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

Third International Workshop on Capturing Scientific Knowledge
Sciknow 2019

Juan Manuel Carrillo Garcia, **Daniel Garijo**, Mark Crowley, Rober Carrillo,
Yolanda Gil and Katherine Borda

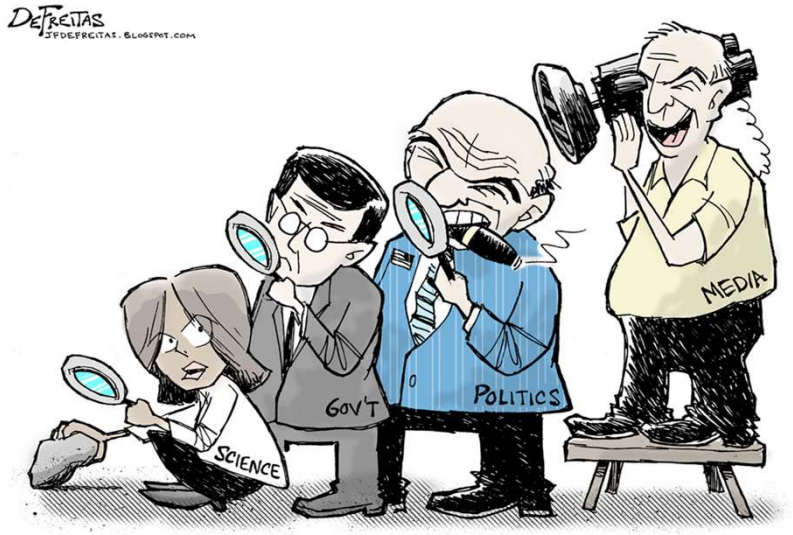
Introduction

Research in Climate Science

- Highly complex, nonlinear dynamics
- Disparate spatial and temporal scales
- Multidisciplinary teams

Therefore, it can benefit from best practices in

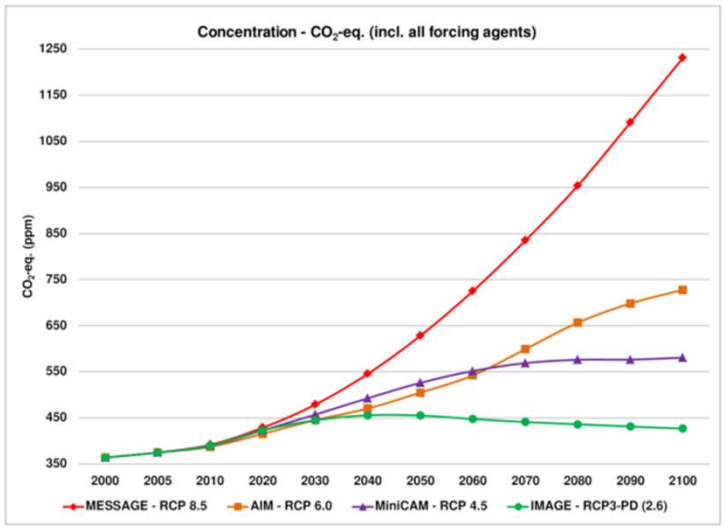
- Data Provenance
- Software interoperability and reusability
- Reproducibility of experiments



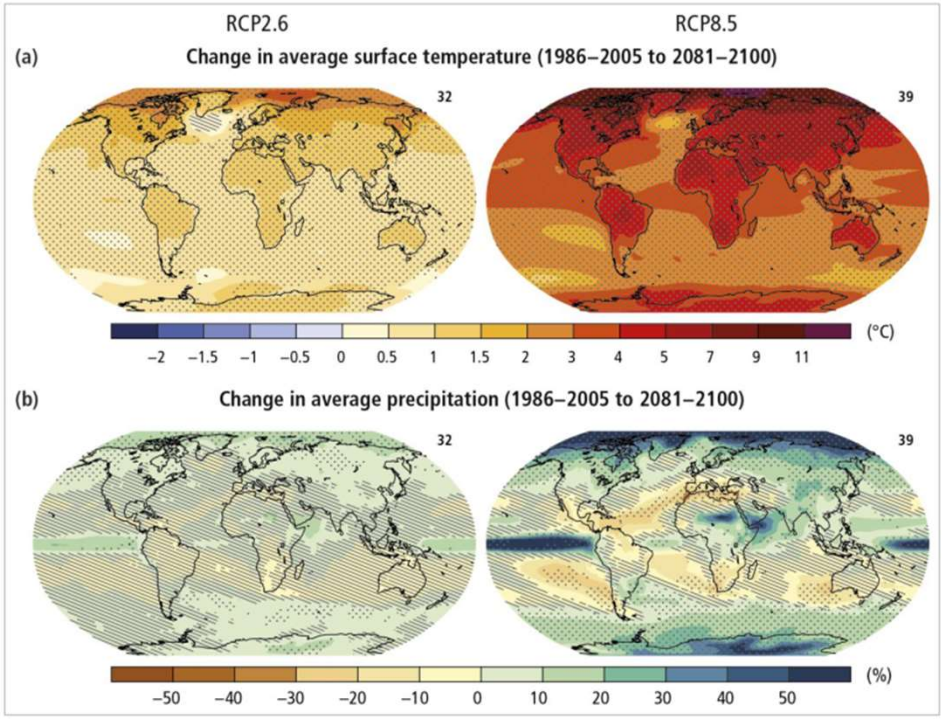
The ultimate goal is to improve Knowledge sharing, and Semantic workflows contribute significantly

Introduction

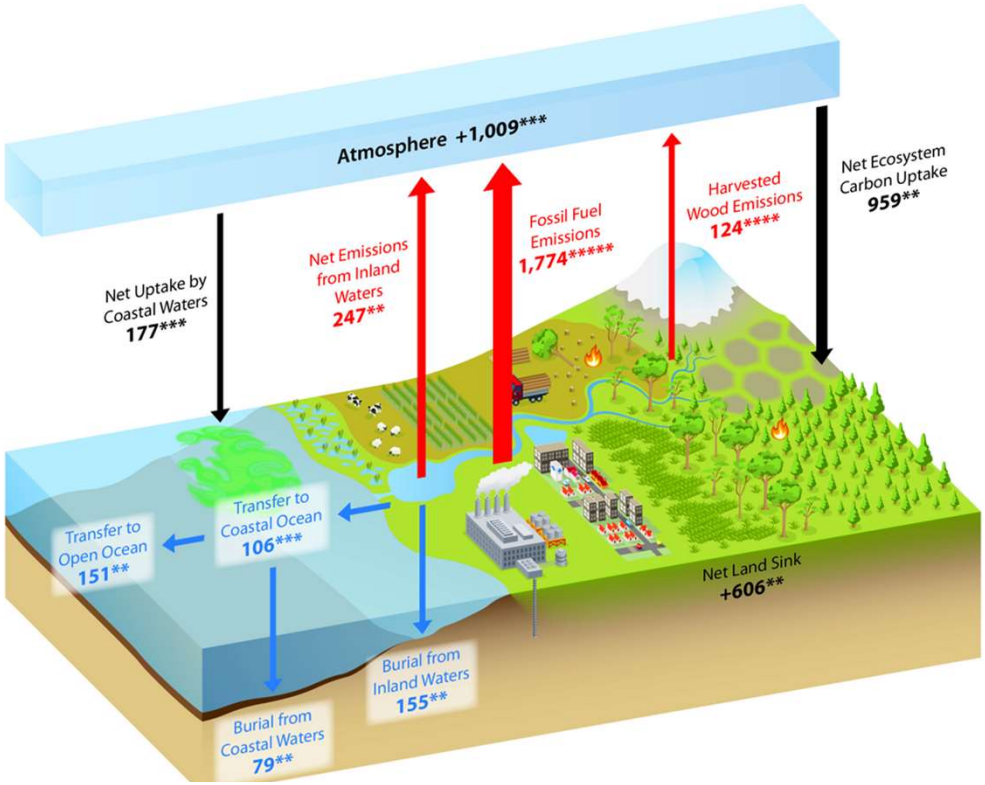
The study of Carbon emissions and storage is key in Climate Science.



Representative Concentration Pathways (RCPs) = Scenarios of Carbon emissions



Introduction

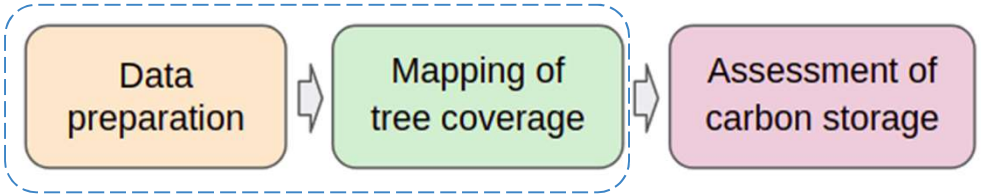


Major Carbon Fluxes of North America. Units are in teragrams of carbon (Tg C) per Year. Source: carbon2018.globalchange.gov

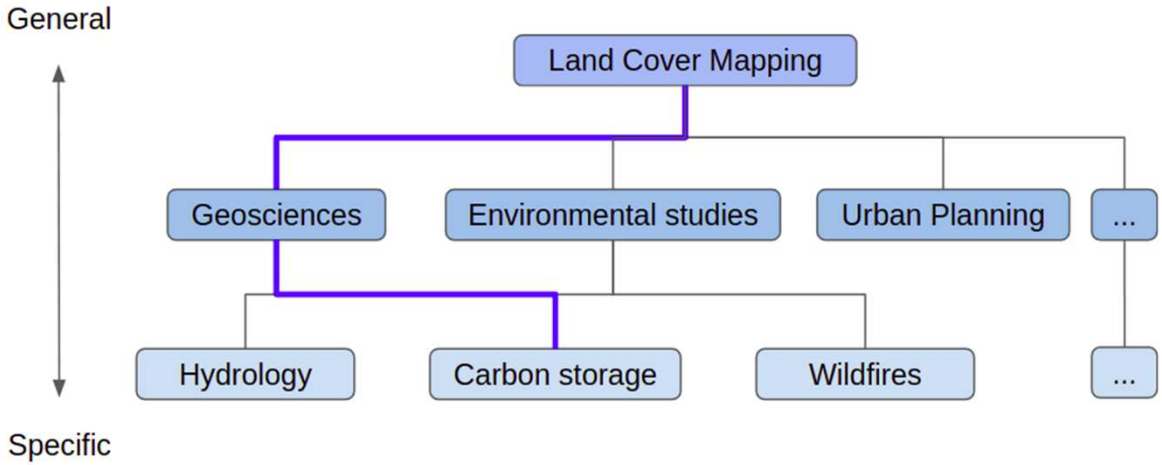
- Carbon storage by **urban trees**, in the form of biomass, is fundamental to mitigate emissions.
- The IPCC provides guidelines for the assessment of carbon stored in trees.
- However, each country determines the implementation details.

Scientific workflow design

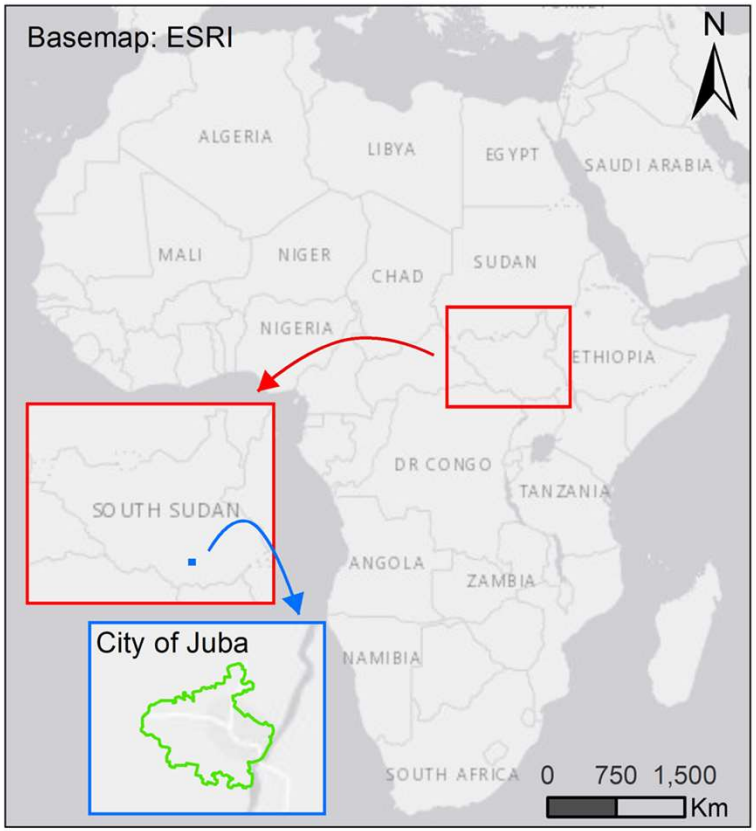
The workflow is designed as multiple interconnected components in WINGS that operate in three consecutive stages.



The components in stages one and two are also useful for other applications.



Area of study and data



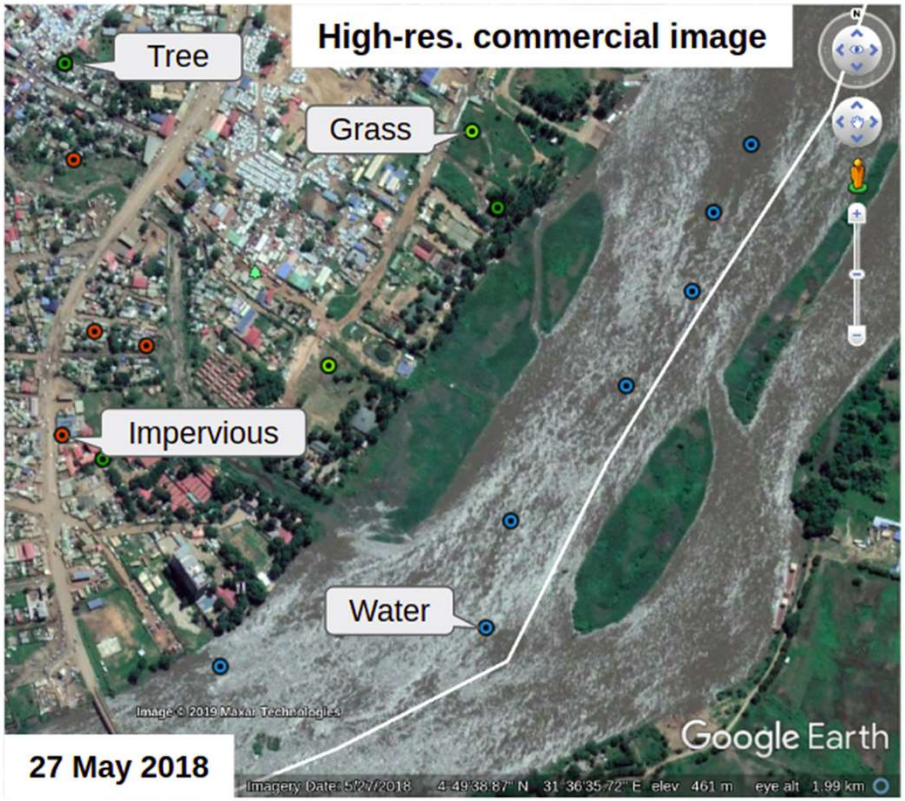
City of Juba in South Sudan.

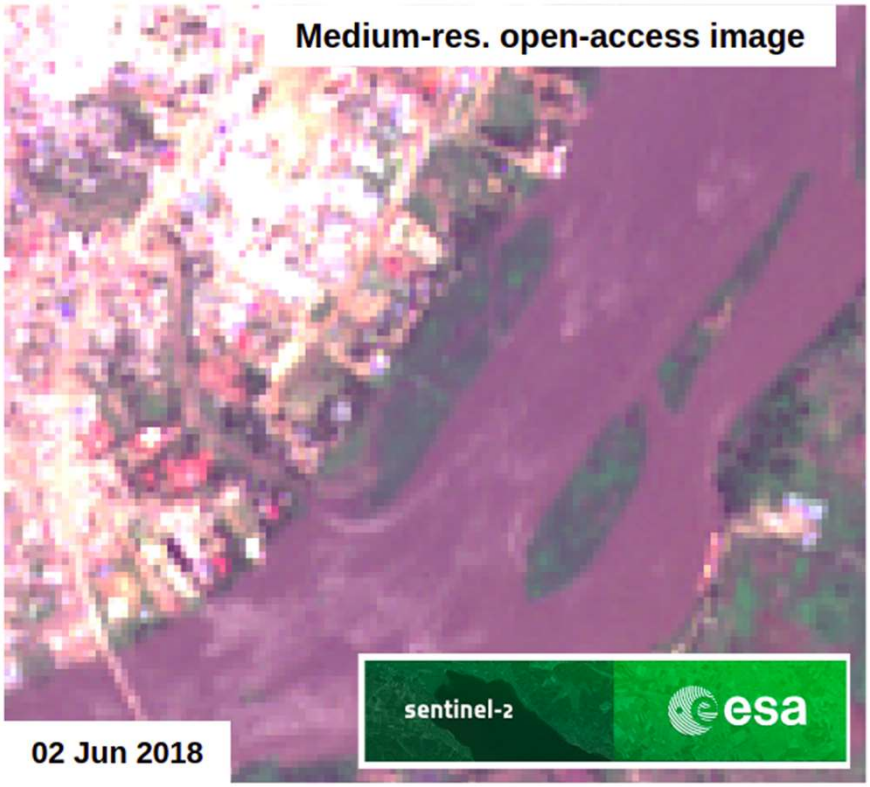
- Capital and main hub for commerce and transportation.
- Population is nearly 386,000

Some current issues

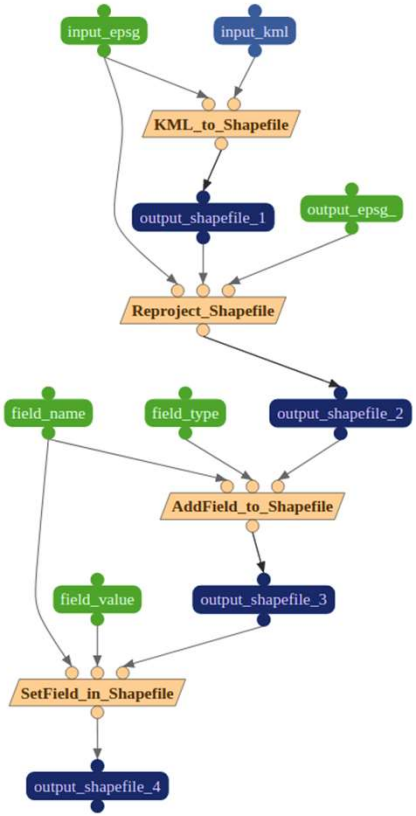
- Political instability
- Poor health services
- Lack of infrastructure

Area of study and data





Implementation and results

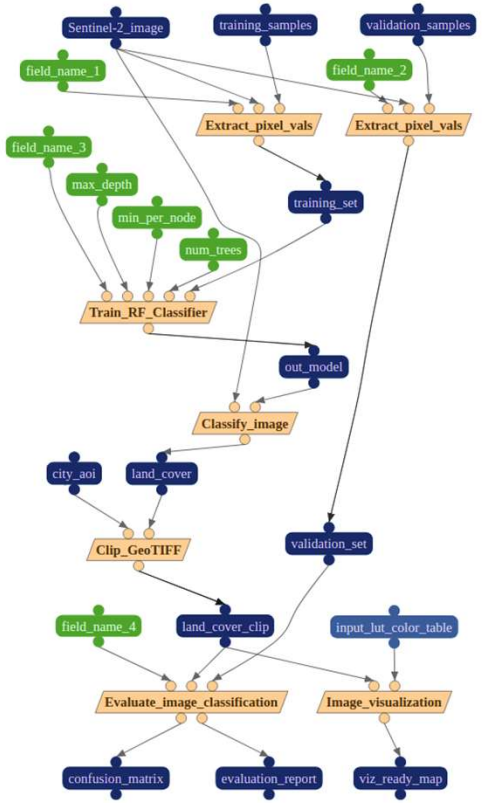


Workflow fragment for data preparation

Conversion of file formats

Transform coordinate system

Prepare file for assignment of labels



Workflow fragment for mapping tree cover

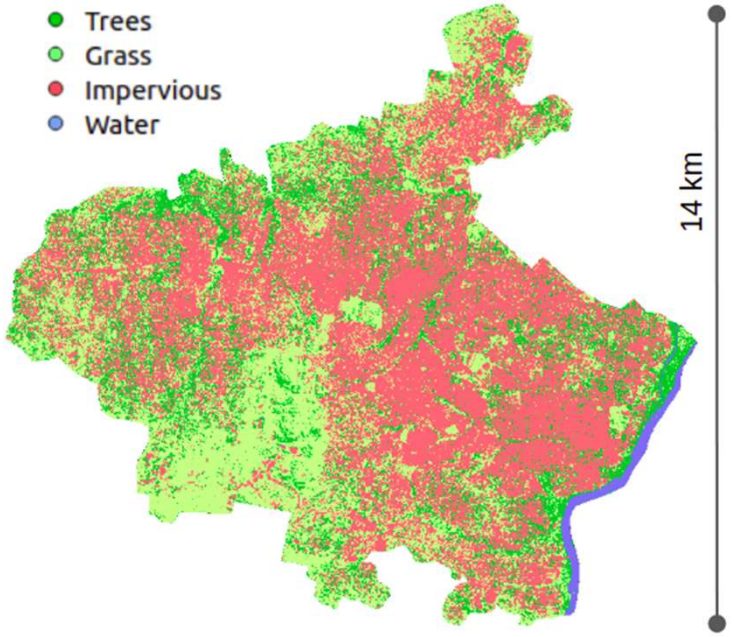
Split sample points in 80% for training and 20% for validation.

Crop satellite image to area of interest

Training of Random Forest image classifier

Implementation and results

Resulting land cover map for the city of Juba



Normalized confusion matrix - Classification with Random Forest

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

IPCC carbon removal factor of 2.9 tonnes of Carbon per hectare of crown cover per year.

Trees in the city of Juba remove **30,506 tonnes** of Carbon per year, roughly the emissions from **6632 buses**.

Conclusions and future work

- We introduce a library of workflow components to perform spatial data transformations, land cover mapping and assessment of carbon storage.
- We use scientific workflows to increase reusability of software components, reproducibility, and transparency of carbon assessment studies.
- Future work will focus on implementation for other locations around the globe and calibration of parameters to improve accuracy.

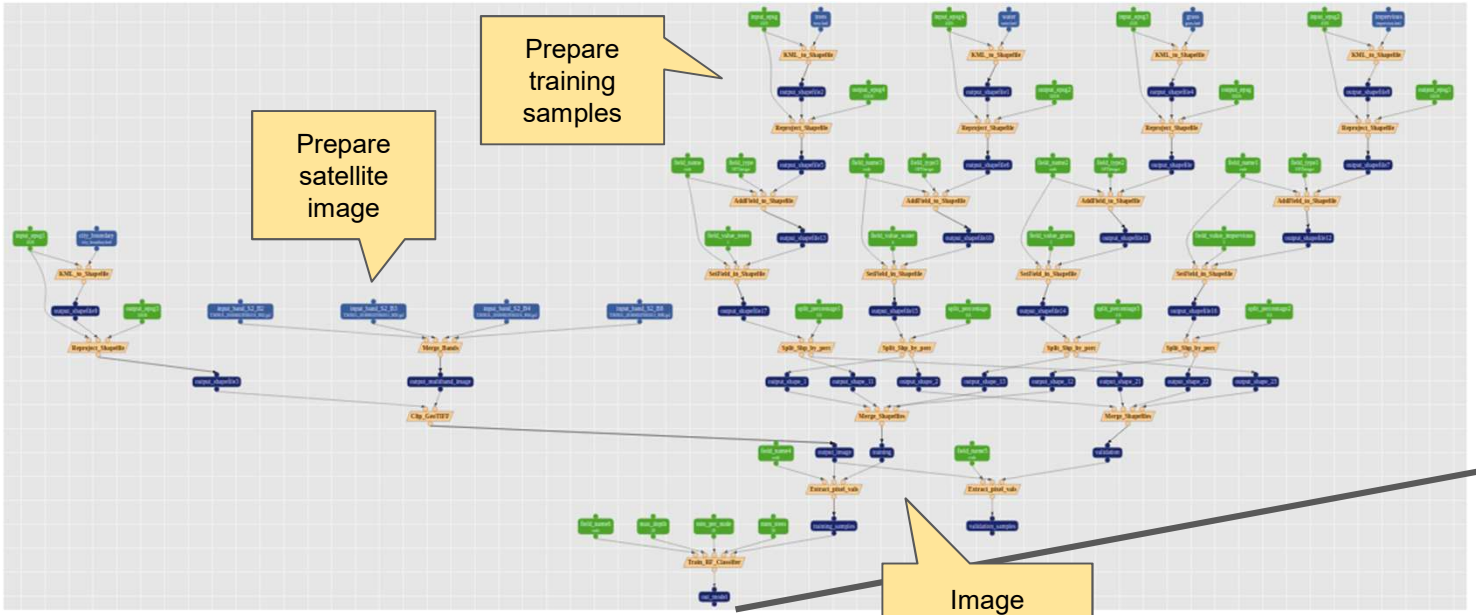
Complementary slides...

Outline

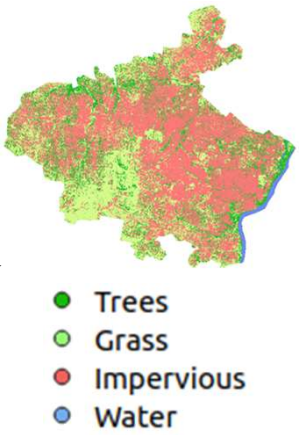
- Introduction
- Scientific workflow design
- Area of study and data
- Implementation and results
- Conclusions and future work

Land cover for the assessment of Carbon storage

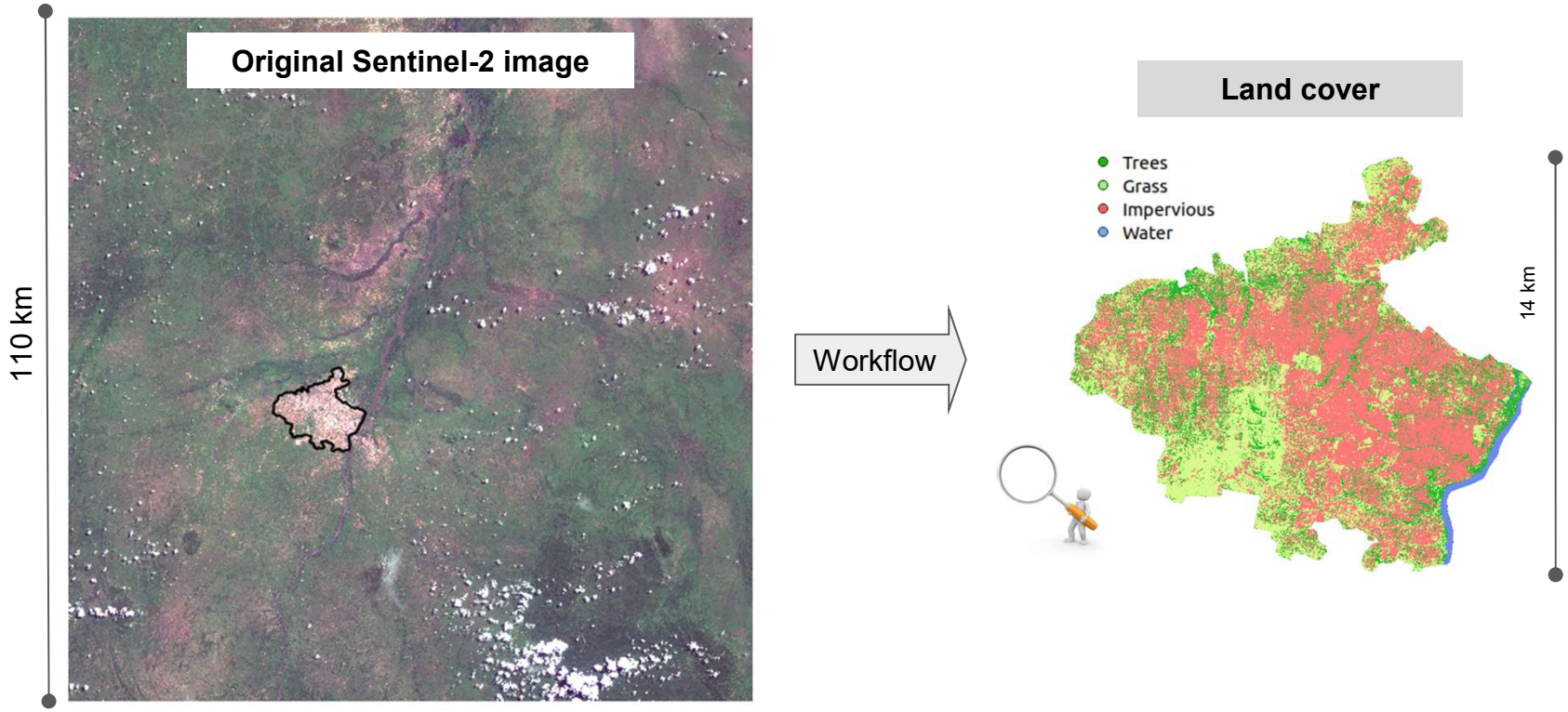
WINGS Workflow



Land cover



Land cover for the assessment of Carbon storage



Example data preparation

Geospatial ETL (Extract Transform Load)

Preparing a Digital Elevation Model DEM

- Convert file formats
- Reproject to a local coordinate system
- Combine multiple DEMs
- Filter polygon of interest from provinces
- Cut to an area of interest

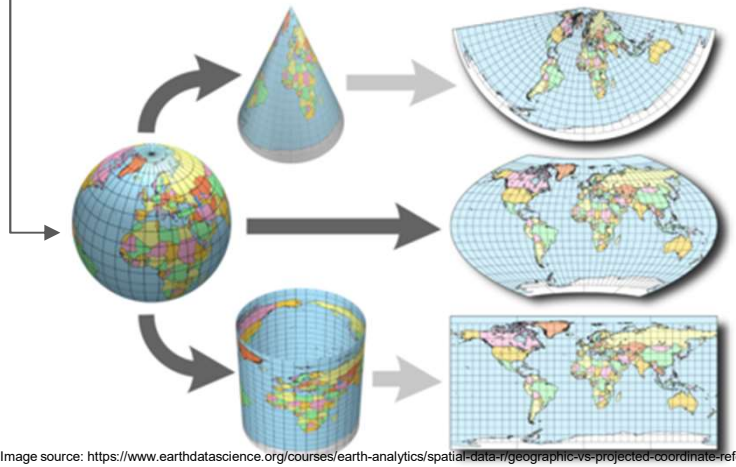
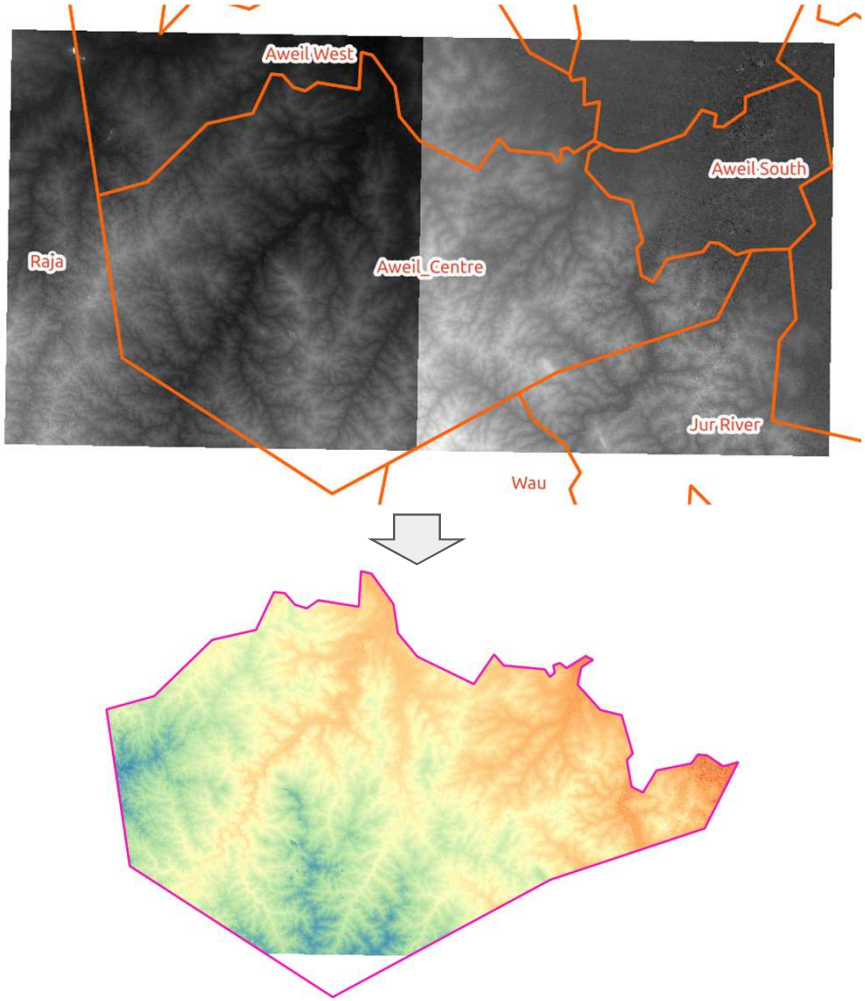


Image source: https://www.earthdatascience.org/courses/earth-analytics/spatial-data-r/geographic-vs-projected-coordinate-reference-systems-UTM/



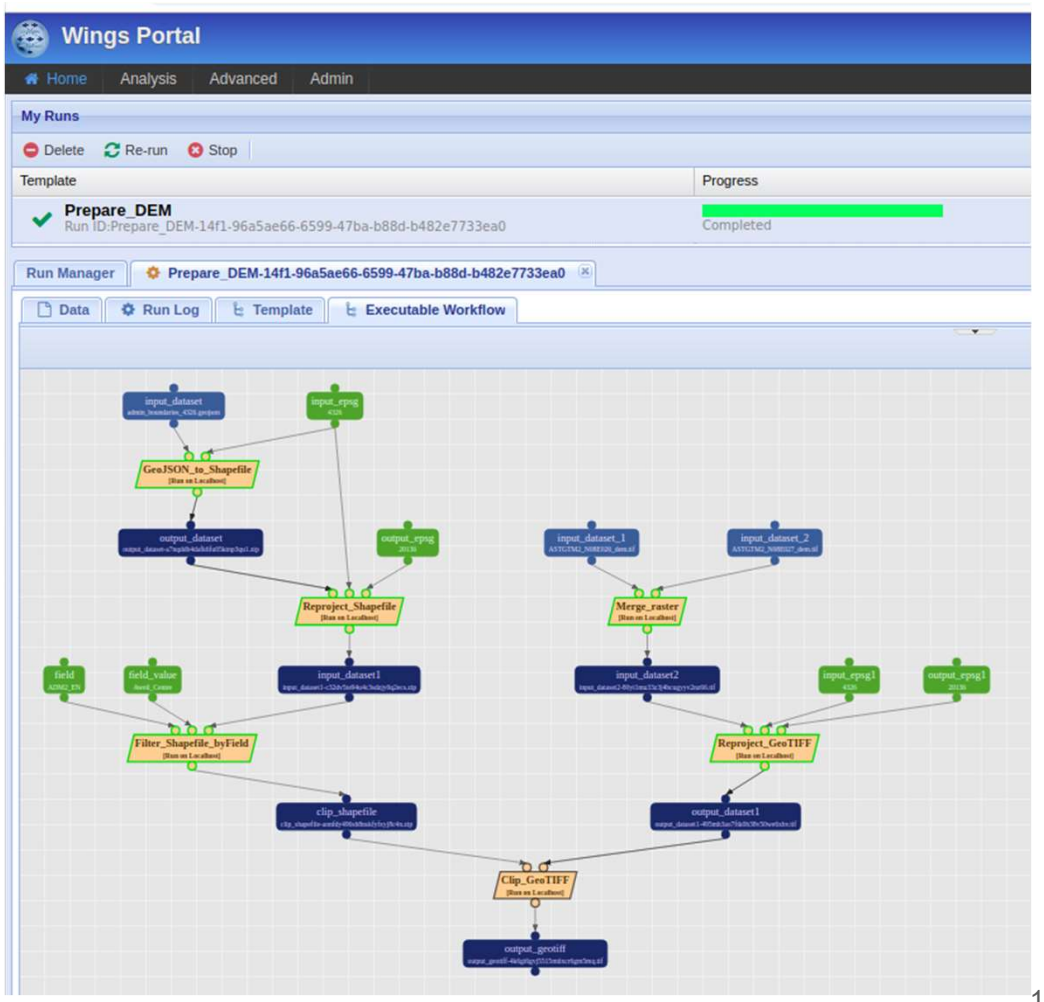
Preparation of DEM

Geospatial ETL (Extract Transform Load)

Implemented as components in WINGS that can be reused for other models

Using GDAL

Image source: https://gdal.org/index.html



Software

Data preparation

Geospatial Data Abstraction Library
gdal.org

Machine Learning

Orfeo ToolBox (OTB)
orfeo-toolbox.org