Learning Diary - CASA0023

Tongmeng Xie

2023/3/16

Table of contents

Preface

Introduction

Welcome to my learning diary page of Remote Sensing Cities and Environment (CASA0023)! This diary is made for the content taught at 2022-2023.

I'm a current Master of Science student at Bartlett Centre for Advanced Spatial Analysis.

This is a learning diary of a Master of Science student at CASA Module CASA0023 (Remote Sensing City Environment).

This learning diary is presented as a Quarto book containing 9 weeks as chapters.

Each week, the content summarises that week's teaching content in section summary, sometimes it emphasises comprehensiveness and therefore appear in overwhelming length. This way, it is easy for a general reader to get lost in finding what's important. Therefore, I also added guides and highlighted what stood out in importance. Future development part can also assist in grasping the big picture of that week's topic.

Application addresses one (or multiple) literature recommended in the module micro-site. It elaborates on parts that interest me and mentions why they are interesting. Also, the contribution and future literature advancement are highlighted.

As for reflection, I try to relate the content in Summary and Application to wider discipline in regard of how those can be of use in future both from my personal perspective, and in how a spatial data scientist might strengthen their arsenal using the content in dealing with Earth Observation data. The selection of content is largely based on how interesting they are, and hopefully the reason why they appear interesting has been illustrated in the their usefulness.

Take a look at the module's micro-site, to better understand what I'm talking about! CASA0023 Remotely Sensing Cities and Environments (andrewmaclachlan.github.io)

There are two weeks that differ in the general structure of this quarto book: Week 2 Portfolio includes only a Xaringan-made and online-hosted slide, Week 4 Policy deals with a policy instead of papers.

1 Week 1 - Getting started with remote sensing

1.1 Summary

- Data types in remote sensing:
  - Passive data: Energy in electromagnetic form (e.g., human eyes)
  - Active data: Energy in addition to illumination (e.g., radar)
- Interaction of EM waves with Earth's surface and atmosphere:

Interaction Type	Components of Earth	Processes Considered	Difficulty
Single	Surface OR Atmosphere	Direct interaction with one component	Most straightforward to analyze
Dual	Surface AND Atmosphere	Absorption, scattering	Requires atmospheric correction for accurate results
Quad	Surface, Atmosphere, Features	Absorption, scattering, reflection, transmission	More challenging due to multiple components involved

1.1.1 Remote Sensing Data Formats

- Raster formats:
  - BIL
  - BSQ
  - BIP
  - GeoTIFF

1.1.2 Four Resolutions in Remote Sensing

1. Spatial resolution:
    - Ranges from 10 cm to several kilometers
2. Spectral resolution:
    - Number of different spectral bands captured
    - Unique spectral signatures for each feature on Earth
    - Atmospheric windows
    - Vegetation: Red edge - infrared bands
        - Application: Analyze infrared bands in cities to identify access to vegetation
3. Radiometric resolution:
    - Resolution of a cell's value
4. Temporal resolution:
    - Usually inversely related to pixel size (spatial resolution)

- Example satellite sensor: MODIS

1.1.3 Bands Explained

- In remote sensing, bands refer to the specific ranges of wavelengths captured by a sensor.
- Each band captures information about different features on Earth's surface.
- Understanding the properties of each band helps in the interpretation and analysis of remote sensing data.

1.1.4 Future Development

Area of Future Development	Description
Wireless communication	Faster and more secure data transmission using electromagnetic waves.
Medical imaging	New techniques using different types of waves for better diagnoses.
New uses for electromagnetic waves	Discovering new applications in energy, environment, and space exploration.
Remote sensing technology	More detailed data gathering using a wider range of wavelengths.

Area of Future Development	Description
Quantum computing	Developing new technologies that use the properties of particles at the quantum level, including electromagnetic waves.

1.2 Application

I explored Butcher (2016) for a better understanding of electromagnectic waves and the application in Remote Sensing.

Type of Wave	Wavelength Range	Frequency Range	Example Applications
Radio Waves	>1mm	<300 GHz	Broadcasting, communication, radar
Microwaves	1mm - 1m	300 MHz - 300 GHz	Cooking, communication, radar
Infrared Waves	700 nm - 1 mm	300 GHz - 430 THz	Thermal imaging, remote controls
Visible Light	400 nm - 700 nm	430 THz - 750 THz	Human vision, photography
Ultraviolet Waves	10 nm - 400 nm	750 THz -30 PHz	Sterilization, fluorescence microscopy
X-Rays	<10 nm	>30 PHz	Medical imaging, airport security
Gamma Rays	<0.01 nm	>30 EHz	Cancer treatment, nuclear medicine

One of the applications really attracted me was the spatial signature of vegetation on the terra, as we could assign features to each end of the spatial signature area see Figure 1.1, such as bare land on the right end of the triangle-like area where red light captured are dense while near-infrared level is low. Heavy vegetation are witnessed at the upper end of the triangle-like area where red light low and near-infrared is high, indicating heavy biomass. As for the left-down corner where both red and near-infrared are low, we can identify wet lands. This is integrated in the NDVI (Normalized Difference Vegetation Index) to estimate vegetation cover.

Spatial signatures can also be used to monitor the health of vegetation by identifying patterns of quavariation in spectral reflectance that are indicative of stress or disease. For example, vegetation that is stressed or diseased may have a different spectral reflectance signature than healthy vegetation, which can be identified using spatial signatures.

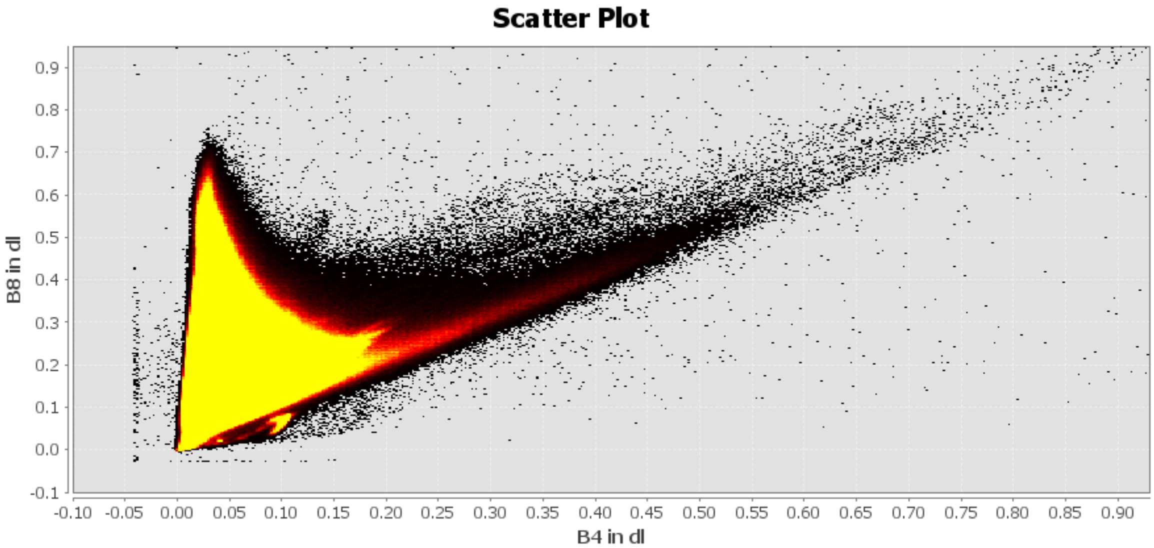


Figure 1.1: Spectral Feature Space, Vegetation On Bands B04 and B08

In addition, spatial signatures can be used to monitor the growth and distribution of vegetation over time by comparing satellite imagery from different dates. This can be useful for understanding the impacts of land use changes, climate change, and other factors on vegetation.

Overall, spatial signatures are a powerful tool for vegetation monitoring, as they can be used to identify and classify different types of vegetation, monitor vegetation health, and track vegetation changes over time.

1.3 Reflection

Having active sensing methods is inspiring as it reminds us that instead of struggling with improving image quality sensed passively using sunlight, we can try altering to artificial signals. Also that sunlight is but another form of electromagnetic wave. This implies a potential to cancel the boundary between natural phenomenon and artificial forms. I feel more encouraged to more actively explore relationships in nature with the devices at hand. I am considering implement active sensors to include SAR data for IoT-based data pipelines. With active sensing data like SAR monitoring forestry real-time, the machine learning model can acquire enough ground truth data from nature for constant adaptation. This could be incorporated into forestry monitoring systems and disaster monitoring systems to cover for both accuracy and rapidness.

One of the challenges I encountered is to navigate the complexities of the interface of SNAP and QGIS. It becomes clear to me that yes implementing several functions in code can be

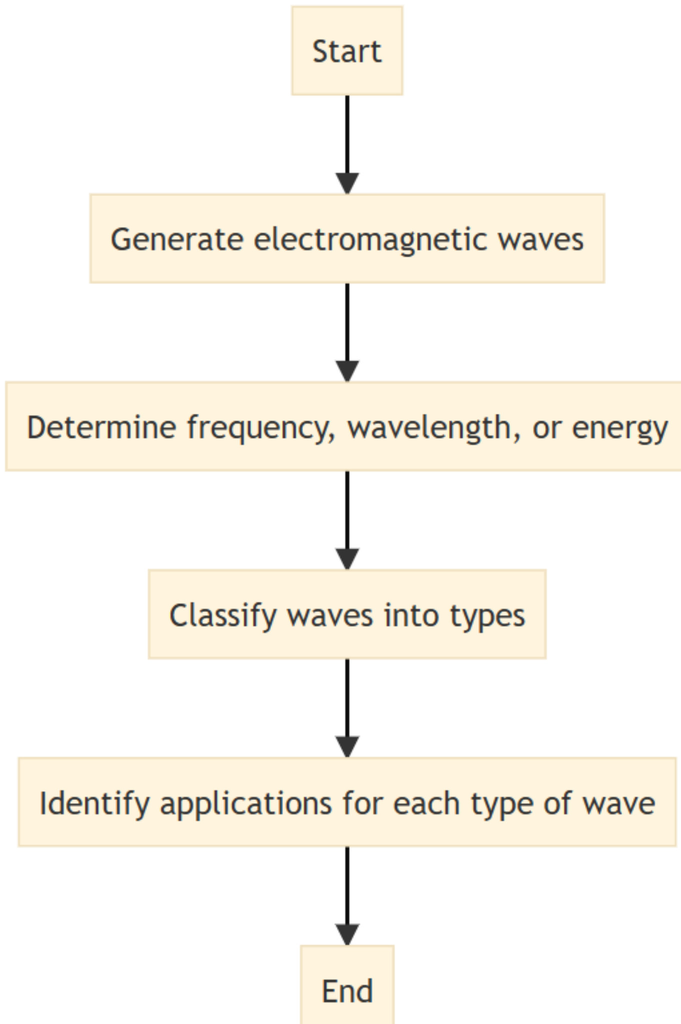


Figure 1.2: workflow of the Electromagnetic Spectrum

challenging, but a software with collective functions as a whole can be mindblowing even when with decent GUIs. Specifically, finding which function falling under which menu consumes a lot of time, and figuring out filling parameters to carry the analysis also took some efforts of iterative validation. Anyway, it's nice to have attempts of designing GUIs for EO data manipulation. Hopefully, with Large Language Model simplifying GIS software designing, we can more easily translate code workflows into user-friendly interfaces and apply our design ideas.

2 Week 2 - Portfolio

Slide about flood prediction and monitoring is here: https://tongmengxie.github.io/Xaringan_slides

2.1 Available instructions

- Introducing Xaringan

3 Week 3 - Remote sensing data

In this week's learning diary, we try to deal with an essential task of remote sensing data workflow, correction

3.1 Summary:

3.1.1 Different Sensors

Across track scanners: Mirror reflects light onto 1 detector. For example, Landsat dataset are captured by this sort

Along track scanners: Basically several detectors pushed along. E.g., Quickbird, SPOT

3.1.2 Geometric Correction

RS data could include image distortions introduced by: View angle, topography, wind and rotation of the earth

We identify Ground Control Points (GCP) in distorted data to match them with local map, correct image, or GPS data from handheld device, but these reference images could also contain distortions and imprecisions.

RMSE is adopted here to measure fitness between images. Use GCPs to minimise RMSE.

Doing geometric correction can shift the original image, so we want to re-sample the final raster by using Nearest Neighbour, Linear, Cubic, Cubic spline re-samplers

3.1.3 Atmosphric Correction

According to Jensen (1986), two factors contribute to environmental attenuation: Atmospheric scattering, topographic attenuation.

There are unnecessary and necessary atmospheric corrections:

necessary ones are:

- Biophysical parameters needed (e.g. temperature, leaf area index, NDVI)
- E.g. .. .NDVI is used in the Africa Famine Early Warning System and Livestock Early Warning System
- Using spectral signatures through time and space

Absorption and scattering can create the haze, i.e. reduces contrast of image.

Scattering can create the "adjacency effect", radiance from pixels nearby mixed into pixel of interest.

3.1.4 Orthorectification Correction

This is a subset of georectification, i.e. giving coords to an image. Particularly Orthorectification means removing distortion so pixels can appear being viewed at nadir (straight down). This requires the support of an Elevation Model to calculate the nadir view for each pixel on a sensor geometry.

To do this: cosine correction, Minnaert correction, Statistical Empirical correction, C Correction (advancing the Cosine). Need radiance (DN to TOA) from sloped terrain, Sun's zenith angle, Sun's incidence angle - cosine of the angle between the solar zenith and the normal line of the slope. Latter two found in angle coefficient files (e.g. Landsat data ANG.txt).

3.1.5 Rdiometric Correction

Corrections to raw satellite imagery can be performed using a method called Dark Object Subtraction (DOS). The logic is that the darkest pixel in the image should be 0 and any value it has is due to the atmosphere. To remove the atmospheric effect, the value from the darkest pixel is subtracted from the rest of the pixels in the image. The calculation involves converting the Digital Number (DN) to radiance, computing the haze value for each band (but not beyond NIR), and subtracting the 1% reflectance value from the radiance. The calculation requires values such as mean exoatmospheric irradiance, solar azimuth, Earth-sun distance, and others, which can be found in sources such as Landsat user manuals.

3.1.6 Joining data sets

Also known as Mosaicking: We feather two images, creating a seamless mosaic, where the diving lien is called seamline.

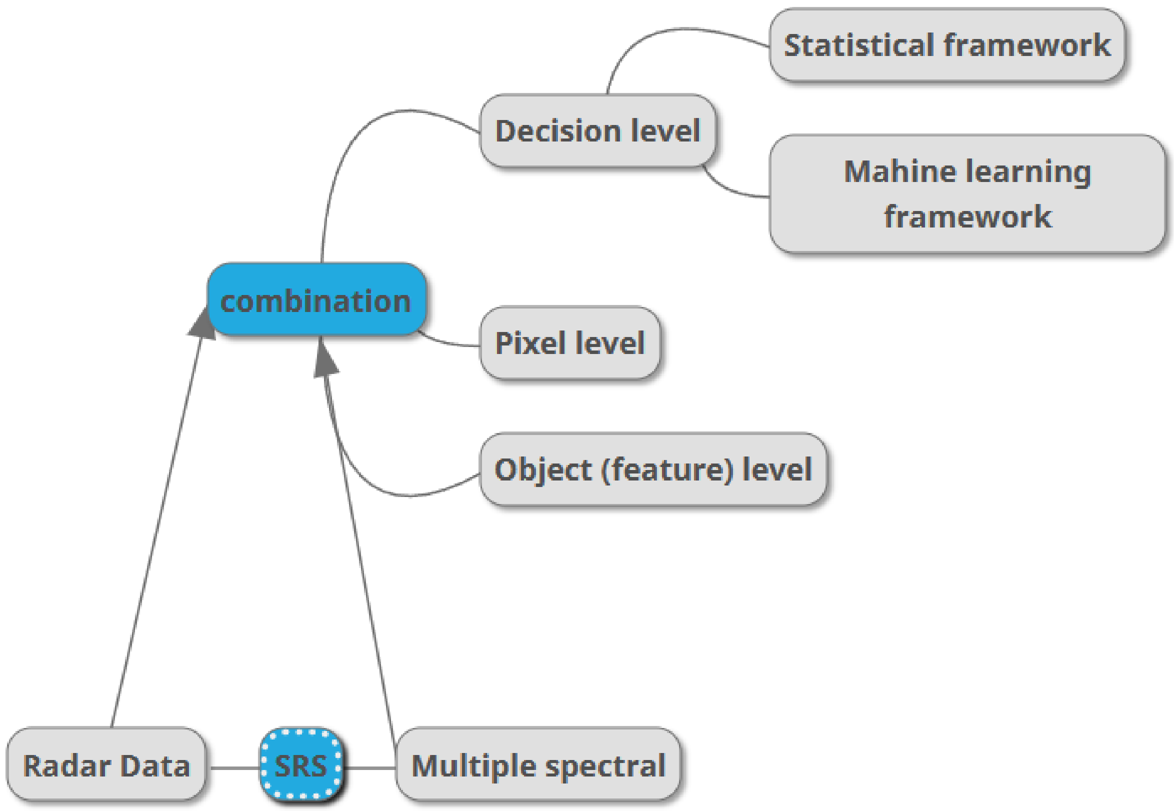
3.1.7 Image Enhancements

Image stretch, Band ratioing, Normalised Burn Ratio, Edge enhancement, Filtering, PCA, Image fusion (see application) etc.

3.1.8 Future development

Type of Correction	Potential Future Developments
Geometric Correction	Automatic GCP identification, more precise resampling methods, handle larger and more complex datasets.
Atmospheric Correction	Incorporate more accurate atmospheric models, improve parameter estimation methods, integrate machine learning algorithms.
Orthorectification Correction	More accurate and efficient algorithms, higher quality elevation models, more advanced image correction techniques.
Radiometric Correction	Develop more advanced methods for parameter estimation, incorporate machine learning algorithms, integrate more detailed and accurate data sources.

3.2 Application - Discussing image fusion in one literature



From literature we delve in the nuances of levels on which we perform image fusion to acquire better results. The integration methods vary as the levels vary (Schulte to Bühne and Pettorelli 2018).

Satellite remote sensing (SRS) can be derived from Multispectral sensors and radar sensors.

Multispectral sensors are passive, merely receiving electromagnetic waves reflected from surface, usually used to reflect chemical properties (such as nitrogen or carbon content and moisture). Usually produces data with comparatively low spatial resolution

Radar ones emit electromagnetic radiation and measure the returning signal, responding to the three-dimensional structure of objects, being sensitive to their orientation, volume and surface roughness. Usually produces data with comparatively high spatial resolution

3.2.1 Image fusion:

1. **decision-level** (SRS integration), where separate predictors are used to estimate a parameter of interest.

2. **object-level (feature-level).** unit: multi-pixel objects. (1) using radar and multispectral imagery is input into an object-based image segmentation algorithm, or (2) segmenting each type of imagery separately before combining them. multi-pixel objects

3. **pixel-level (Observation-level)**, where pixel values are combined to derive a fused image with new pixel values, either in the spatial or the temporal domain.

(2. and 3. derive entirely new predictors.)

Schematic overview of multispectral-radar SRS data fusion techniques. The parameter of interest can be a categorical variable, like land cover, or a continuous variable, like species richness. In pixel-level fusion, the original pixel values of radar and multispectral imagery are combined to yield new, derived pixel values. Object-based fusion refers to (1) using radar and multispectral imagery is input into an object-based image segmentation algorithm, or (2) segmenting each type of imagery separately before combining them. Finally, decision-level fusion corresponds to the process of quantitatively combining multispectral and radar imagery to derive the parameter of interest (by e.g. combining them in a regression model, or classification algorithm)

3.2.2 Implementation Approaches

pixel-level

1. Component substitution techniques: such as principal component analysis (PCA), Intensity-hue-saturation (IHS).

2. PCA is the only pixel-level image fusion technique that cannot be applied to imagery with different spatial resolutions, and the only that allows unlimited image numbers.

3. IHS fusion. Three images with lower spatial resolution (typically multispectral data) are integrated with a single image with high spatial resolution (typically radar) to retain the radiometry but increase the spatial resolution of the former. Facilitate visual interpretation by combining resulting images into a single RGB image.

4. Multi-resolution analysis, such as **Wavelet transformation. Decompose multispectral and radar imagery into their respective low- and high-frequency components

5. Arithmetic fusion techniques: such as the Brovey transform algorithm. Unlikely to be appropriate for multispectral-radar SRS image fusion.

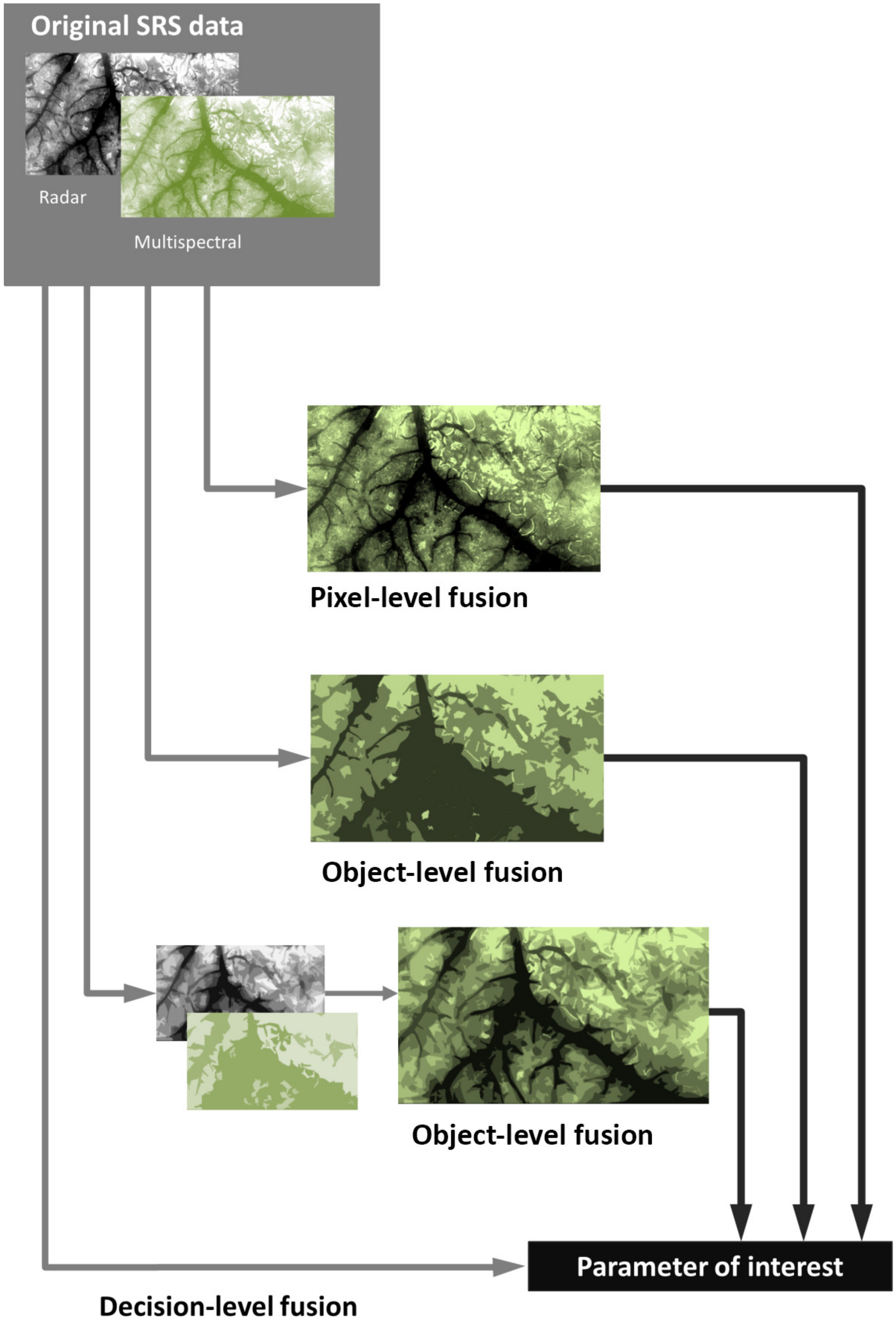


Figure 3.1: Credit: Schulte to Bühne and Pettorelli (2018)

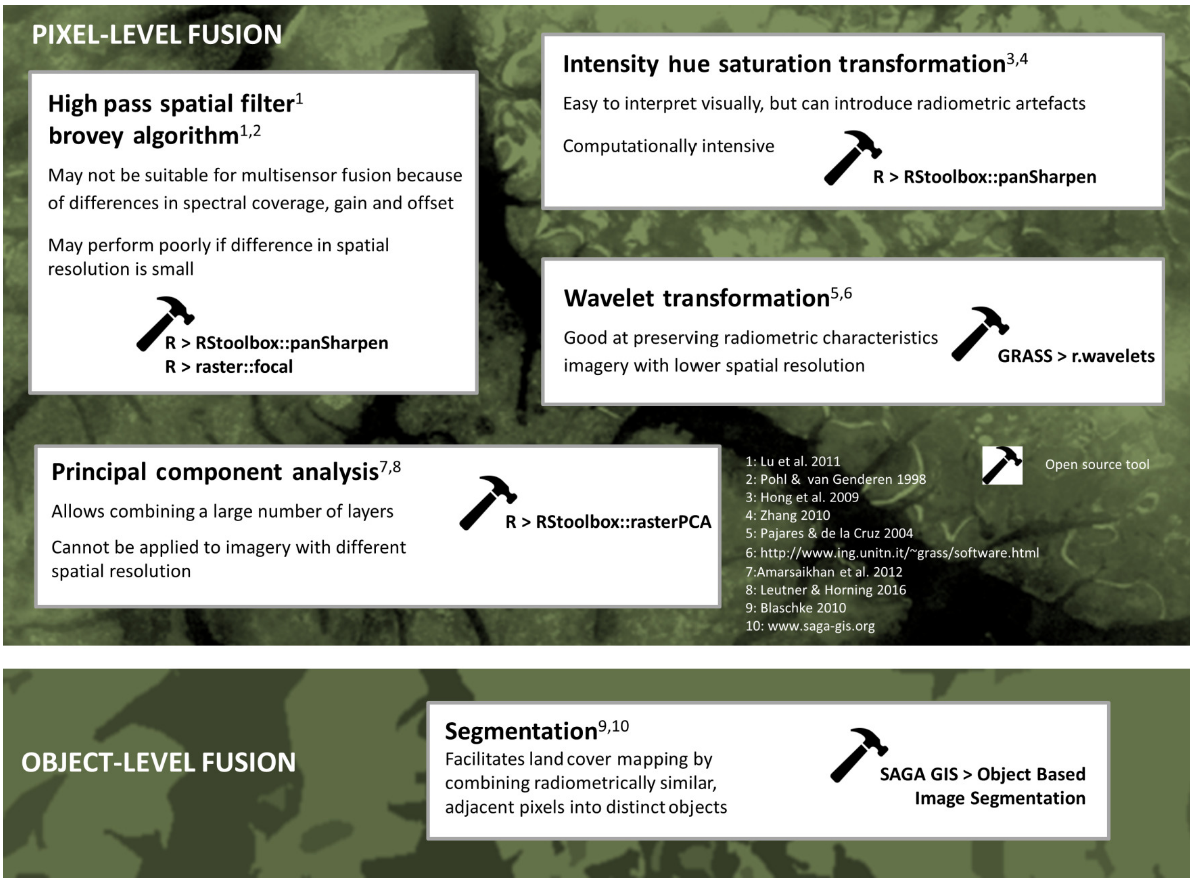


Figure 3.2: Credit: Schulte to Bühne and Pettorelli (2018)

Object-level: Based on brightness and intensity values of each pixel, as well as its spatial context, objects such as lines, shapes or textures are extracted.

1. **image segmentation:** Demands that multispectral and radar SRS images are with the same spatial resolution

2. *extracting objects separately and combining in a feature map*

Object-based fusion reduces all multispectral and radar information into a single layer of discrete objects, which are often relatively easy to relate to ecological features.

Decision-level fusion: Quantitative decision-making frameworks—such as a regression, a quantitative model or a classification algorithm.

3.3 Reflection

Data correction, Data fusion and Image enhancement SRS data fusion can increase the quality of SRS (Satellite Remote sensing)-derived parameters for application in terrain detection, urban analysis, ecology and conservation (Schulte to Bühne and Pettorelli 2018). It is thus important to explore how best to capitalise on recent technological developments and changes in SRS data availability. It is exctiing to apply solid machine learning methods to this area and it is marvelous to see the progress reflected by the increasing number of software supporting this application. The improvement of image quality enables new research designs in ecology and conservation areas and reignite previously greyed-out options.

The application of data correction, data fusion, and image enhancement techniques to SRS data can greatly improve the accuracy and reliability of SRS-derived parameters, which can then be used in various fields, including terrain detection, urban analysis, ecology, and conservation. With the rapid advancements in technology and the increasing availability of SRS data, there is a growing opportunity to leverage the latest machine learning techniques in this area. The development of new software tools to support these applications is a testament to the progress being made in this field. By enhancing the quality of the SRS data, researchers are able to design more robust and informative studies, unlocking new insights and avenues for exploration in ecology and conservation. This, in turn, has the potential to lead to breakthroughs and innovations in these fields, making a significant impact on the world around us.

4 Week4 - Policy Case Study in New York City

OneNYC-2050-Summary.pdf (cityofnewyork.us)

4.1 Summary

4.2 Application

The initiatives provided cover a wide range of areas, including education, small business support, community resilience, infrastructure, transportation, and sustainability. One of the key applications of these initiatives is to improve the quality of life for New Yorkers, particularly those in underrepresented communities. For example, the initiatives aimed at increasing the number of New Yorkers earning a high school equivalency diploma and connecting underrepresented groups to construction jobs created by City investments are designed to provide greater economic opportunities and upward mobility.

Similarly, the initiatives aimed at enhancing walkability and accessibility and improving the sustainability and efficiency of air travel are designed to improve the physical infrastructure of the city and make it more accessible and sustainable for all residents. Another application of these initiatives is to promote equity and inclusivity in the city.

Many of the initiatives are specifically targeted at underrepresented communities, such as the initiatives aimed at providing paid internships and professional development opportunities to cultural workers and supporting the growth and retention of small businesses.

By providing resources and support to these communities, the city aims to promote greater equity and inclusivity and reduce disparities in access to resources and opportunities.

The initiatives also reflect a commitment to sustainability and resilience. Many of the initiatives are aimed at improving the city's infrastructure and transportation systems to make them more sustainable and resilient in the face of climate change and other challenges. For example, the initiatives aimed at investing in innovative and resilient transportation networks and enhancing walkability and accessibility are designed to reduce emissions and congestion and promote sustainable modes of transportation.

Overall, the initiatives outlined in the text reflect a comprehensive and multi-faceted approach to improving life in New York City. While there is still much work to be done, these initiatives

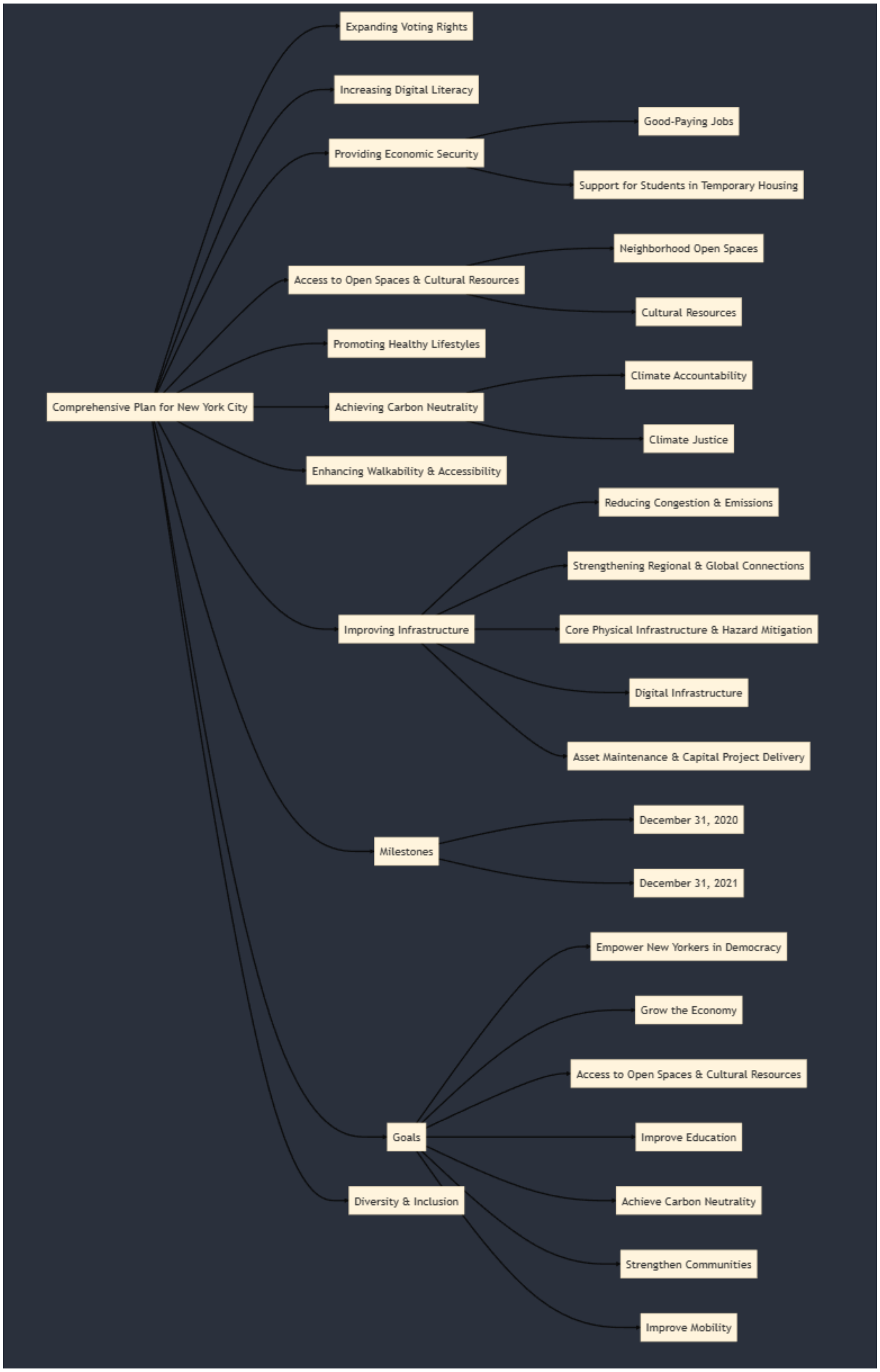


Figure 4.1: workflow of the Electromagnetic Spectrum

represent an important step forward in promoting equity, sustainability, and resilience in the city. As future literature advancements are made, it will be important to continue to evaluate and refine these approaches to ensure that they are effective and responsive to the needs of all New Yorkers.

4.3 Reflection

The initiatives outlined in the text cover critical components of urban planning and development, including infrastructure, transportation, education, and sustainability. The initiatives outlined in the text emphasize the importance of investing in the city's data infrastructure and establishing a citywide data catalog, among other things. This requires a deep understanding of data management and analysis techniques, as well as the ability to work with large and complex datasets. Spatial data scientists and deep learning solution engineers can use these skills to develop innovative solutions for urban planning and development, such as predictive models for traffic flow or energy consumption.

Many of the initiatives outlined in the text involve the use of GIS products and services, as well as the expansion of walkability and accessibility in the city. This requires a deep understanding of geospatial data analysis techniques, as well as the ability to work with mapping and visualization tools. Spatial data scientists and deep learning solution engineers can use these skills to develop innovative solutions for urban planning and development, such as interactive maps that show the most walkable routes in the city.

In addition to these skills, the content, data, and tools presented in the text are highly relevant to the broader discipline of urban planning and development. Spatial data scientists and deep learning solution engineers can use these resources to develop innovative solutions for a wide range of urban challenges, from improving transportation networks to promoting sustainability and resilience. By leveraging these resources, these professionals can help to create more livable and equitable cities that meet the needs of all residents. Looking to the future, it will be important for spatial data scientists and deep learning solution engineers to continue to stay up-to-date with the latest advancements in data management, analysis, and visualization techniques.

5 Week 5 - An introduction to Google Earth Engine

This week introduces **Google Earth Engine (GEE)**, a geospatial processing service that allows for planetary scale analysis of massive datasets in seconds.

Basics:

- The set up of GEE, its terms and jargon, and client vs server side operations, see Table 1
- How GEE uses Javascript and how mapping functions are used instead of loops
- The concept of scale in GEE, which refers to both the volume of analysis and pixel resolution
- How GEE aggregates the image to fit a 256x256 grid.

Objects and methods in GEE are introduced:

- E.g. geometries, features, feature collections, and
- Various data reduction techniques (e.g., reducing images, reducing images by region(s), reducing images by neighborhood).

Also, the types of analyses that can be performed in GEE are briefly covered.

5.1 Summary

Introduced GEE Basics, Objects, Geometries and applications.

Table 1: Terms, Jargon, and Processes Related to Google Earth Engine

Category	Term/Aspect	Definition
Basics	Google Earth Engine	A geospatial processing service that allows geospatial analysis at scale.
Basics	Image	Refers to raster data in GEE and has bands.
Basics	Feature	Refers to vector data in GEE and has geometry and attributes.
Basics	ImageCollection	A stack of images in GEE.

Category	Term/Aspect	Definition
Basics	FeatureCollection	A stack of features (lots of polygons) in GEE.
Basics	Proxy objects	GEE objects that are stored on the server and have no data in the script.
Objects	Earth Engine Objects	Objects in GEE starting with “ee”.
Objects	Images (Rasters)	GEE object representing raster data with bands.
Objects	Feature	GEE object representing vector data with geometry and attributes.
Objects	ImageCollection	A stack of images in GEE.
Objects	FeatureCollection	A stack of features (lots of polygons) in GEE.
Objects	Joins	Combining two FeatureCollections with a common property.
Objects	Arrays	Used to store and manipulate lists of values.
Objects	Chart	Used to visualize data in GEE.
Geometry	Point	A single location represented by its longitude and latitude.
Geometry	Line	A series of connected points representing a linear feature.
Geometry	Polygon	A closed shape with three or more sides, represented by a series of connected lines forming a closed loop.
Geometry	MultiPolygon	A collection of polygons, where each polygon is represented as a list of coordinate tuples defining its vertices.
Geometry	MultiGeometry	A collection of different types of geometries.
Processes	Reducing	Summarizing data over a specified dimension or property.
Processes	Filtering	Reducing data to a specific subset based on a specified condition.
Processes	Mapping	Applying a function to every element of a collection in GEE.
Processes	Scaling	Refers to the pixel resolution in GEE. The scale is set by the output, not the input, and Earth Engine selects the pyramid with the closest scale to analysis.
Client/Server	Client Side	Refers to the browser side of GEE.
Client/Server	Server Side	Refers to the side of GEE where data is stored.
Client/Server	Looping	Looping is not recommended for objects on the server side.
Client/Server	Mapping	Instead of loops, mapping is used in GEE to apply a function to everything on the server.

5.1.1 GEE Basics

JavaScript, where objects are dictionaries:

- We have ee (EarthEngine), a powerful package. Anything starting with ee (proxy objects)

are stored on the server.
- Problems:
  - We don't iterate the data on the server; instead, we map (using a mapping function) them into objects (variables) so we only load them once.
  - There are also some sort of server-wide functions.
  - Avoid using loops in GEE on the server-side, as mapping can automatically detect the number of loops needed.

Scale:

- Pixel resolution, set by the output.
- GEE does resampling, aggregating your input to a 256*256, mainly down-sampling.

5.1.2 GEE Objects

Objects:

- Images (Rasters), geometry, ImageCol, features, featureCol, joins, arrays, chart.

5.1.3 GEE Processes and Applications/Outputs

GEE applications:

- Reducing types.
- Different to filterBounds() that filters the area of interest, to do zonal statistics, we have reduceRegion(), where regions are subcategories of the area of interest.
- Also, we have reduceNeighborhood(), which is a bit like a kind of image enhancement.

Linear Regressions:

- In a scenario of visualising precipitation, we can do a multivariate multiple linear regression where both independent variables (time) and dependent (precip, temp) variables are multiple.
- Something about constant bound.

Joins:

- In GEE, everything, e.g. within a buffer, intersect, etc. needs the mediation of Join (apply()).
- To perform joins, we need to put data into Filter().

Classifiers:

- Per-pixel

- sub-pixel

Table 2: GEE Processes and Applications/Outputs

Process	Description
Geometry operations	Spatial operations such as union, intersection, buffer, and distance analysis
Joins	Combining two feature collections based on a shared attribute value
Zonal statistics	Computing statistics for a region or set of regions such as mean, median, and mode of pixel values within a feature or a collection of features
Filtering	Filtering of images or specific values based on criteria such as date range, location, and attribute value
Machine learning	Using statistical and machine learning algorithms for classification, clustering, and prediction tasks
Deep learning	A subset of machine, using Deep Neural Networks

5.1.4 Advantages and Limitations

Pros	Cons
1. Large-scale data processing	1. Limited to Google's data catalog
2. Access to vast satellite imagery library	2. Steeper learning curve for beginners
3. Real-time data analysis capabilities	3. Requires coding skills (JavaScript, Python)
4. Cloud-based platform	4. Limited customization options
5. Free for non-commercial use	5. Data export restrictions
6. Easy data sharing and collaboration	6. Dependent on internet connectivity

*No support for phase data, needs SNAP.

5.1.5 Trend

See also Section Application for details and references.

1. Enhancement user interface: GEE might introduce a more user-friendly interface to lower the entry barrier for beginners and non-programmers, making it more accessible to a wider audience.
2. Integration with machine learning and AI: GEE could expand its integration with advanced machine learning and AI algorithms, enabling users to derive more sophisticated insights from geospatial data.

3. Customisable solutions: GEE may introduce more customization options for users, allowing them to develop tailored geospatial analysis tools and applications.
4. Better support for commercial use: GEE could offer more comprehensive support and licensing options for commercial users, helping businesses harness the full potential of geospatial data analysis.

5.2 Application

Literature choice: Gorelick et al. (2017).

This week's recommended literature mainly are documentation support for GEE and literature, even papers. Therefore the contribution of the literature will be in more general senses. An overview of Google Earth Engine's capabilities and applications, as well as its potential to address societal issues.

They also discuss potential future developments, including expanding Earth Engine's data catalog, improving its user interface, and increasing collaboration with other organizations.

- Expanding Earth Engine's data catalog: currently includes a wide range of geospatial datasets but could be expanded to include additional sources of data (Gorelick et al. 2017).
- Improving the user interface: make it more intuitive and user-friendly, particularly for non-expert users.
- Increasing collaboration with other organizations: Collaboration with other organizations, both in terms of data sharing and joint research projects, is also an important area for future development.
- Ongoing research into new analysis techniques and algorithms: Ongoing research into new analysis techniques and algorithms will continue to expand Earth Engine's capabilities and applications (Moore and Hansen 2011).

5.3 Reflection

GEE-using skills can be a valuable asset for a spatial data scientist, as it allows for complex spatial analysis at scale. Traditional GIS software are eclipsed when it comes to both efficiency and scale.

GEE's unique and efficient way of conducting analysis flows is interesting, such as the introduction of concepts like client vs server-side operations and data reduction techniques. These was required by GEE's feature of carrying out analyses on massive datasets (Gorelick et al. 2017). For those interested in BigData technology, the strategies (server/client split, no looping on server, etc.) applied by Google here is a very resourceful one and worth learning. The user end also has to learn to adopt good practices for reducing data range, which has been simplified to a series of reduction and filtering functions, e.g. `ImageCollection.filterDate()`, `image.reduceNeighborhood()`(Google 2023b).

GEE's combination with machine learning is also promising in regard of automating complex analysis tasks, as Machine Learning APIs offered by GEE support Supervised and Unsupervised Classification, and Regression (Google 2023a). According to Saad El Imanni et al. (2023), as a subtask of intelligent agriculture, weeds detection task sees an impressive performance (overall accuracy reached 96.87%) when GEE and Machine learning are combined.

In conclusion, I believe familiarity with GEE will add to one's machine learning workflow in dealing with EO data, and, more generally, incredibly large datasets. The design of GEE also opens an era of web-service based big-data handling. Its designs in alleviating computation on client side and getting rid of for-loop on server-side inspires service designers to make distinct standards for code practice based on the server-client split. Besides, the sheer amount and diverse categories of data available on GEE saves experts from burdensome data collection process, so they can focus more on EO data processing, analysis and storytelling.

6 Wk6 Classification

6.1 Summary

Information	Summary
Purpose of classification	To subset data into classes or values, such as landcover or estimating values like GCSE scores or pollution.
Different classification methods	Essentially slice the data in different ways.
Complexity of classification methods	They can often be made to appear more complicated than they are.
Controlling classifiers	Can be done using hyperparameters.
Desired outcome of classifiers	Can range from a single tree to a decision hyperplane boundary in multiple dimensions.

6.1.1 ML methods in EO data classification

Table 1: Supervised Classification Methods

Method	Description
Maximum Likelihood	A statistical method used to estimate the parameters of a probability distribution based on observed data.
Support Vector Machines (SVM)	A supervised learning algorithm that finds the best hyperplane to separate data into different classes.

Table 2: Unsupervised Classification Methods

Method	Description
Density Slicing	Divides the range of pixel values into equal intervals and assigns each interval a unique class value.
Parallelpiped	Uses a set of user-defined ranges for each band to define class boundaries in multi-dimensional space.

Method	Description
Minimum Distance to Mean	Assigns each pixel to the class with the closest mean value in multi-dimensional space.
Nearest Neighbor	Assigns each pixel to the class of its nearest neighbor in multi-dimensional space.

Table 3: Other Machine Learning Methods

Method	Description
Artificial Neural Networks (ANN)	A set of algorithms inspired by the structure and function of biological neural networks, used for pattern recognition and prediction tasks.

6.1.2 Pros and cons - Supervised vs. Unsupervised

Table 1: Supervised vs. Unsupervised Classification

Classification Type	Definition	Method
Supervised	Classifier learns patterns in the data and uses that to place labels onto new data. Pattern vector is used to classify the image. Usually, pixels are treated in isolation but as we have seen - contextual (neighboring pixels), objects (polygons), texture.	Pattern recognition or machine learning
Unsupervised	Identifies land cover classes that aren't known a priori (before) and tells the computer to cluster based on info it has (e.g. bands) and label the clusters.	Density slicing, parallelpiped, minimum distance to mean, nearest neighbor, neural networks, machine learning / expert systems*

6.1.3 Overfitting

- Bias refers to the difference between the predicted value and the true value. When a model has high bias, it is too simple and may underfit the data. On the other hand, when a model has low bias, it may overfit the data.

- Variance, on the other hand, refers to the variability of a model for a given point. When a model has high variance, it is too complex and may overfit the data. This means that it will perform well on the training data but poorly on new data.

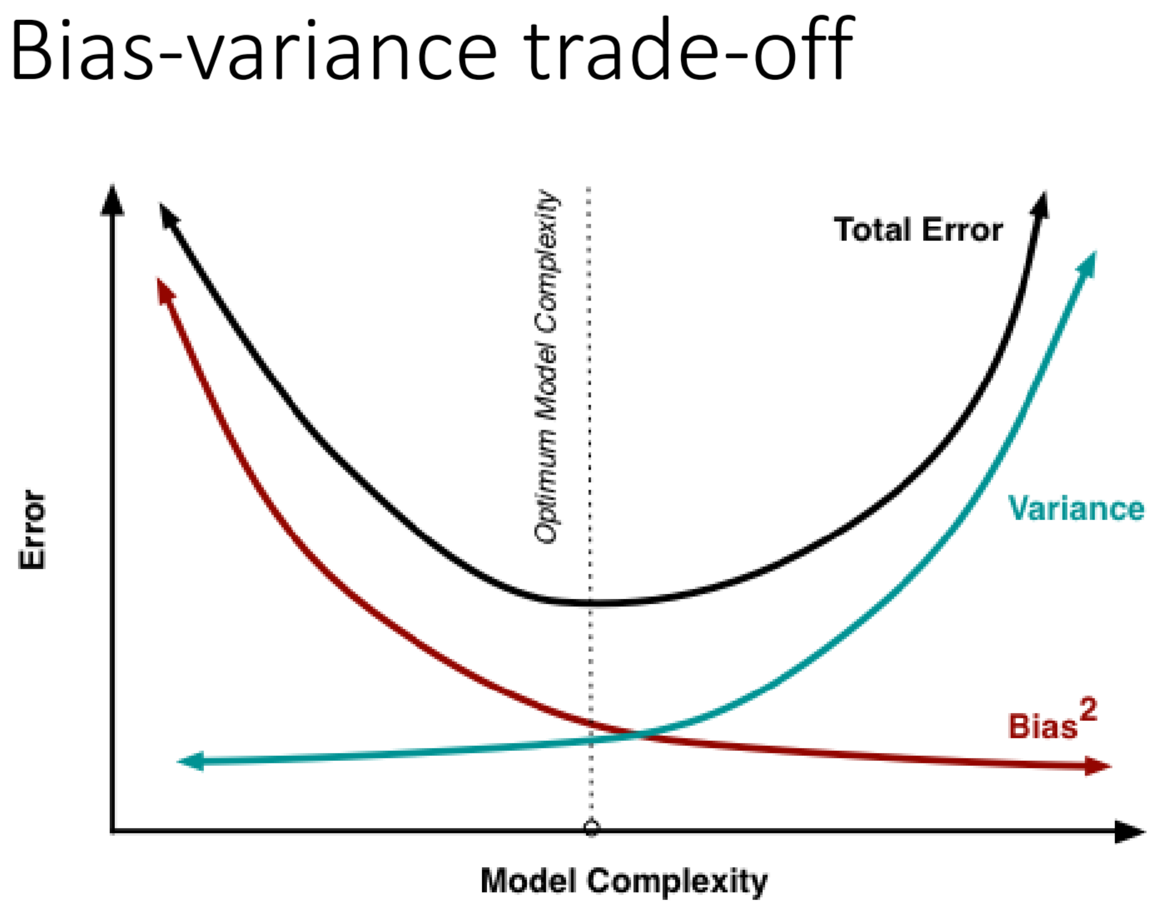


Figure 6.1: Credit: CASA0006

In general, overfitting occurs when there is a trade-off between bias and variance. A model with high bias and low variance will underfit the data, while a model with low bias and high variance will overfit the data. The goal is to find a balance between bias and variance that results in good performance on both training and test data.

6.1.4 Outlook on the development of EO data Classification

Earth Observation (EO) data classification is continually evolving, with new technologies and techniques leading to several anticipated future developments:

Error_diff = testing error – training error

Training error \ Error_diff	Low	High
Low	Low-bias & low-variance Good balance	Low-bias & high-variance Overfitting
High	High-bias & low-variance Underfitting	*Improvement needed*

(optional) However, it is INCORRECT to say:
- Training error is an estimate of bias
- error_diff is an estimate of variance.

19

Figure 6.2: Credit: CASA0006

1. Multi-source data fusion:

    - Integrating data from multiple sources like satellite imagery, LiDAR, and ground-based sensors will become more prevalent. This fusion enhances classification accuracy and offers comprehensive Earth's surface information, improving decision-making and monitoring. For example, the European Union's **Copernicus Programme** could use this in providing free data from various satellite missions and sensors for environmental monitoring, disaster management, and urban planning.

2. Multi-temporal analysis:

    - Sophisticated multi-temporal analysis techniques will be increasingly used to monitor changes in land cover, vegetation, and other features over time. This enables accurate and efficient change detection and monitoring of phenomena like urbanization, deforestation, and climate change. This aligns with the **REDD+** initiative under the United Nations Framework Convention on Climate Change (UNFCCC), which uses multi-temporal analysis to monitor forest cover changes and evaluate policy effectiveness in reducing greenhouse gas emissions from deforestation and forest degradation.

3. Cloud-based processing:

    - The growth of cloud-based platforms, such as Google Earth Engine, allows for more efficient and scalable EO data processing workflows. This accessibility enables researchers and organizations to innovate in classification techniques and applications. The National Oceanic and Atmospheric Administration's (NOAA) **Big Data Project** aims to make vast amounts of environmental data accessible and usable in the cloud, fostering innovation in developing new applications and services.

These advancements will provide comprehensive and timely information about Earth's surface, informing policies and strategies in areas such as environmental management, disaster response, and urban planning.

6.2 Application - ****Support vector machines for classification in remote sensing****

Deep Learning methods can have universally good performance across Computer Vision tasks, e.g. Earth Observation data classification, not to mention techniques like transfer learning (pre-trained model plus large data for a specific task) and meta learning (large universal pretrained model plus small amount of task-specific data) can further strengthen the accuracy.

However, in this week's literature [https://www.notion.so/Wk6-Classification-98918c59eebc4b869f45ec22d152965
SVM was demonstrated to generate better result with smaller data amount.

This vastly contributes to the particular task of remote sensing image classification. Also, we can derive insights of how elegant choice of model (SVM in this case) for downstream tasks can outperform blindly stacking (make ensemble of) popular neural networks.

6.2.1 Support Vector Machine

This model basically attempts at maximising margins of fitting lines that are trying to classify points. To do this, it finds an optimal hyperplane ("lines" extended in dimensionality). In the sense that it projects data into higher dimensions, it resembles kernel methods. Some even categorise SVM as one of kernel methods.

Oh, higher dimensions! Sounds computation-intense? But this is already an alleviation of computation compared to Neural Networks.

Besides, the amount of required data and scale of model weights are severely reduced, making it easier for both algorithm engineers to train and Remote Sensing experts to use.

6.2.2 Rationale Behind the Paper

The paper here deals with small amount of data, with ground truth selected using a random sampling procedure. To effectively use small data, it adopts SVM and achieved high classification accuracy with high dimensional data.

The author also delves into the detailed problems encountered using SVM. For classification task, two-class or multi-class problems are separately discussed. Usually, Earth Observation data falls within the multi-class one. The 'one against one' and the 'one against the rest' strategies for generating multi-class SVMs are compared in this study. The "one-against-one" method proved to be better for multi-class.

Remote-sensing data often have different spectral, spatial and temporal resolutions, which pose challenges for traditional classification methods. SVM can overcome these challenges by mapping the data into a higher-dimensional feature space where a linear separator can be found. This way, It can

- help identify and map different land cover types and changes over time
- assist in monitoring and managing natural resources, such as forests, water, soil, etc.
- provide valuable information for disaster management, such as flood detection, fire risk assessment, landslide susceptibility, etc.
- support various applications in agriculture, urban planning, climate change studies, biodiversity conservation, etc.
- reduce the cost and time of field surveys and data collection

6.2.3 Future Advancement

- Combing ANN and SVM:
  - A hybrid regression support vector machine-convolutional neural network (HRSVM-CNN) classifier can be used for object-based classification of high-resolution remote sensing images [Full article: Object based classification of high resolution remote sensing image using HRSVM-CNN classifier (tandfonline.com)]. In this approach, the image data is preprocessed and segmented, and then features are extracted using techniques such as local ternary patterns (LTrP), color histograms, gray-level co-occurrence matrices (GLCM), gray-level difference method (GLDM), edge features, and shape features. These extracted features are then classified using the HRSVM-CNN classifier.
- SVM can go further in its advantages:
  - Transfer learning involves using a pre-trained model that has been trained on a large dataset to extract features from the data. These extracted features can then be used as input to an SVM classifier trained on a smaller dataset [TL-SVM: A transfer learning algorithm | Semantic Scholar]. This approach can take advantage of the ability of the pre-trained model to learn complex representations of the data and the ability of SVMs to find good decision boundaries even when the data is not linearly separable.
  - Active learning, which involves iteratively selecting the most informative samples from a pool of unlabeled data and adding them to the training set. This can help to improve the performance of an SVM classifier even when the amount of labeled data is very limited.

6.3 Reflection

This week, I have been absent from the lecture and workshop, because I was at Data Dive CUSP London, which is quite an opportunity for meeting people in data science sector and honing skills of utilising data science skills to tell a story addressing real-world problems. My group explored “Does built-environment have an influence on Londoners’ mental health”, where we tried to utilise latest deep learning methods like attention mechanism and U-map for analysis and visualisation for a high ‘technical complexity’ mark.

I have been obsessed with Neural Networks during my undergraduate years: How can I distill this General Pretrained Model to be locally implementable to be my personal poem-composing assistant? How can I deploy this open-source Object Detection model on an ARM (Advanced RISC Machines) built in an IoT camera to detect high-street footfall? But the SVM (Support vector machines) introduced in this week’s literature really proves how simpler models (without neural classifiers @Benediktsson *et al.*, Tso and Mather) can achieve performance no worse than the SOTA Deep Learning models in specific tasks like ****classification in remote sensing****.

This insights, elegant choice of model (SVM in this case) for downstream tasks can outperform blindly stacking (make ensemble of) popular neural networks, unveil new potentials for me when optimsing model performance, e.g., when doing transfer learning on a pretrained model on a classification task, I might consider experimenting with SVM in parallel with hyper-tuning Neural network, and seek possibilities of combining the two.

Despite the usefulness of this diversity of classification models, always

- Pay attention to their assumptions,
- Check carefully if our data and problem align with these assumptions.
  - If not, process accordingly to satisfy them or switch methods.

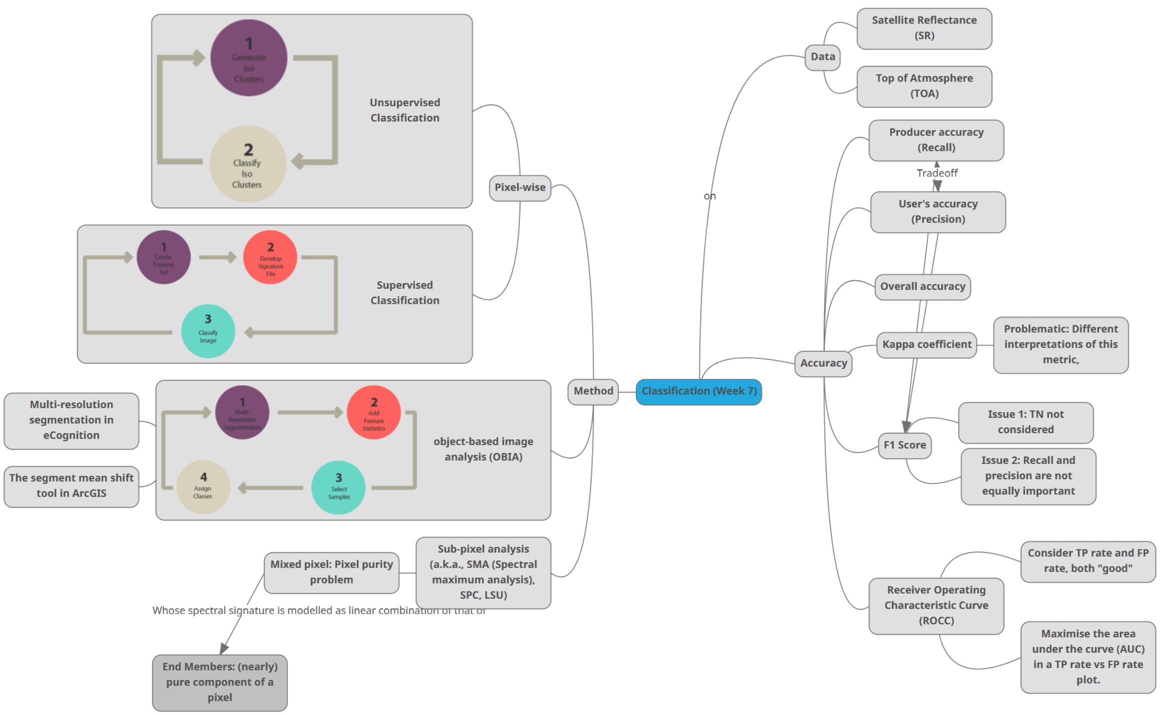
Especially, when combining different Machine Learning methods, e.g. in module GISS(CASA0005), we used the result of KNN models to decide parameters (min_point and radius) for DBSCAN, always look into data to ensure the assumptions are met. The disparity in alignment with model assumptions can have impact on the whole data pipeline.

7 Week7 - Classification and Accuracy

This week's learning diary continues that from Week 6 in addressing the big problem in Remote Sensingm, i.e. classification within Earth Observation data. Also, accuracy metrics are discussed.

7.1 Summary

The summary of lecture content as well as practical outcomes. See **?@fig-mindmap** for an overview. If certain words are intelligible due to resolution issues, hopefully you can right click and "open in new page" to get a better view since this is a .SVG file.



7.1.1 Data

- Surface Reflectance (SR)
- Top of Atmosphere (TOA)

A mixed way of doing urban recognition

7.1.2 OBIA (object-based image analysis)

Instead of a per-pixel approach, we adopt an object-based image analysis (OBIA), where you have to manually create objects.

SLIC (***Simple Linear Iterative Clustering***) (2012): No ground truth

Descent, similarity (Homogeneity)

7.1.3 Sub-pixel analysis

SMA (Spectral maximum analysis), SPC, LSU

Through a series of manipulation of material, we acquire a list describing the broken-down land cover of that pixel

7.1.3.1 Pixel purity

Endmember: an important concept in spectral mixture analysis

In remote sensing, an end member refers to a pure or nearly pure material or component that is present within a mixed pixel.

In spectral mixture analysis, the spectral signature of a mixed pixel is modelled as a linear combination of the spectral signatures of the constituent endmembers, with each end member being assigned a proportion or fraction that represents its contribution to the overall reflectance or radiance of the mixed pixel.

7.1.4 Accuracy assessment

- Producer accuracy: Recall
- User's accuracy: Precision
- Overall accuracy: not equivalent to F1
- Kappa coefficient: [0, 1], measures how good the classification is compared to random distribution e.g. Poisson. Different interpretations of this metric, problematic

Make a tradeoff between Producer accuracy and User accuracy, by shifting the decision boundary.

- F1: issue: TN not considered; Recall and precision are not equally important yet equally weighted in F1
- Receiver Operating Characteristic Curve: True positive rate and false positive rate are all good. We want to maximise the area under the curve in a True positive rate vs false positive rate plot.

7.1.5 Workflow

- (Potentially use unsupervised classification to understand your data
- Class definition (Potentially use unsupervised classification)
- Preprocessing
- Training
- Pixel Assignment
- Accuracy assessment

Pseudo-invariant features to be trained on to make your model robust to time-space changes

Pseudo-invariant features are often used as reference targets or calibration sites in remote sensing to account for changes in sensor or atmospheric conditions and to reduce the effects of noise and calibration drift on image data. These features have relatively constant spectral properties over time and space, and can therefore serve as a stable reference for monitoring changes in other features or materials within an image or scene.

A flow chart can be seen in Figure 7.1 :

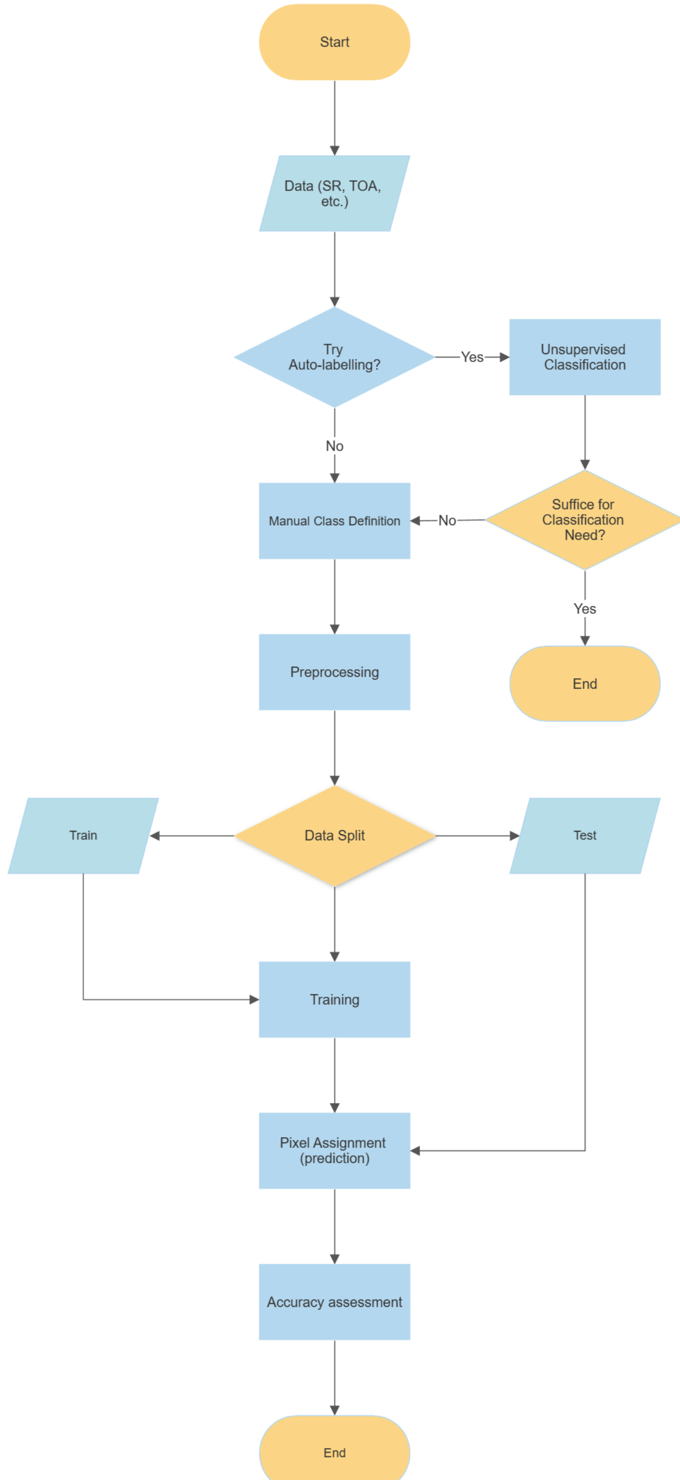


Figure 7.1: Classification Workflow, courtesy: myself

7.1.6 A "Sneak preview" (Analogous to Data Leakage in ML)

Waldo Tobler's first law of geography indicates that if training and testing are spatially close, the training can cause the problem of a sneak preview.

7.1.6.1 Spatial Cross Validation

Similar to cross-validation but adds clustering to folds.

In spatial cross-validation, the data are split into spatially contiguous blocks or subsets, rather than randomly shuffled subsets as in traditional cross-validation. This is done to ensure that the model is tested on data that are spatially distinct from the data used to train the model and to account for spatial autocorrelation and other spatial dependencies in the data.

7.1.7 Approaches to deal with Spatial Autocorrelation

Object-based image classification

Moran's I (Spatial Cross Validation)

7.2 Application - to be completed

In remote sensing, it is often challenging to accurately classify mixed pixels, which contain a combination of different materials or components. Endmembers refer to pure or nearly pure materials that are present within a mixed pixel. By modelling the spectral signature of a mixed pixel as a linear combination of the spectral signatures of the constituent endmembers, we can determine the contribution of each endmember to the overall reflectance or radiance of the mixed pixel.

This approach can be very useful in urban recognition, where it is essential to accurately classify the different land covers present within a pixel. Furthermore, it can also help us to understand the composition of the land cover in a given area, which can have important implications for environmental monitoring and management. This approach has been applied in various studies to estimate urban land cover, such as the work by Zhang et al. (2018) that utilised endmember extraction to detect urban impervious surfaces.

7.3 Reflection

The workflow of Classification of Surface Reflectance and Top of Atmosphere data intrigues me, as it differs from, yet shares certain features with traditional computer vision tasks like image classification and object detection (in regards of treating pixels as objects/units). Alternatively, Surface Reflectance Classification can be treated as a downstream task for both aforementioned ML tasks, due to its uniqueness in dealing with high-precision satellite data and unseparability (worth debating) from EO processing workflow (calibration etc.). Also, uncertainty of classification genres derived from unsupervised labelling can also be an issue.

Optionally, in supplementation to manual labelling, automated labelling workflow (e.g. roboflow) can be introduced to curtail repeatitive works in image labelling(Nair, Paul, and Jacob 2018). However, manual labelling is not replaceable at current time (Robison 2018) and the pinning down of ground truth seems to always need human intervention in addition to machine automation.

The introduction of production accuracy and user accuracy is also interesting, as these terminologies are designed presupposing a customer/producer split, dwarfing precision-recall in readability. The treadeoff to be made between the two is crucial, and this is problematically handled by introducing F1 score with two competitive components, and improved by introducing Receiver Operating Characteristic Curve (ROCC) with two "good" indicators, true positive rate and false positive rate.

8 Week8 - Temperature and Policy

The "policy" section occupies two weeks, mainly trying to introduce how to fit EO data workflow into current policies. To do this, you have to identify the gaps, e.g., that between the overarching global policies, metropolitan plans and local plans. Or, the gap within policies like missing locations in the Singapore one.

8.1 Summary

8.1.1 Urban Heating Islands (UHI) problem and plans

8.1.1.1 Causes

Urban areas have comparatively higher atmospheric and surface temperatures than surrounding rural areas, mainly due to

1. More dark surfaces that retain heat
2. Less vegetation that cools the environment

Also, there are other contributors to the heat:

Contributor	Correlation with UHI phenomenon
Sky View Factor (SVF)	Positive
Air speed	Negative
Heavy cloud cover	Positive
Cyclic solar radiation	Positive
Building material type	Varies
Anthropogenic energy	Positive

8.1.1.2 Cost

The cost of the Urban Heat Island can be divided into social, environmental, and economic costs.

Type of Cost	Examples
Social Costs	Population-adjusted excess mortality rates, heat-related deaths
Environmental Costs	Increase in fossil fuel usage
Economic Costs	Loss of Gross Domestic Product (GDP)

For instance, under a low greenhouse gas scenario, the percent GDP lost from UHI is estimated to be 0.71% in 2050.

8.1.1.3 Plans

8.1.1.3.1 Global

8.1.1.3.2 Local

City	Initiatives	Technology
Singapore	Green buildings	Urban greenery
Medellin	Green Corridors	Urban greenery
Sydney	Turn Down the Heat Strategy and Action Plan	Reflective roofs/pavements/sidewalks, Cool roads trial in Western Sydney

8.2 Application

MacLachlan et al. (2021) The document outlines a sub-city urban planning modeling approach using open-source tools to measure and monitor localised urban heat island (UHI) mitigation targets. The methodology involves comparing temperature dynamics of low- and high-density census areas using Earth observation data and determining optimal placement of greening elements in proposed plans using a data-driven model. The document concludes that this approach can be universally integrated into urban planning regulation frameworks to mitigate the localized UHI effect and ensure long-term city sustainability. Also it discusses the impact of low population density on housing in Perth, Australia, and the resulting need for strategic land zonation and sustainability targets. #### Why Data-driven approach

8.2.1 Policy limitations

- lacking **specificities** for combating adverse temperature effects at the local level (**sub-city**), therefore not planning practicality

- no **consistency** in planning implementation **methodologies** or steps for **assessing** progress toward UHI reduction targets
- lackage of empirical **evidence for optimizing** UHI mitigation strategies

8.2.2 Data to drive the new approach

- Earth Observation (**EO**) data can be processed to identify (un)sustainable urban development through aerial assessments of **land cover change**
- temperature
- elevation

8.2.3 Advantages for the Data-driven approach

- EO data can produce **consistent information** necessary for restricting unsustainable development
- **monitor** UHI effects based on associated land-temperature dynamics
- Assessment at finer spatial scales (e.g., block subdivisions)

8.2.4 Methodology

8.2.4.1 Temperature Modeling

- Modeled temperature every 3 hours using SOLWEIG model between 2008 and 2010

- Inputs generated from meteorological data, land cover, building DSM, ground DTM, and vegetation canopy REM in QGIS

- SVF computed from vegetation canopy REM, building DSM, and ground DTM using UMEP plugin

---

- Building wall heights and aspect generated from DSM and DTM using UMEP plugin

---

8.2.4.2 Data-Driven Tree Placement

- Site selected for modeling temperature in the City of Fremantle

---

- Three scenarios processed: current urban footprint, proposed changes, and proposed redevelopment with no trees

---

- Highest temperatures identified and used to redesign tree placement

---

- 15 trees distributed according to original design aspects

- Updated vegetation canopy REM reflected new tree locations

- Analysis re-run to compare temperature across redevelopment site

---

- Modeled all scenarios accounting for influence of neighboring landscape features

---

Factors assessed	Effect on UHI effect

8.2.5 Result

Table 1: Results of assessments of urban design factors on UHI effect in Perth.

Factors assessed	Effect on UHI effect
Vegetation cover	Negative correlation with UHI effect
Canopy cover	Negative correlation with UHI effect
Building density	Positive correlation with UHI effect
Building height	Positive correlation with UHI effect
Albedo	Negative correlation with UHI effect
Land use	Negative correlation with UHI effect
Urban sprawl	Positive correlation with UHI effect

Table 2: Total population and population density per 0.1 km2 between 2011 and 2016 for Subiaco and Currambine SA1s as defined by the ABS.

SA1	**Total Population (2011)**	**Total Population (2016)**	**Population Density per 0.1 km2 (2011)**	**Population Density per 0.1 km2 (2016)**
Subiaco	N/A	N/A	325	712
Currambine	N/A	N/A	310	324

8.3 Reflection

Basically about pedestrian economy. Though there have been many retail modes like KFC and other American fast food that favour more cars, the experience from Europe tells that economy vitality can have a boost with pedestrian-dominated areas (Case study: Superblocks, Barcelona). Besides the interesting economic insights, this week I learned how to bring my focus upon specific areas in urban policies like UHI. Also, it seems to be good practice to narrow down one's scope from global to local, and finally to a specific city policy, identifying gaps in each layer, and finally introducing Earth Observation data flow.

9 Week 9 - Synthetic Aperture Radar (SAR) data

9.1 Summary

9.1.1 A quick overview

This week addresses problems in

- The object of using Synthetic Aperture Radar (SAR)

Detecting changes in the Earth's surface over time

- The advantages of SAR for change detection
    - see through clouds
    - high temporal resolution
- Techniques for change detection with SAR?
    - ratio
    - log ratios between two images
    - t-tests
- fused with other data?

  Yes, with optical data using techniques such as

    1. principal component analysis
    2. object-based image analysis
    3. intensity fusion
- Applications?
    - monitoring land use changes
    - detecting deforestation
    - identifying urban growth pattern

9.1.2 SAR fundamentals

- Definition: Synthetic Aperture Radar (SAR) is a type of radar that uses microwave signals to create high-resolution images of the Earth's surface.
- Advantages:
  - Operates in all weather conditions
  - Penetrates through clouds and vegetation cover.
- Limitations
  - Sensitive to surface roughness; limited spatial resolution.
- Processing Techniques
  - Interferometry: combines multiple SAR images to create 3D maps of the Earth's surface.
  - Polarimetry: analyzes the polarization properties of reflected signals to extract additional information about surface features.

  The relationship between different surfaces and their sensitivity to polarizations in SAR data

Surface Type	Scattering Mechanism	Most Sensitive Polarization
Rough (bare earth)	Rough Scattering	Vertical-Vertical (VV)
Leaves	Volume Scattering	Vertical-Horizontal (VH) or Horizontal-Vertical (HV)
Trees / Buildings	Double Bounce	Horizontal-Horizontal (HH)

- Applications:
  - Environmental monitoring, disaster response, urban planning, military surveillance, and more.

9.1.3 Practical change detection with SAR

Topic	Key Points
Advantages of SAR for Change Detection	Can see through clouds unlike optical sensors; high temporal resolution.

Topic	Key Points
Change Detection Techniques	Ratio or log ratios between two images; t-tests; standard deviation.
Fusion of SAR and Optical Data	Principal component analysis; object-based image analysis; intensity fusion.
Applications of Change Detection with SAR	Monitoring land use changes, detecting deforestation, identifying urban growth patterns, and more.

9.1.4 Possible future developments

Resolution, accuracy, real-time-ness and data scale in SAR might see advancements.

- Improved resolution and accuracy:
  - Urban planning: Improved resolution and accuracy can influence local zoning regulations and urban growth management by providing detailed information on land use changes and the built environment. For example, high-resolution SAR data can be used to assess the effectiveness of urban containment policies or to identify areas where infrastructure investments are needed.
- Data processing capabilities
  - Disaster response: The ability to process larger datasets and monitor Earth's surface in near real-time can inform global policies regarding disaster management, such as the Sendai Framework for Disaster Risk Reduction. Boosted rapidness and data capability of SAR can better response to natural disasters, like earthquakes or hurricanes. This allows for quicker design of resources deployment strategy, and more comprehensive information in managing affected areas.
- New SAR applications
  - Agricultural: Enhanced SAR facilitates its use in change detection and monitoring. This will support policies like the European Union's Common Agricultural Policy (CAP), by providing data on crop health, irrigation needs, and land use changes.
  - Forestry (deforestation and reforestation tracking)
  - Disaster response (flood and landslide monitoring)
  - Environmental management: SAR data can inform policies related to wetland and coastal zone management, such as the Ramsar Convention on Wetlands and the United Nations Convention on the Law of the Sea (UNCLOS). By monitoring changes in these sensitive areas, policymakers can evaluate the effectiveness of existing regulations and develop new strategies to protect vital ecosystems.
- More matured machine learning and artificial intelligence:

- Advanced algorithms that are yet to be developed or need further maturity for SAR data analysis could include:

Algorithm	Pros	Cons
Improved Unsupervised Change Detection Algorithms	- No need for labeled training data. - Can discover unknown or unexpected changes.	- May struggle with complex or subtle changes. - Can be sensitive to noise and variations in the data.
Multi-Modal Fusion Algorithms	- Combines SAR data with other sources (e.g., optical, hyperspectral) for better feature identification. - Can exploit the complementary strengths of different data types.	- Requires data synchronization and registration, which can be challenging. - May involve increased complexity and computational cost.
Graph-based Change Detection Algorithms	- Can model complex spatial relationships between features. - Robust against noise and speckle effects in SAR data.	- Computationally expensive, especially for large-scale datasets. - May require tuning of hyperparameters.

- Improved overall accuracy of change detection can support climate change adaptation efforts at both local and global levels, including the United Nations Framework Convention on Climate Change (UNFCCC) and the Paris Agreement. Predictive models based on SAR data can help policymakers identify areas at risk of flooding, coastal erosion, or other climate-related impacts, enabling the targeted adaptation strategies.

Advancements in SAR technology, combined with the integration of machine learning and AI, will enhance change detection capabilities, enabling new analysis avenues in sectors including agriculture, forestry, disaster management, and environmental protection, ultimately influencing policymaking and promoting more informed decision-making processes.

9.2 Application

This week I chose to explore the application of SAR in wetland classification and real-time monitoring, as well as its future advancement, in the context of a literature recommended by the module webpage (Dabboor and Brisco 2019).

Wetland acts as a kidney to Earth reminding one of the old positivist tradition in French philosophy, taking abstract structures as organisms. The chance to explore this wholist view with an analytic approach is thrilling, as these two paradigms that seem to be falling in a prevailing antithesis can actually sparkle inspiration and exhibit harmony of inclusion of each other.

9.2.1 Wetland classification

Wetland classification has been a daunting task utilising traditional air photo and filed visits. Ever since the launch of ERTS in 1972, this task has been expecting the evolution of methods through applying Earth Observation data.

The SOTA of this tasks, as implied in the literature, has been an Object-based classification:

- Combining multi-source data: optical and SAR
- Analysing using a machine-learning classification model
- Incorporating a Digital Elevation Model (DEM)
- Aims at identifying terrain suitability for wetlands and surface water

See @ for a workflow

This method achieves over 90% of accuracy and is useful in that its core method utilises SAR data's feature of "seeing under the water" to better identify flooded vegetation class.

How did this amazing feature come about? Is that an emerging effect due to unprecedented combination of data? Or is it determined by the characteristics innate to SAR data?

The answer is the latter: The flooded vegetation tends to produce a double bounce scattering mechanism, which increases the intensity of the backscatter, making HH polarization to be the best for this due to the enhanced penetration in vegetation (**Jahncke2018MappingWI?**).

This workflow utilising different scattering effect between water and vegetation is one of the main contributors to the SOTA performance, laying the ground for further development and iteration on this method: various angle-choosing strategies, machine-learning algorithm combinations, etc.

9.2.2 Future Advancement

- Resolution in time and space can see chance of enhancement. Spaceborne SAR remote sensing technology being the essential tool for effective wetland observation, its improvement can be expected to reflect on enhancement of wetland observation in temporal and spatial resolution, e.g., the RCM is expected to provide SAR imagery in a spatial resolution ranging from 1 m to 100 m, in a revisit time of only 4 days (Dabboor and Brisco 2019).

This can better our understanding of climate change in wetlands and water quality, allowing ecosystem managers and decision makers to have sufficient information regarding wetland preservation

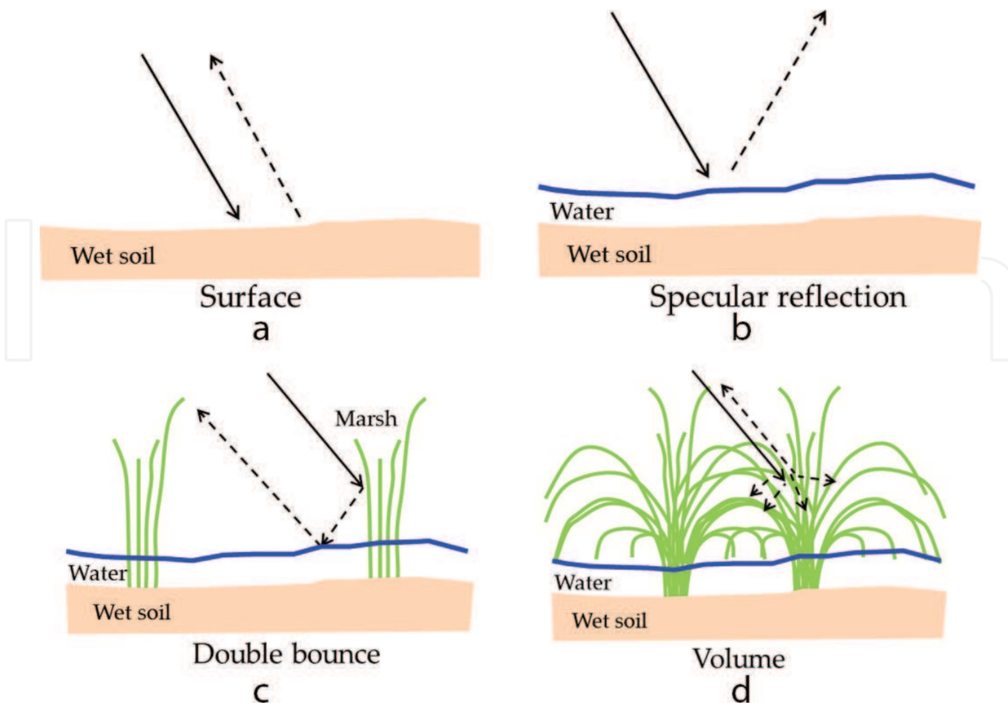


Figure 7. (a) Surface scattering mechanism from wet soil, (b) radar signal reflection from shallow open water, (c) double bounce scattering mechanism from signal interaction with vegetation stems and water surface and (d) volume scattering due to random scattering within the dense flooded vegetation canopy.

Figure 9.1: Credit: Dabboor and Brisco (2019)

- More sensors with more data forms, as well as improved data quality can be anticipated in the future. The integration of SAR imagery with optical and topographic data from multiple sensors was shown in Dubeau et al. (2017) to be necessary for improved wetland mapping and classification during the growing season.

However, the integration of SAR imagery and LiDAR data did not improve significantly the classification accuracy of wetland in Millard and Richardson (2013).

- The effectiveness of machine learning algorithms for automated wetland classification can expect further development. E.g., Graph-based Change Detection Algorithms can model complex spatial relationships between features and is robust against noise and speckle effects in SAR data.

This shift toward the automated machine learning algorithms comes to fulfill the requirement for operational wetland monitoring systems.

9.3 Reflection

SAR data with its amazing all-weather capability and real-time monitoring feature heralds the possibilities of real-time monitoring, which aligns with the disaster monitoring tasks. But there remain concerns in this approach:

- How real-time is real time?
  - Does a 5-seconds delay in data transmission fail an essential decision-making?
- How to quantify tolerance thresholds in regards of delay, error and bias?
  - A sharp cutoff or a fuzzy one?
- What degree of accuracy? Managing response time of
  - Data acquirement, transmission and calculation
  - Human-based decision-making

  Incurs extra cost. The estimation is almost always NOT deterministic.

Thinking about these questions can prepare one for awareness of risks when applying seemingly fancy technologies to real-world problems where people lives are concerned and efficiency of the solution needs elaborate calculation.

10 Summary

This module significantly boosted my knowledge and practice experience in fitting Remote Sensing into data pipeline and framing the workflow as a process of informing policies. I had my undergraduate degree including GIS training and GIS software development, while I see this module as a chance for systematically delving into Earth Observation data, as well as simplifying the process on platforms like GEE. As a result, my understanding of EO data incorporation process and technical fundamentals are deepened by tense practice in case studies and wide literature exposure.

Especially, the workflow of Google Earth Engine gives me a template of handling big-data and designing online EO data workflow, including the fast deployment of machine learning models in EO image classification tasks introduced in Week 6 and Week 7.

Also, the exemplar ways of approaching policy gaps between plans and execution by introducing EO data and remote sensing will guide me to utilise technical expertise to participate in policy framing and improvement. The group presentation also gave me a chance to do so in designing a flood risk management solution in Kuala Lumpur. The outcome can be viewed here, thanks to all team members' collaborated efforts.

Thanks to Andy, who delivered this amazing module!

References

Butcher, Ginger. 2016. *Tour of the Electromagnetic Spectrum*. Third edition. Washington, DC: National Aeronautics; Space Administration.

Dabboor, Mohammed, and Brian Brisco. 2019. "Wetland Monitoring and Mapping Using Synthetic Aperture Radar." In *Wetlands Management - Assessing Risk and Sustainable Solutions*, edited by Didem Gökçe. IntechOpen. https://doi.org/10.5772/intechopen.80224.

Dubeau, Pierre, Douglas J. King, Dikaso Gojamme Unbushe, and Lisa-Maria Rebelo. 2017. "Mapping the Dabus Wetlands, Ethiopia, Using Random Forest Classification of Landsat, PALSAR and Topographic Data." *Remote Sensing* 9 (10). https://doi.org/10.3390/rs9101056.

Google. 2023a. "Machine Learning in Earth Engine | Google Earth Engine." *Google Developers*. https://developers.google.com/earth-engine/guides/machine-learning.

———. 2023b. "Reducer Overview | Google Earth Engine." *Google Developers*. https://developers.google.com/earth-engine/guides/reducers_intro.

Gorelick, Noel, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. 2017. "Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone." *Remote Sensing of Environment*, Big Remotely Sensed Data: Tools, applications and experiences, 202 (December): 18–27. https://doi.org/10.1016/j.rse.2017.06.031.

Jensen, J. Robert. 1986. "Introductory Digital Image Processing: A Remote Sensing Perspective." In.

MacLachlan, Andrew, Eloise Biggs, Gareth Roberts, and Bryan Boruff. 2021. "Sustainable City Planning: A Data-Driven Approach for Mitigating Urban Heat." *Frontiers in Built Environment* 6. https://www.frontiersin.org/articles/10.3389/fbuil.2020.519599.

Millard, Koreen, and Murray Richardson. 2013. "Wetland Mapping with LiDAR Derivatives, SAR Polarimetric Decompositions, and LiDAR–SAR Fusion Using a Random Forest Classifier." *Canadian Journal of Remote Sensing* 39 (4): 290–307. https://doi.org/10.5589/m13-038.

Moore, R., and M. Hansen. 2011. "Google Earth Engine: A New Cloud-Computing Platform for Global-Scale Earth Observation Data and Analysis." *AGU Fall Meeting Abstracts*, December, 02.

Nair, Rahul, P. J. Paul, and K. Poulose Jacob. 2018. "Automated Image Annotation: A Survey." *Journal of Imaging* 4 (3): 37.

Robison, Keela. 2018. "The Future of Image Annotation: Human in the Loop." https://lionbridge.ai/articles/the-future-of-image-annotation-human-in-the-loop/.

Saad El Imanni, Hajar, Abderrazak El Harti, El Mostafa Bachaoui, Hicham Mouncif, Fatine Eddassouqui, Mohamed Achraf Hasnai, and Moulay Ismail Zinelabidine. 2023. "Multispectral UAV Data for Detection of Weeds in a Citrus Farm Using Machine Learning and Google Earth Engine: Case Study of Morocco." *Remote Sensing Applications: Society and Environment* 30 (April): 100941. https://doi.org/10.1016/j.rsase.2023.100941.

Schulte to Bühne, Henrike, and Nathalie Pettorelli. 2018. "Better Together: Integrating and Fusing Multispectral and Radar Satellite Imagery to Inform Biodiversity Monitoring, Ecological Research and Conservation Science." *Methods in Ecology and Evolution* 9 (4): 849–65. https://doi.org/10.1111/2041-210X.12942.