mekong program (see <https://servir.adpc.net>), along with consortium members Spatial Informatics Group (SIG), Deltares, and Stockholm Environment Institute (SEI). In 2018, ADPC hosted a training workshop focusing on the use of SAR for evaluating forest structure, led by Paul Siqueira. Representatives of ADPC attended the training workshop hosted at ICIMOD on SAR data processing and the application of SAR for forest monitoring, respectively, led by Franz Meyer and Josef Kellndorfer. ADPC is especially interested in incorporating SAR data into its RLCMS for the Mekong region. Supported by the SilvaCarbon program, ADPC has also begun to provide trainings to its focus countries in the application of SAR for forest monitoring. In FY2019, ADPC plans to explore delivering trainings on the use of SAR data for applications crop-type mapping. In addition, in the upcoming year, SAR will be used for water and water-related disaster applications, SAR-focused trainings, and uptake into other SERVIR-Mekong services.

### 8.3.5 SERVIR-AMAZONIA

At the time of writing, the SERVIR-Amazonia hub

was in the process of being established. It is not possible to describe the capabilities or needs of the institutions that will implement SERVIR-Amazonia. Nevertheless, the institutions in the region have experience with using SAR data. For instance, in the 1970s, the RADAM Project (Radar of the Amazon) utilized aerial radar surveys to map approximately 8.5 million km<sup>2</sup> of the Amazon (Carvalho 1984). That project was extended to the RADMBRASIL project, and the Ministry of Mines and Energy (Geog. Serv. of Brazil) hosted the project's output data. In more recent years, Colombia's Institute of Hydrology, Meteorology, and Environmental Studies (IDEAM, in Spanish) has piloted the use of Sentinel-1 data for monitoring changes in forest cover. The ongoing (2018–2021) previously mentioned NASA-funded SAR-CBC project is also working with institutions in Colombia and Ecuador to develop their SAR analysis capabilities. Depending on the outcomes of the service planning process involving SERVIR-Amazonia, given the existing SAR capabilities in South America and cloud cover limitations over the Amazon, it is certainly possible that SERVIR-Amazonia may also focus on strengthening the use of SAR for monitoring the Amazon's forests.

## 8.4 Other Considerations

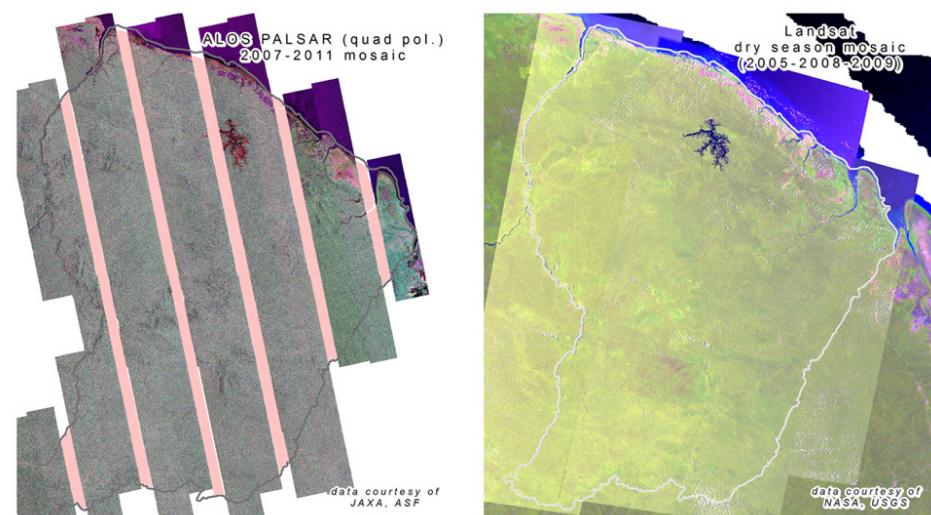
Having previously explored in detail both how SAR is currently and will likely continue to be used for forest monitoring, it is also useful to reflect on how SAR's potential future contributions to other application areas beyond forest monitoring. It is also useful to examine, to some extent, how SAR data can and will be integrated with other types of remote sensing data, such as multispectral imagery and lidar data.

### 8.4.1 OTHER APPLICATION AREAS

As indicated in the previous section on the involvement of SERVIR hubs in SAR capacity building activities in their regions, those hubs also have interest in the use of radar data for other applications. Those applications include, but are not limited to, the list presented in **Table 8.1**. The table addresses how the applications relate to GEO's societal benefit areas (SBAs) and the United Nations' Sustainable SDGs. It goes without saying that via the "Earth Observations in Service of the 2030 Agenda for Sustainable Development" activity

GEO SBA	APPLICABLE SDG	APPLICATION
Biodiversity & ecosystem sustainability	Life on land	Mapping ecosystems, estimating habitat suitability and species richness by ecosystem type
Disaster resilience	Sustainable cities and communities	Estimating elevation for modeling flood susceptibility (disaster mitigation and preparedness) Mapping flooded areas (disaster response) Mapping historic flood extents (disaster mitigation and preparedness) Estimating ground deformation caused by earthquakes and landslides (disaster response)
Energy & mineral resource management	Industry, innovation, and infrastructure	Mapping oil spills Monitoring reservoir extents and levels
Food security & sustainable agriculture	Zero hunger	Crop type identification based on vegetative phenological changes (and surface water for flooded rice) reflected in backscatter changes Evaluating drought impacts by estimating changes in crop biomass Start of Season (SoS)-based backscatter signatures of crop emergence
Sustainable urban development	Sustainable cities and communities	Mapping urban expansion Monitoring subsidence due to development
Water resource management	Clean water and sanitation	Mapping changes in water body extent over time

**Table 8.1** Other uses for SAR data, and how they relate to GEO's SBAs and the SDGs.



**Figure 8.5** Contrast between polarimetric ALOS PALSAR data (left) and Landsat multispectral imagery (right) of French Guiana (Cherrington 2016).

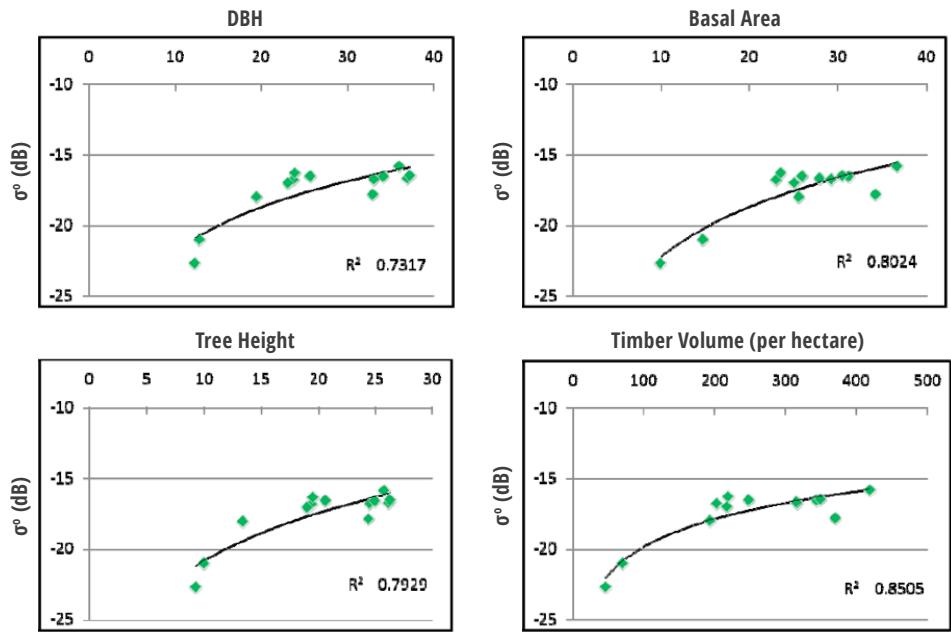
in GEO's 2017–2019 Work Programme—an initiative known as EO4SDG—work is already being done to highlight how Earth observation data, including SAR, can be used to support countries' implementation of the SDGs. A 2017 report on the EO4SDG initiative likewise highlights the relevance of SAR to SDGs 2, 6, 11, and 15 (GEO 2017).

### 8.4.2 LIMITATIONS AND INTEGRATION WITH OTHER TYPES OF REMOTELY SENSED DATA

While the preceding chapters have highlighted the potential of SAR data for forest monitoring, an important caveat is that SAR data are not necessarily suited for all forest applications. For instance, one application not highlighted in this handbook—likely

because of the level of difficulty—is the use of SAR imagery for distinguishing forest types as defined by unique tree communities. As an example, from **Figure 8.5**, it is apparent that various forest formations (e.g., montane forests) are visible in multispectral optical data, but do not show strong contrasts with other forest types in polarimetric SAR imagery. And in recognizing that limitation, it must be recalled that while L-band SAR data (such as the data displayed) are sensitive to ecosystem structure, the multispectral imagery, in contrast, will be sensitive to factors which vary even more greatly than structure, such as variable combinations of leaf spectral properties, leaf area index, leaf angle distributions, among others (Gastellu-Etchegorry et al. 2015). This emphasizes previously raised points about the value of complementing SAR data with other data types such as optical and lidar data.

While lidar is treated as the gold standard in terms of its ability to map the spatial distribution of the biomass of forests, SAR is sometimes—erroneously—treated as merely a proxy for lidar. Nevertheless, in terms of the complementarity between SAR and lidar, SAR backscatter—which usually covers large spatial extents, relative to smaller lidar coverages—has been used on occasion for extrapolating from lidar data. Related to that, and returning to the contrast between multispectral data and SAR backscatter, it is important to recall that radar backscatter is sensitive to forest structure, compared to, say, forests' species composition, which multispectral imagery are more sensitive to. As illustrated in the comparisons shown in **Figure 8.6**, for instance, it can be seen that there are strong correlations between L-band HV backscatter and various forest structural parameters such as diameter at breast height (DBH), basal area, tree height, and timber volume. This reinforces the concept that radar backscatter is representative of some general-



**Figure 8.6** Relationship between forest stand parameters and ALOS PALSAR HV backscatter (Holecz 2011, [https://www.eorc.jaxa.jp/ALOS/kyoto/jan2011\\_kc15/pdf/1-10\\_kc15\\_sarmp.pdf](https://www.eorc.jaxa.jp/ALOS/kyoto/jan2011_kc15/pdf/1-10_kc15_sarmp.pdf))

ized forest structure parameter. Furthermore, with growing availability of SAR data and a growing body of research to accompany that data, it will be interesting to see how future scientific advances might impact what are currently limitations to using SAR data. That will be particularly relevant for upcoming L-band and P-band missions like NISAR and Biomass, the latter for which there is no precedent. Beyond this brief discussion of research challenges, the other key limitation is capacity to apply SAR, which have been discussed through the future capacity building efforts in this chapter.

## 8.5 Final Thoughts

In this chapter, the contents of the practical applications encountered throughout this handbook have been summarized. Future SAR developments, such as the possibility of increasing data availability, new ways to access SAR data, and developments

in desktop- and cloud-based processing environments were also addressed. A summary has been provided for recent and future capacity building initiatives and programs, as well as prime additional application areas for SAR, from the perspective of SERVIR regions. Finally, as discussed throughout this handbook, the forestry sector should not see SAR as a solution to be considered in isolation. Challenges and opportunities abound in blending SAR data with other remote sensing and ground-based techniques for forestry and biomass applications.

This handbook serves to create a practical resource for applying SAR data for forestry and biomass estimation. In support of GFOI and SilvaCarbon, the global network of SERVIR hubs are seen as important resource centers to articulate and address new environmental monitoring challenges and to help strengthen the capacity to apply SAR for forest monitoring.

## 8.6 References

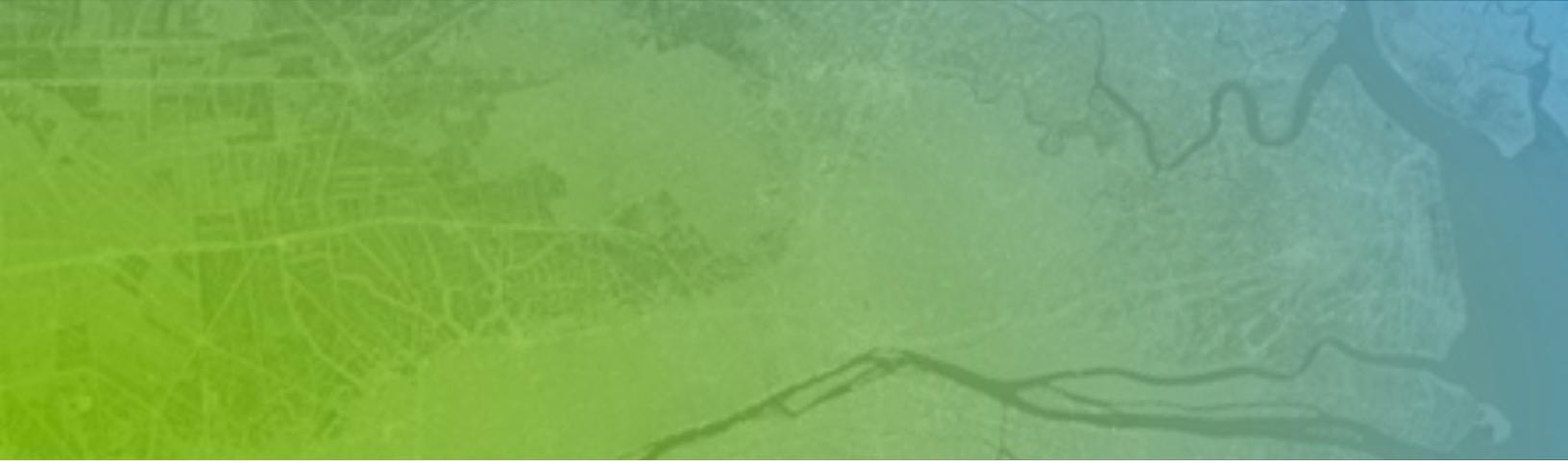
- Asian Disaster Preparedness Center (ADPC). 2018. SERVIR-Mekong Annual Work Plan, FY 2019. Bangkok, Thailand. 59 pp.
- Albinet, C., Whitehurst, A.S., Laur, H., Murphy, K.J., Frommknecht, B., Scipal, K., Mitchell, A.E., Jai, B., Ramachandran, R. and Lavalle, M., 2018. ESA-NASA multi-Mission Analysis Platform for improving global aboveground terrestrial carbon dynamics. European Geophysical Union 2018. <https://meetingorganizer.copernicus.org/EGU2018/EGU2018-6626.pdf>
- Belgiu, M., and Dragut, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114: 24-31. <http://dx.doi.org/10.1016/j.isprsjprs.2016.01.011>
- Blumenfield, J., 2018. "Getting Ready for NISAR—and for Managing Big Data using the Commercial Cloud." NASA Earthdata / EOSDIS. Accessed 17 Dec 2018. <https://earthdata.nasa.gov/getting-ready-for-nisar>
- Carvalho, J.C.M. 1984. *The conservation of nature in the Brazilian Amazonia*. In: Sioli, H. (Ed.) *The Amazon: limnology and landscape ecology of a mighty tropical river and its basin*. pp. 707-736. Dr. W. Junk Publishers, Dordrecht, Holland.
- Chen, L., Ren, C., Zhang, B., Wang, Z., and Xi, Y., 2018. Estimation of Forest Above-Ground Biomass by Geographically Weighted Regression and Machine Learning with Sentinel Imagery. *Remote Sensing* 9(10) 582. <https://doi.org/10.3390/rs9100582>
- Cherrington, E.A. 2016. Towards ecologically consistent remote sensing mapping of tree communities in French Guiana: Are forest types identifiable from spatio-temporal canopy reflectance patterns? PhD dissertation. AgroParisTech / Technische Universität Dresden. 142 pp. <https://tel.archives-ouvertes.fr/tel-01486533>
- Gastellu-Etchegorry J.P., Yin T., Lauret N., Cajfinger T., Gregoire T., Grau E., Feret J.B. 2015. Discrete Anisotropic Radiative Transfer (DART 5) for Modeling Airborne and Satellite Spectroradiometer and LiDAR Acquisitions of Natural and Urban Landscapes. *Remote Sensing*, 7 (2): 1667-1701. doi:10.3390/rs70201667.
- Group on Earth Observations (GEO). 2017. *Earth Observations in support of the 2030 Agenda for Sustainable Development*. GEO / EO4SDG initiative. 34 pp. [https://www.earthobservations.org/documents/publications/201703\\_geo\\_eo\\_for\\_2030\\_agenda.pdf](https://www.earthobservations.org/documents/publications/201703_geo_eo_for_2030_agenda.pdf)
- Geological Service of Brazil. <http://www.cprm.gov.br>. Accessed November 2018.
- Gorelick N., Hancher M., Dixon M., Ilyushenko S., Thau D., Moore R. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202: 18-27.
- Hird J.N., DeLancey E.R., McDermid G.J., Kariyeva J. 2017. Google Earth Engine, Open-Access Satellite Data, and Machine Learning in Support of Large-Area Probabilistic Wetland Mapping. *Remote Sensing*, 9: doi: 10.3390/rs9121315
- Holecz, F. 2011. Product Delivery Report for K7C Phase 2. Kyoto & Carbon Initiative, Science Team Meeting # 15. PowerPoint presentation. 15 pp. [https://www.eorc.jaxa.jp/ALOS/kyoto/jan2011\\_kc15/pdf/1-10\\_kc15\\_sarmap.pdf](https://www.eorc.jaxa.jp/ALOS/kyoto/jan2011_kc15/pdf/1-10_kc15_sarmap.pdf)
- International Centre for Integrated Mountain Development (ICIMOD). 2018. SERVIR-Hindu Kush Himalaya Annual Work Plan: FY 2019. Cooperative Agreement Number: AID-EGEE-10-15-00002. Kathmandu, Nepal. 70 pp.
- Lillesand T., Kiefer R., Chipman J. 2011. *Remote Sensing and Image Interpretation*. 6th Edition. Wiley India. ISBN-13: 978-8126532230.
- Meyer, F., Anderson, E., Flores Cordova, A., Cherrington, E., Griffin, R., Rosen, P., Owen S., Arko, S., Galindo Garcia, G., Mayorga Torres, T., and Limaye, A. 2018. A Capacity Building Center for the Use of SAR in Decision Making. American Geophysical Union Fall Meeting 2018. Washington, DC.
- Ndikumana E., Ho Tong Minh D., Baghdadi N., et al. 2018. Deep Recurrent Neural Network for Agricultural Classification using multitemporal SAR Sentinel-1 for Camargue, France. *Remote Sensing*, 10: doi: 10.3390/rs10081217.
- Regional Centre of Mapping of Resources for Development (RCMRD). 2018. SERVIR-Eastern and Southern Africa Annual Work Plan: FY 2019. Cooperative Agreement Number: AID-EGEE-10-15-00002. Nairobi, Kenya. 63 pp.
- Rosen, P.A., Gurrola, E.M., Agram, P., Cohen, J., Lavalle, M., Riel, B.V., Fattah, H., Aivazis, M.A.G., Simons, M., and Buckley, S.M. 2018. The InSAR Scientific Computing Environment 3.0: A Flexible Framework for NISAR Operational and User-Led Science Processing. In: 2018 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2018). IEEE, Piscataway, NJ, pp. 4897-4900. ISBN 978-1-5386-7150-4. <http://resolver.caltech.edu/CaltechAUTHORS:20181108-155745164>
- SERCO. 2018. *Sentinel Data Access 2017 Annual Report*. Technical report to the European Space Agency. COPE-SERCO-RP-17-0186. 96 pp. [https://scihub.copernicus.eu/twiki/pub/SciHubWebPortal/AnnualReport2017/COPE-SERCO-RP-17-0186\\_-\\_Sentinel\\_Data\\_Access\\_Annual\\_Report\\_2017-Final\\_v1.4.1.pdf](https://scihub.copernicus.eu/twiki/pub/SciHubWebPortal/AnnualReport2017/COPE-SERCO-RP-17-0186_-_Sentinel_Data_Access_Annual_Report_2017-Final_v1.4.1.pdf)
- Shelestov A., Lavreniuk M., Kussul N., et al. 2017. Exploring Google Earth Engine Platform for Big Data Processing: Classification of Multi-Temporal Satellite Imagery for Crop Mapping. *Frontiers in Earth Science*, 5:17. doi: 10.3389/feart.2017.00017
- Tetra Tech. 2018. SERVIR West Africa Annual Work Plan: Fiscal Year 2019. Contract No. AID-OAA-I-13-00058, Task Order No: AID-624-T0-16-00001. Washington, DC. 112 pp.
- Woodhouse, I. 2005. *Introduction to Microwave Remote Sensing*. Boca Raton: CRC Press. ISBN-13: 978-0415271233. 400 pp. <https://doi.org/10.1201/9781315272573>

# ABOUT THE EDITORS

**AFRICA I. FLORES-ANDERSON** is a Research Scientist at the Earth System Science Center in the University of Alabama in Huntsville, with extensive experience in the applied use of satellite remote sensing for environmental monitoring. She has worked with SERVIR since 2008 starting in Central America, and currently leads the Land Cover Land Use Change and Ecosystems thematic portfolio of SERVIR-Global from the SERVIR Science Coordination Office at NASA Marshall Space Flight Center. She is also the SERVIR-Amazonia Science Coordination Lead. Flores' research focuses in the applied use of satellite remote sensing for forest monitoring, water quality and ecological forecasting.

She is the NASA representative for the SilvaCarbon initiative, and is leading the collaboration between SERVIR and SilvaCarbon to create applied knowledge and capacity throughout the SERVIR network on the use of Synthetic Aperture Radar (SAR) for forest monitoring and biomass estimation.

**KELSEY E. HERNDON** is the Regional Science Associate for Amazonia at the SERVIR Science Coordination Office at NASA Marshall Space Flight Center in Huntsville, Alabama. She is a remote sensing specialist and social scientist whose research interests include water resources, integrating remote sensing to address issues of natural resource management (specifically water scarcity), and mitigating local/regional conflicts over natural resources. Her research at SERVIR has focused on the long term dynamics of ephemeral water bodies in the West African Sahel, and the implications of political, economic, and cultural practices on their use. She is broadly interested in the role of formal and informal institutions in natural resource management, as well as in addressing the challenges that accompany incorporating various scales of social, cultural, and economic data with remote sensing data. Kelsey has an MS in Earth System Science from the University of Alabama in Huntsville and an MA in Anthropology from the University of Alabama.



**DR. RAJESH BAHDUR THAPA** works at ICIMOD and leads the Capacity Building Programme of SERVIR-HKH, MENRIS and the Group on Land Use, Land Cover Change and Ecosystem Services, Geospatial Solutions. He has over twenty years of experience working in Asia, including in Nepal, Thailand, and Japan. Prior to joining ICIMOD, he was a researcher at Japan Aerospace Exploration Agency (JAXA) working for the Japanese Earth observing missions ALOS, ALOS-2, and PiSAR-L2. He was also a visiting professor at the University of Tsukuba, Japan. He has extensive expertise in SAR, optical, and LiDAR remote sensing data processing and analysis for various applications. His research focus is on monitoring and assessment of terrestrial environments, including forest, agriculture, urban, and disasters thematic areas, as well as the dissemination of Earth observation findings to policy makers, practitioners, university students, and stakeholders through education, capacity building workshops, conferences, and publications. He has conducted extensive research and fieldwork in Indonesia, Japan, Nepal, Thailand, and Vietnam. He holds a PhD in Geoenvironmental Science from University of Tsukuba (2009), a MSc in Remote Sensing and GIS from the Asian Institute of Technology (2003), and a master's degree in Geography from Tribhuvan University (1998).

**DR. EMIL CERRINGTON** is a forest ecologist and remote sensing scientist whose research interests include mapping of forest and land cover types in Central America and French Guiana. While much of his research has focused on the use of multispectral data, he has also dabbled in the use of Synthetic Aperture Radar (SAR) for forest mapping in Central America using ALOS PALSAR, and Sentinel-1 imagery for detecting forest cover change, and has given workshops focusing on the latter. In his role as a research scientist with the University of Alabama in Huntsville, he currently serves as a co-investigator on the SAR virtual Capacity Building Center (SAR-CBC) project led by Principal Investigator Franz Meyer. He also serves as the Regional Science Coordination Lead for West Africa at the NASA SERVIR Science Coordination Office.