Semantic Workflows and Machine Learning for the Assessment of Carbon Storage by Urban Trees

Juan Carrillo
University of Waterloo
Waterloo, Canada
jmcarril@uwaterloo.ca

Daniel Garijo
Information Sciences Institute
University of Southern California
USA
dgarijo@isi.edu

Mark Crowley
University of Waterloo
Waterloo, Canada
mark.crowley@uwaterloo.ca

Rober Carrillo
United World Colleges
Bogota, Colombia
rober.carrillo@uwcim.net

Yolanda Gil
Information Sciences Institute
University of Southern California
USA
gil@isi.edu

Katherine Borda
Waterloo, Canada
kbordac@gmail.com

ABSTRACT

Climate science is critical for understanding both the causes and consequences of changes in global temperatures and has become imperative for decisive policy-making. However, climate science studies commonly require addressing complex interoperability issues between data, software, and experimental approaches from multiple fields. Scientific workflow systems provide unparalleled advantages to address these issues, including reproducibility of experiments, provenance capture, software reusability and knowledge sharing. In this paper, we introduce a novel workflow with a series of connected components to perform spatial data preparation, classification of satellite imagery with machine learning algorithms, and assessment of carbon stored by urban trees. To the best of our knowledge, this is the first study that estimates carbon storage for a region in Africa following the guidelines from the Intergovernmental Panel on Climate Change (IPCC).

KEYWORDS

Reproducibility, scientific workflows, machine learning, land cover mapping, carbon assessment, Sentinel-2

1 INTRODUCTION

Climate science requires modeling natural and man-made processes that are highly complex, exhibit non-linear dynamics and possess disparate spatial and temporal scales. Handling this complexity requires a holistic approach among multiple disciplines [29], but scientists from different fields may also need to use domain-specific data sources, methods, and computational models. The integration of their knowledge and experiments is a challenging task [7], especially when a study is expected to provide actionable insights for decision making at regional and local scale.

Scientific workflows have emerged as an integrated solution to manage this challenge, as they capture the computational steps and data dependencies required to carry out a computational experiment [40]. Scientific workflows ease data handling (metadata, provenance), component versioning (parametrization, calibration) and have a clear separation between workflow design and workflow execution [4] [43]. One of the major advantages of scientific workflow systems is their role in improving the reproducibility of scientific studies. Reproducibility plays a critical role in climate sciences due to their impact in our society [38]. In fact, due to issues with the documentation of experiments, some climate science studies have been re-examined lately due to their impact in global policy-making [19] and water resources management [25].

In this paper we describe the process we followed to design and implement reusable scientific workflows for the climate sciences. In particular, we focus on carbon storage assessment by using urban trees, a common requirement for cities to reduce their carbon emissions globally. Our contributions include the development of a library of components for preparing geospatial data by doing coordinate transformations, the integration of machine learning components to classify trees in satellite images and the creation of workflows for carbon storage assessment in cities. In order to implement these workflows, we use the Workflow Instance Generation and Selection (WINGS) system [18], which has been successfully used for applications in domains ranging from Genomics [17] to Geosciences [16].

The paper starts by giving an overview of previous research on the assessment of carbon storage by urban trees according to the guidelines published by the Intergovernmental Panel on Climate Change (IPCC), highlighting the advantages and limitations of the most recent methods. We then describe the design considerations of our scientific workflows as well as their experimental implementation and evaluation. The paper continues with a discussion of results and suggested future work.

2 BACKGROUND: CARBON EMISSIONS AND STORAGE

There is an increasing interest among scientific organizations and national governments regarding Carbon emissions and their role in climate science [13]. The consequences of higher concentrations of carbon gases in the atmosphere are now more explicit and international organizations such as the Intergovernmental Panel on Climate Change (IPCC) are leading initiatives to monitor national efforts to lower emissions and increase carbon storage [10]. Monitoring carbon emissions is fundamental to inform government policies in topics such as renewable energy, transportation, and manufacturing technologies. Similarly, the assessment of carbon storage is equally important to guide hazard mitigation efforts [20].

Most studies in climate science require domain knowledge to design and run the experiments [23]. But the increasing concerns about the changing climate require strategies to streamline the replication of assessment studies, the reusability of data, methods, and results. The use of scientific workflows can significantly improve the implementation and reproducibility of carbon assessment studies, with additional gains in data and model sharing as well as knowledge capture through semantic representations.

2.1 Carbon assessment by urban trees

Urban trees provide a natural and cost-effective alternative to capture and store carbon in cities. Having trees in densely populated areas also improve human health and biodiversity and provide benefits for flood prevention and reduced cooling costs, among other benefits [26]. In 2003, the IPCC published the *Good Practice Guidance for Land Use, Land-Use Change and Forestry* [33] and in 2006 the *IPCC Guidelines for National Greenhouse Gas Inventories* [11]. These guidelines suggest the use of area covered by trees, shrubs, and herbaceous (perennial) plants to determine the amount of carbon stored as biomass in settlements. However, due to the limited availability of detailed data the majority of published studies focus only on tree cover. While these two documents describe the stages of an assessment study, including aspects such as data collection and uncertainty estimation, they suggest governments to deal with minor implementation details according to their technical capacity and available resources.

Published carbon assessment studies use different combinations of data, methods, and software. These studies can be divided into two major groups according to the data collection approaches and models they use. Assessments in the first group use a statistics point sampling technique to estimate tree density from aerial imagery [28][32][31]. This method is easy to implement and only requires imagery for sample areas, but the outcome is a percentage value that does not describe the spatial distribution of trees. The second group of methods use LiDAR, aerial or satellite imagery to provide a comprehensive assessment of tree coverage, including their spatial distribution [42][37][8][35]. However, the second approach requires imagery for the complete the area of study as well as the configuration of more complex methods for detection or classification of urban trees. Later in this document we introduce our own method, which belongs to the second group and uses freely available satellite imagery and a carefully designed workflow to facilitate implementation and reuse.

One common limitation of carbon assessment studies is the lack of a systematic approach to share data, models, software, and results. Regardless of the specific data source, technology, or processing method, most reports only contain descriptions of the work done, which are not enough to replicate the experiments or reuse the software tools [34].

Turning the information from those reports into actionable knowledge becomes a cumbersome task, especially for scientists in developing countries where technical capacity and resources are particularly limited [1].

3 SCIENTIFIC WORKFLOW DESIGN

With the advantages and limitations of published methods in mind, we design a new workflow to efficiently determine tree cover for urban areas We start by presenting the advantages of knowledge capture systems to represent models in geosciences as scientific workflows and then describe how we leverage previous research on carbon assessment and tree mapping to design our own workflow.

The workflow is designed as multiple interconnected components in WINGS that operate in three consecutive stages as seen in Figure 1, data preprocessing, mapping of tree coverage, and assessment of carbon storage. Our workflow is based on previous work in which high resolution satellite imagery are used to produce land cover maps over urban areas [35] [37]. However, we focus on using freely available medium resolution satellite imagery from the Sentinel-2 sensors [9] to facilitate replication by other researchers.

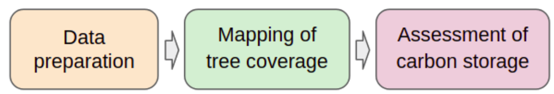

Figure 1: Stages of our carbon assessment workflow

Spatial data preprocessing stage involves common operations for experiments across earth sciences, such as transformation of coordinate systems and conversion between file formats. We have designed all these data preprocessing steps as reusable building blocks so they can be included in other workflows. These preprocessing operations are also known as Extract Transform Load (ETL) tasks and are implemented using the Geospatial Data Abstraction Library GDAL [15].

We map the tree coverage using satellite image classification and design multiple components to train machine learning algorithms, classify the image over an area of interest, produce a visualization ready tree cover map, and determine the resulting accuracy. The Machine Learning algorithms we implement are Random Forests and Support Vector Machines, both available in the Orfeo Remote Sensing toolbox [21]. Random Forest [36] [5] and Support Vector Machines [41] [6] are well-known methods for pixel-based image classification in Remote Sensing due to their straightforward implementation and calibration, as well as its documented robustness and accuracy.

We train the algorithms using sample points collected through visual inspection as described further in this document. Our workflow generates a map that includes other land cover categories such as water, grass and built areas; which may additionally serve for other use cases in disciplines such as hydrology, planning, and forestry, just to name a few (as seen in Figure 2). Moreover, this map is generated in a standard format for further use in Geographic Information Systems or other scientific platforms.

The assessment of carbon storage is completed following the IPCC guidelines to calculate carbon stored based on urban canopy area. In the calculation we multiply the canopy area by a conversion

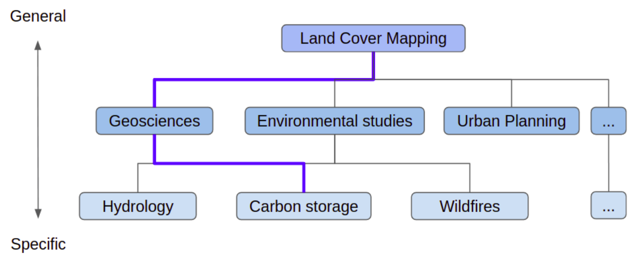

Figure 2: Some applications of land cover mapping

factor to estimate carbon stored in the form of biomass. Since no values are published specifically for Africa (our initial region of interest) we use a default value suggested by the IPCC.

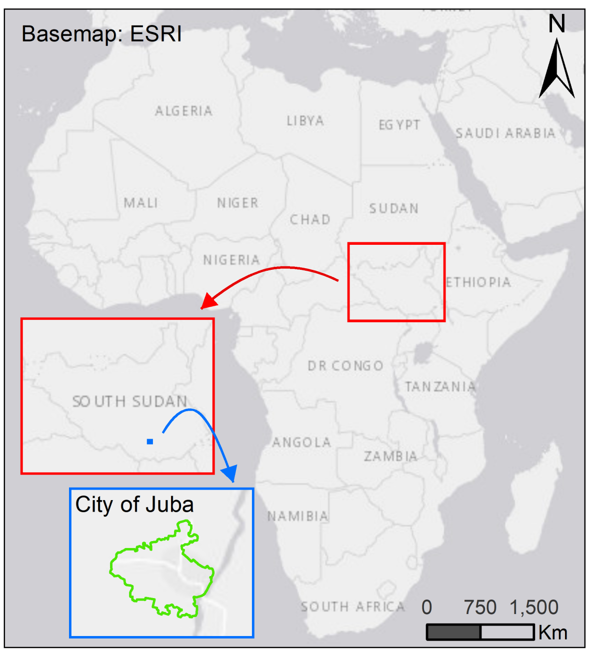

Figure 3: City of Juba in South Sudan, Africa

4 AREA OF STUDY AND DATA

Our area of study is the city of Juba in South Sudan. It is the current capital of the country and serves as its main commercial and transportation hub with an estimated population of nearly 386,000 inhabitants [2]. Juba is located in the southern region of the country and has an extension of 103 km^2 according to the urban boundary retrieved in July 2019 from Open Street Maps [22]. Figure 3 shows the location of Juba and South Sudan in the African continent.

The country of South Sudan currently faces multiple issues, including political instability [39], poor health services [27], and a lack of infrastructure, especially for storage and distribution of water [30]. Few mapping projects have been conducted in Africa, and especially in South Sudan [3][24], which hinder the work of humanitarian and non government organizations; therefore, any study aiming to understand the characteristics of the rural or urban territories will likely produce positive outcomes in the short and long term.

The data we use includes 4000 sample points digitized through visual inspection using Google Earth high resolution satellite imagery. The digitization is performed by trained individuals and even though this task is not implemented as one of the workflow components we follow best practices in Geographic Information Systems [14] and complete random checks to ensure data accuracy and consistency. The specific steps are as follows: First, 4000 random points are generated within the city boundary and used as spatial reference to ensure that the actual sample points are spatially distributed across the city. Then, the person who is digitizing places sample points closer to the random points but over locations where examples of each land category are seen. The four land categories are trees, grass, impervious, and water, with 1000 points per category to have a balanced dataset. Impervious includes all built or bare areas not covered by vegetation, water, or agriculture. Examples of impervious areas are constructions, parking lots, roads, airport runways, and rocky surfaces.

Additionally, we use a multi-spectral satellite image from the Sentinel-2 sensor made available by the European Space Agency (ESA) [9]. While this image originally includes 13 bands we use only the four bands with a pixel size equivalent to 10 meters on the ground. The first three bands collect data in the visible light spectrum and the fourth in the near infrared spectrum, with the latter being particularly useful for detecting vegetation.

Figure 4 shows some sample points of the four land cover categories in a neighborhood close to the White Nile river in the city of Juba, overlaid on high resolution satellite imagery from Google Earth and the same area as seen on the medium resolution Sentinel-2 image.

5 IMPLEMENTATION AND RESULTS

The workflow is implemented as a series of software components in WINGS. Additionally, each component is designed to run a particular task and configured with relevant parameters to calibrate their functionality. Components are coded as Python scripts that serve as a high level interface for us to use libraries such as GDAL for data preprocessing and Orfeo command line tools for satellite image classification.

These components are connected through intermediate datasets that in turn are outputs and inputs for the previous and following components. We created 14 components in total, with eight dedicated for data preparation, five for mapping of tree coverage, and one for carbon assessment. The use of components as modular pieces of software to accomplish specific data processing tasks creates opportunities for reusability across a variety of models in Earth Sciences and Geospatial technologies.

The eight components designed for data preparation allow researchers to handle datasets in the most common file formats and transform them according to the particular goals of the study. Some operations correspond to sub-setting the spatial extent of the data, changing its coordinate reference system, and editing the attributes

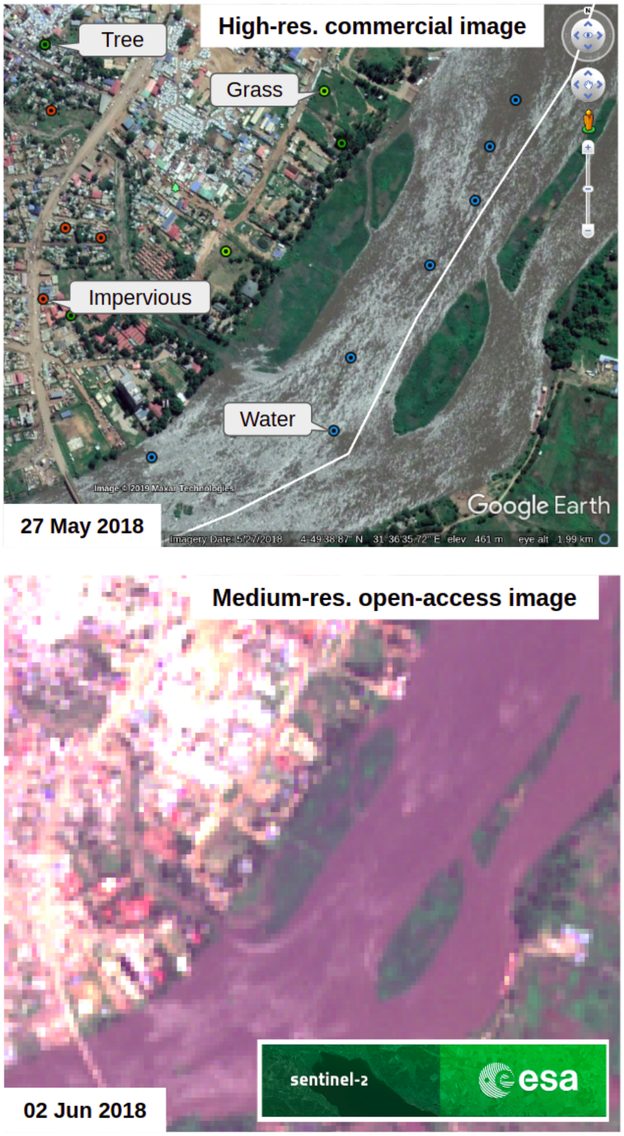

Figure 4: Sample points and satellite imagery of Juba

for tabular datasets. When running the workflow for a new area of study, the task of setting parameters such as the coordinate reference system is facilitated by the WINGS system, which suggest the most appropriate value according to the geographic region as configured by the workflow designer. Figure 5 shows a fragment of our workflow where we use multiple data preparation components to perform a format transformation and reprojecting a file.

Mapping of tree coverage includes four components focused on training the Machine Learning algorithms to classify the Sentinel-2 satellite image using the sample points digitized through visual inspection. Initially, one component extracts the pixel values of the satellite image for the 4000 sample locations. Next, another component uses 80% of these values as training and 20% as validation data to train the Random Forest and Support Vector Machine

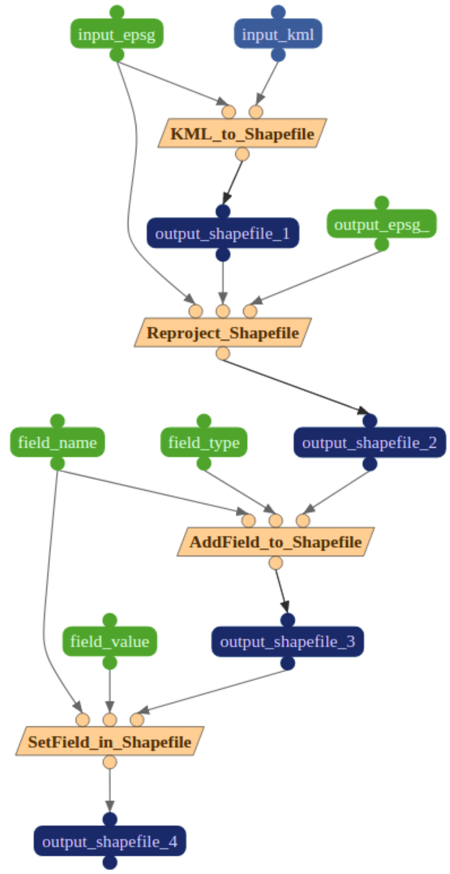

Figure 5: Workflow fragment for data preparation of sample points

algorithms. Then we use the trained algorithms to perform pixel level classification of the Sentinel-2 image for the entire area of interest and output the tree cover map. Subsequently, one more component evaluates the classification accuracy. Figure 6 shows a fragment of our workflow where we use multiple components to map tree cover.

Table 1 shows the normalized confusion matrix resulting from the evaluation of the Random Forest classification algorithm. We see that the tree cover category is the one with the lowest accuracy with only 54% of the trees classified correctly in the test set. This is likely the result of using medium resolution satellite imagery (Sentinel-2) with a ground pixel size of 10 m, in other words, the canopy area of a tree should be about 100 m^2 to be easily identifiable in at least one pixel, without considering boundary issues between adjacent pixels. The grass and impervious land cover categories exhibit a comparable accuracy of 65% and 73% consequently. While areas corresponding to these two categories show a slightly better accuracy they are still hard to differentiate, presumably due to grass patches and house rooftops with a size smaller than the area of a pixel (100 m^2). For the water land cover category the algorithm reaches an almost perfect accuracy, which is anticipated due to the significant difference in the way it reflects the light compared to

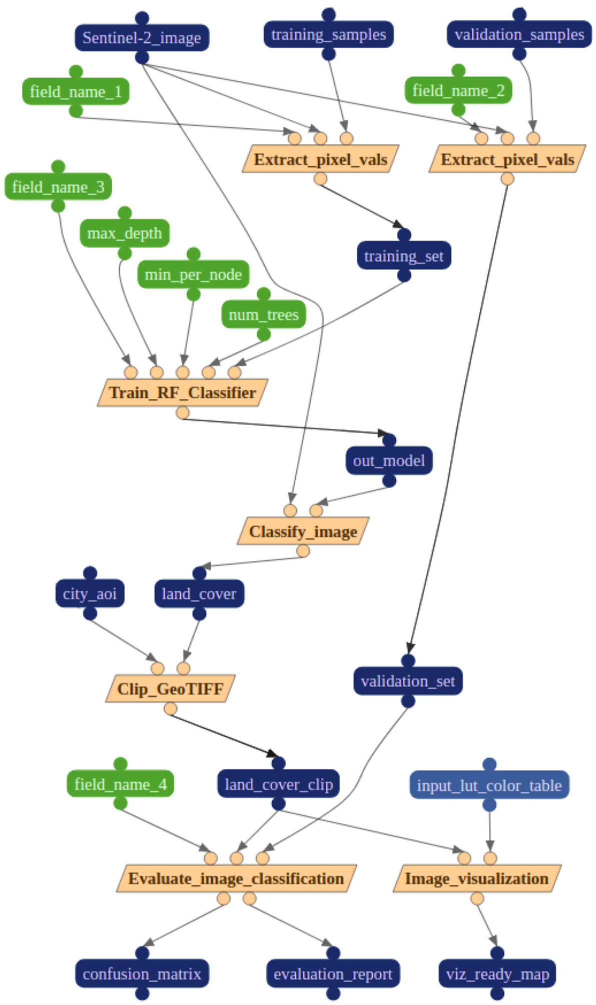

Figure 6: Workflow fragment for mapping tree cover

Table 1: Normalized confusion matrix

Actual class	Predicted class: Trees	Grass	Imp.	Water
Trees	0.54	0.18	0.28	0
Grass	0.32	0.65	0.14	0
Impervious	0.17	0.09	0.73	0
Water	0.01	0	0	0.99

the other three land cover categories, especially in the near infrared band (B4).

The assessment of carbon storage is performed by one component that receives the tree cover map shown in Figure 7, calculates the total canopy area and uses a carbon removal factor to determine the total amount of carbon stored as biomass in the city of Juba. We use a default carbon removal factor of 2.9 tonnes C (ha crown

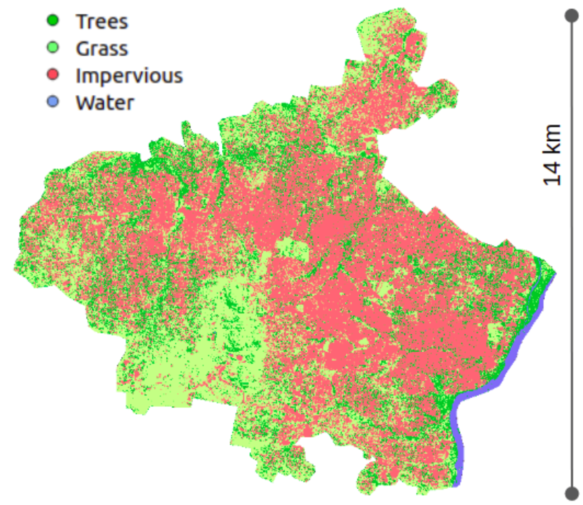

Figure 7: Resulting land cover map for the city of Juba

cover)$^{-1}$ yr $^{-1}$, which is the value suggested by the IPCC for Tier 2a studies [11]. Multiplying the default carbon removal factor by the total tree cover area (10,519 ha) from the previous stage, we calculate that trees in the city of Juba remove 30,506 tonnes C yr $^{-1}$. This amount is equivalent to the carbon dioxide emitted by 6632 passenger vehicles per year [12]. This value along with the classification accuracy from the previous stage are valuable information for countries to prepare their carbon assessment reports according to the IPCC guidelines. The code used to create the geospatial transformations [1] and the carbon assessment workflows [2] is available online.

6 CONCLUSIONS AND FUTURE WORK

In this paper we introduced our work to create a library of workflow components to perform spatial data transformations, land cover mapping and assessment of carbon storage. By leveraging scientific workflows, we aim to ease the reusability of these components in other workflows and the reproducibility and transparency of carbon assessment studies. Our future work will focus on two main areas: first, we aim to test our workflow using data from other locations around the globe, which requires additional training data points. Second, we will focus on calibration of the parameters for the classifiers to improve our classification accuracy during the land cover mapping stage.

7 ACKNOWLEDGMENTS

Juan Carrillo was financially supported for this project by the Mitacs Globalink Research Award and the University of Waterloo Machine Learning Lab. Juan Carrillo gives special thanks to Mark Crowley, Daniel Garijo, and Yolanda Gil for their mentoring and valuable contributions during this research project.

[1] https://github.com/jmcarrillog/geospatial-etl
[2] https://github.com/jmcarrillog/machine-learning-workflow-for-carbon-assessment

REFERENCES

[1] Udara Willhelm Abeydeera, Lebunu Hewage, Jayantha Wadu Mesthrige, and Tharushi Imalka Samarasinghalage. 2019. Global Research on Carbon Emissions: A Scientometric Review. *Sustainability* 11, 14 (2019), 3972.

[2] Central Intelligence Agency. 2018. The World Factbook: South Sudan. https://www.cia.gov/library/publications/the-world-factbook/geos/od.html

[3] Sabina N Baariu, Galcano C Mulaku, David N Siriba, et al. 2019. State of Cartographic Services among the East African Community Member States. *Journal of Geographic Information System* 11, 01 (2019), 56.

[4] Derik Barseghian, Ilkay Altintas, Matthew B. Jones, Daniel Crawl, Nathan Potter, James Gallagher, Peter Cornillon, Mark Schildhauer, Elizabeth T. Borer, Eric W. Seabloom, and Parviez R. Hosseini. 2010. Workflows and extensions to the Kepler scientific workflow system to support environmental sensor data access and analysis. *Ecological Informatics* 5, 1 (2010), 42 – 50. https://doi.org/10.1016/j.ecoinf.2009.08.008 Special Issue: Advances in environmental information management.

[5] Mariana Belgiu and Lucian Drăguţ. 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing* 114 (2016), 24–31.

[6] Gabriele Cavallaro, Morris Riedel, Matthias Richerzhagen, Jón Atli Benediktsson, and Antonio Plaza. 2015. On understanding big data impacts in remotely sensed image classification using support vector machine methods. *IEEE journal of selected topics in applied earth observations and remote sensing* 8, 10 (2015), 4634–4646.

[7] Joseph David Daron, Kate Sutherland, Christopher Jack, and Bruce C. Hewitson. 2015. The role of regional climate projections in managing complex socio-ecological systems. *Regional Environmental Change* 15, 1 (01 Jan 2015), 1–12. https://doi.org/10.1007/s10113-014-0631-y

[8] Zoe G Davies, Jill L Edmondson, Andreas Heinemeyer, Jonathan R Leake, and Kevin J Gaston. 2011. Mapping an urban ecosystem service: quantifying above-ground carbon storage at a city-wide scale. *Journal of applied ecology* 48, 5 (2011), 1125–1134.

[9] Matthias Drusch, Umberto Del Bello, Sébastien Carlier, Olivier Colin, Veronica Fernandez, Ferran Gascon, Bianca Hoersch, Claudia Isola, Paolo Laberinti, Philippe Martimort, et al. 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote sensing of Environment* 120 (2012), 25–36.

[10] Ottmar Edenhofer. 2015. *Climate change 2014: mitigation of climate change*. Vol. 3. Cambridge University Press.

[11] Simon Eggleston, Leandro Buendia, Kyoko Miwa, Todd Ngara, and Kiyoto Tanabe. 2006. *2006 IPCC guidelines for national greenhouse gas inventories*. Vol. 5. Institute for Global Environmental Strategies Hayama, Japan.

[12] United States Environmental Protection Agency EPA. 2019. Greenhouse Gas Emissions from a Typical Passenger Vehicle. https://www.epa.gov/greenvehicles/greenhouse-gas-emissions-typical-passenger-vehicle

[13] Christiana Figueres, Hans Joachim Schellnhuber, Gail Whiteman, Johan Rockström, Anthony Hobley, and Stefan Rahmstorf. 2017. Three years to safeguard our climate. *Nature News* 546, 7660 (2017), 593.

[14] American Society for Photogrammetric Engineering and Remote Sensing (ASPRS). 2015. ASPRS positional accuracy standards for digital geospatial data (Edition 1, Version 1.0., November, 2014). *Photogrammetric Engineering & Remote Sensing* 81, 3 (2015), A1–A26.

[15] GDAL/OGR contributors. 2019. *GDAL/OGR Geospatial Data Abstraction software Library*. Open Source Geospatial Foundation. https://gdal.org

[16] Yolanda Gil, Ewa Deelman, Ibrahim Demir, Chris Duffy, Marlon Pierce, Suresh Marru, and Gerry Wiener. 2012. Designing a Roadmap for Workflow Cyberinfrastructure in the Geosciences: From Big Data to the Long Tail. American Geophysical Union Fall Meeting, San Francisco, CA.

[17] Yolanda Gil, Varun Ratnakar, Ewa Deelman, and Christopher Mason. 2012. Using Semantic Workflows for Genome-Scale Analysis. International Conference on Intelligent Systems for Molecular Biology (ISMB), Long Beach, CA.

[18] Yolanda Gil, Varun Ratnakar, Jihie Kim, Pedro Antonio Gonzalez-Calero, Paul Groth, Joshua Moody, and Ewa Deelman. 2011. Wings: Intelligent Workflow-Based Design of Computational Experiments. *IEEE Intelligent Systems* 26, 1 (2011). http://www.isi.edu/~gil/papers/gil-etal-ieee-is-11.pdf

[19] Kenneth Gillingham, William Nordhaus, David Anthoff, Geoffrey Blanford, Valentina Bosetti, Peter Christensen, Haewon McJeon, and John Reilly. 2018. Modeling uncertainty in integrated assessment of climate change: A multimodel comparison. *Journal of the Association of Environmental and Resource Economists* 5, 4 (2018), 791–826.

[20] Clair Gough. 2016. *Carbon capture and its storage: an integrated assessment*. Routledge.

[21] Manuel Grizonnet, Julien Michel, Victor Poughon, Jordi Inglada, Mickaël Savinaud, and Rémi Cresson. 2017. Orfeo ToolBox: Open source processing of remote sensing images. *Open Geospatial Data, Software and Standards* 2, 1 (2017), 15.

[22] Mordechai Haklay and Patrick Weber. 2008. Openstreetmap: User-generated street maps. *IEEE Pervasive Computing* 7, 4 (2008), 12–18.

[23] Frank P Incropera. 2016. *Climate change: a wicked problem: complexity and uncertainty at the intersection of science, economics, politics, and human behavior*. Cambridge University Press.

[24] Ayako Kagawa and Guillaume Le Sourda. 2018. Mapping the world: cartographic and geographic visualization by the United Nations Geospatial Information Section (formerly Cartographic Section). In *Proceedings of the ICA*, Vol. 1. 58.

[25] ZW Kundzewicz, V Krysanova, RE Benestad, Ø Hov, M Piniewski, and IM Otto. 2018. Uncertainty in climate change impacts on water resources. *Environmental Science & Policy* 79 (2018), 1–8.

[26] SJ Livesley, EG McPherson, and C Calfapietra. 2016. The urban forest and ecosystem services: impacts on urban water, heat, and pollution cycles at the tree, street, and city scale. *Journal of environmental quality* 45, 1 (2016), 119–124.

[27] Peter M Macharia, Paul O Ouma, Ezekiel G Gogo, Robert W Snow, and Abdisalan M Noor. 2017. Spatial accessibility to basic public health services in South Sudan. *Geospatial health* 12, 1 (2017), 510.

[28] Mark McGovern and Jon Pasher. 2016. Canadian urban tree canopy cover and carbon sequestration status and change 1990–2012. *Urban forestry & urban greening* 20 (2016), 227–232.

[29] Alison M Meadow, Daniel B Ferguson, Zack Guido, Alexandra Horangic, Gigi Owen, and Tamara Wall. 2015. Moving toward the deliberate coproduction of climate science knowledge. *Weather, Climate, and Society* 7, 2 (2015), 179–191.

[30] Ayman Nassif, Ann Stewart, Millie Mutepfe, and Petros Christou. 2016. Infrastructure needs in sub Saharan Africa with particular reference to South Sudan. In *Third Australasia and South-East Asia Structural Engineering and Construction Conference: ASEA-SEC-3*. ISEC Press.

[31] David J Nowak, Eric J Greenfield, Robert E Hoehn, and Elizabeth Lapoint. 2013. Carbon storage and sequestration by trees in urban and community areas of the United States. *Environmental pollution* 178 (2013), 229–236.

[32] Jon Pasher, Mark McGovern, Michael Khoury, and Jason Duffe. 2014. Assessing carbon storage and sequestration by Canada's urban forests using high resolution earth observation data. *Urban forestry & urban greening* 13, 3 (2014), 484–494.

[33] Jim Penman, Michael Gytarsky, Taka Hiraishi, Thelma Krug, Dina Kruger, Riitta Pipatti, Leandro Buendia, Kyoko Miwa, Todd Ngara, Kiyoto Tanabe, et al. 2003. Good practice guidance for land use, land-use change and forestry. *Good practice guidance for land use, land-use change and forestry.* (2003).

[34] Gillian Petrokofsky, Hideki Kanamaru, Frédéric Achard, Scott J Goetz, Hans Joosten, Peter Holmgren, Aleksi Lehtonen, Mary CS Menton, Andrew S Pullin, and Martin Wattenbach. 2012. Comparison of methods for measuring and assessing carbon stocks and carbon stock changes in terrestrial carbon pools. How do the accuracy and precision of current methods compare? A systematic review protocol. *Environmental Evidence* 1, 1 (2012), 6.

[35] Steve M Raciti, Lucy R Hutyra, and Jared D Newell. 2014. Mapping carbon storage in urban trees with multi-source remote sensing data: Relationships between biomass, land use, and demographics in Boston neighborhoods. *Science of the Total Environment* 500 (2014), 72–83.

[36] Victor Francisco Rodriguez-Galiano, Bardan Ghimire, John Rogan, Mario Chica-Olmo, and Juan Pedro Rigol-Sanchez. 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 67 (2012), 93–104.

[37] Johannes Schreyer, Jan Tigges, Tobia Lakes, and Galina Churkina. 2014. Using airborne LiDAR and QuickBird data for modelling urban tree carbon storage and its distributionâA case study of Berlin. *Remote Sensing* 6, 11 (2014), 10636–10655.

[38] Michael Schulson. 2018. Science's Reproducibility Crisis Is Being Used as Political Ammunition. *Wired* (2018). https://www.wired.com/story/sciences-reproducibility-crisis-is-being-used-as-political-ammunition/

[39] Gunnar M Sørbø and Abdel Ghaffar M Ahmed. 2013. *Sudan divided: continuing conflict in a contested state*. Springer.

[40] Ian J. Taylor, Ewa Deelman, Dennis B. Gannon, and Matthew Shields. 2014. *Workflows for e-Science: Scientific Workflows for Grids*. Springer Publishing Company, Incorporated.

[41] Phan Thanh Noi and Martin Kappas. 2018. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors* 18, 1 (2018), 18.

[42] Jan Tigges, Galina Churkina, and Tobia Lakes. 2017. Modeling above-ground carbon storage: a remote sensing approach to derive individual tree species information in urban settings. *Urban ecosystems* 20, 1 (2017), 97–111.

[43] Katherine Wolstencroft, Robert Haines, Donal Fellows, Alan Williams, David Withers, Stuart Owen, Stian Soiland-Reyes, Ian Dunlop, Aleksandra Nenadic, Paul Fisher, et al. 2013. The Taverna workflow suite: designing and executing workflows of Web Services on the desktop, web or in the cloud. *Nucleic acids research* 41, W1 (2013), W557–W561.