Mapping Global Land Cover with Sentinel-2 and AI: From Deep Learning to Data Exploration

8 May 2025

By: Tania Diamanta

Mapping Our Changing Planet: The Role of Land Use and Land Cover Intelligence

Harnessing Satellite Data and Deep Learning to Understand and Respond to Environmental Impacts




Human-driven land use and land cover (LULC) change has transformed ecosystems worldwide.

- Drivers: Agricultural expansion, urbanisation, mining, logging, livestock grazing, etc.
- These changes impact biodiversity, climate, and sustainable development
- Accurate, up-to-date LULC maps are essential for monitoring and policy-making
- However, global, high-resolution, time-series LULC data wasn't available (in 2020)

Sentinel-2 satellites + AI + cloud computing enabled scalable LULC mapping

- Sentinel-2 offers 10m resolution + high revisit rate → ideal for LULC monitoring
- Deep learning segmentation model trained on 5+ billion human-labeled pixels
- Cloud computing reduced training time from 1.2 million hours → 7 days
- Enables annual global LULC maps, accessible to analysts and decision-makers

Highlight

A powerful fusion of satellite imagery, deep learning, and cloud computing now makes global, high-resolution, annual LULC maps possible — empowering better decisions for people and the planet.

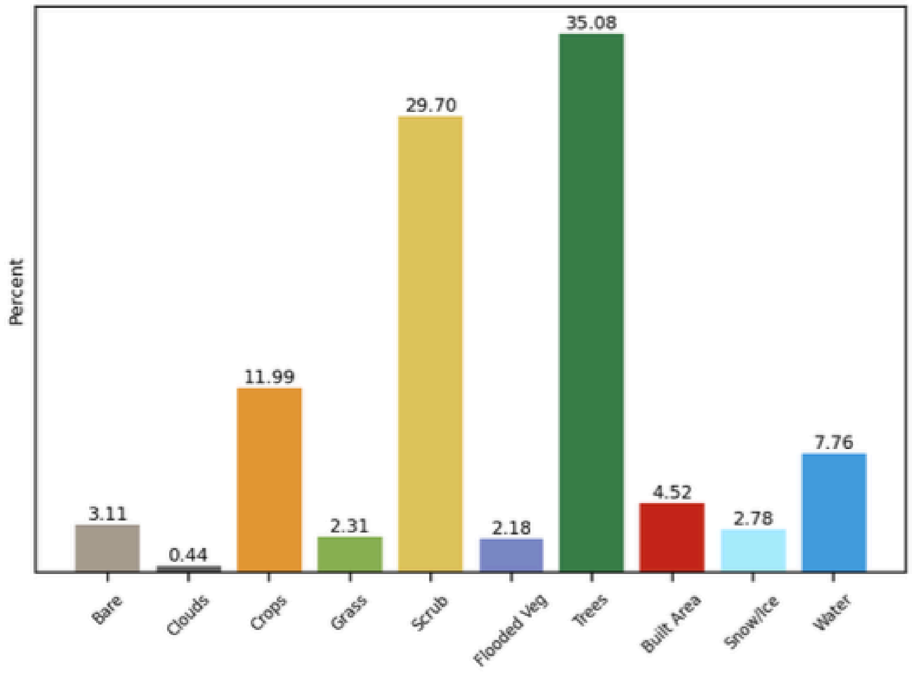
How Does the Model Work?

Building a Global Land Cover Map with Deep Learning and Sentinel-2 Imagery

Input

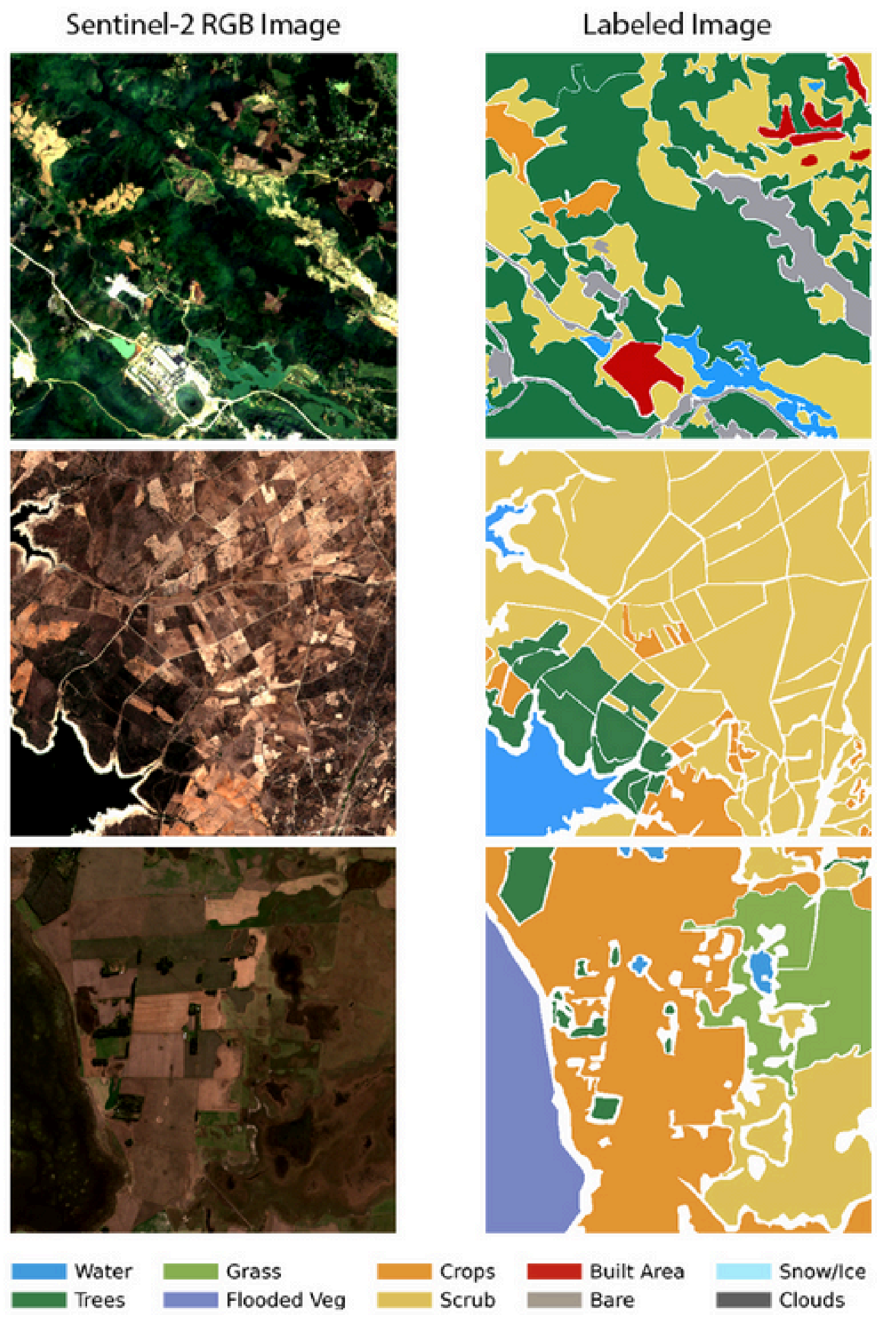
Human labelled areas (Train)

- There were >24,000 5km x 5km areas from satelite images, manually labelled into 10 categories: water, trees, grass, flooded vegetation, crops, scrub/shrub, built area, bare ground, snow/ice, and cloud. In total, that's over 5 billion pixels!
- These areas were randomly sampled from 14 distinct biomes across the globe



Satelite images (Test)

Obtained from Sentinel-2 Satelites



Process

Deep learning (UNet)

- Build a UNet deep learning model to label every pixel in a satellite image into one of 10 land cover categories
- Used 6 spectral bands (R, G, B, NIR, SWIR1, SWIR2)

Train the model in the Cloud

- It learns by comparing its predictions with the human labels
- To handle class imbalance: inverse log-weighting
- To avoid overfitting:
  - Data augmentation (flipping images to teach variety)
  - Dropout (randomly turning off 20% neurons during training)
- The model was trained for 100 epochs with learning rate

Output

Global LULC maps with pixel-level classification at 10m resolution

- Overall accuracy 85%
  - Most accurate: water, trees, crops, built-area
  - Challenges: grass, flooded vegetation, bare grounds, scrubs
- Accessible from Microsoft Planetary Computer → **next discussion!**

Accessing the Global LULC Dataset from Microsoft Planetary Computer

Exploring Land Cover Change Using Sentinel-2 Data and Python

1 Setting Up the Environment

```python
# system
import os
import warnings
warnings.filterwarnings(action='ignore')
```

```python
# data handling
import pystac_client
import odc.stac
from pystac.extensions.item_assets import ItemAssetsExtension

import geopandas as gpd
import rasterio as rio
import numpy as np
import pandas as pd
from shapely.geometry import Polygon
import xarray as xr
import rioxarray
```

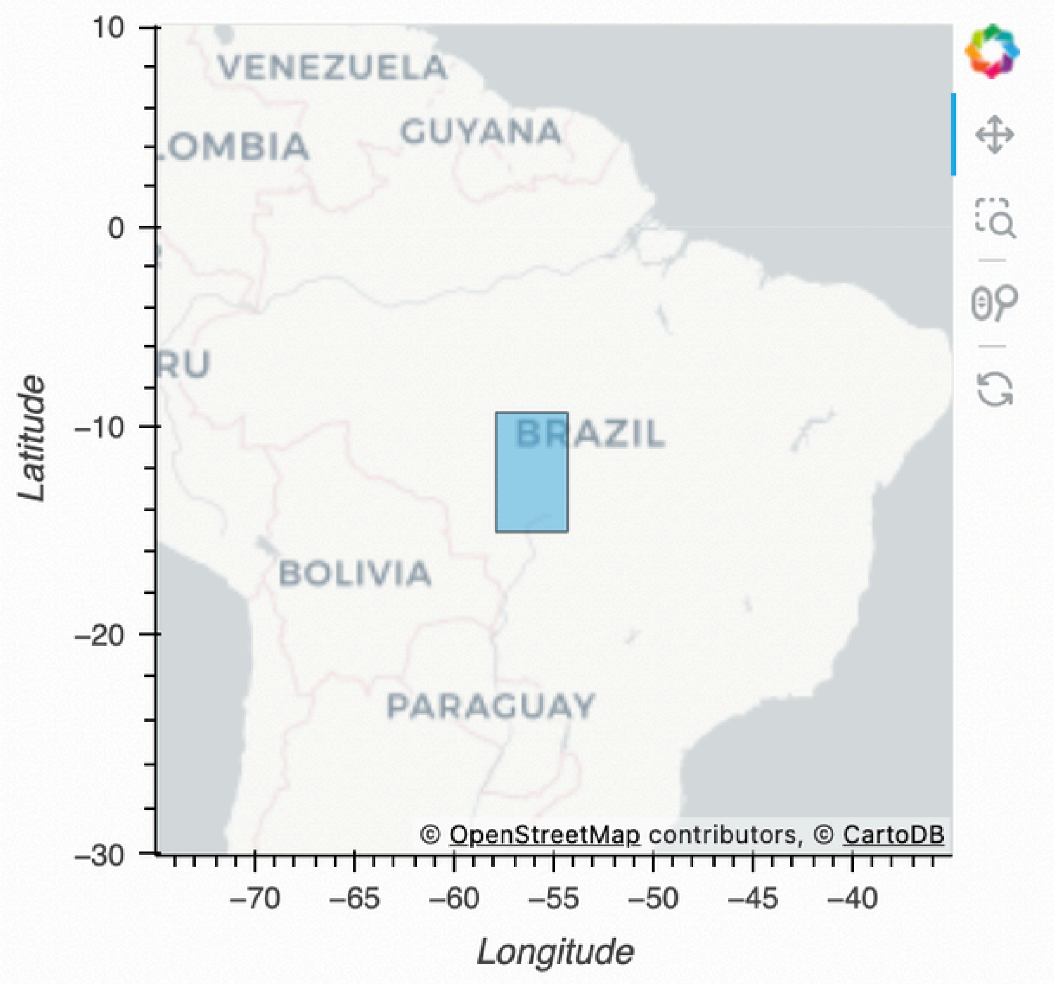
```python
# visualisation
import matplotlib.pyplot as plt
import matplotlib.colors as mplc
import holoviews as hv
import hvplot.pandas
from holoviews import opts, dim
```

```
odc-stac                  0.4.0                 pyhd8ed1ab_0    conda-forge
geopandas                 1.0.1                 pyhd8ed1ab_3    conda-forge
pystac                    1.13.0                pyhd8ed1ab_0    conda-forge
pystac-client             0.8.6                 pyhd8ed1ab_0    conda-forge
python                    3.10.17           h6cefb37_0_cpython    conda-forge
rasterio                  1.4.3             py310h524c300_1     conda-forge
```

2 Define the Study Area

For example: Mato Grosso, Brazil

- Define a rectangular bounding box for the study area
  - Use center point at longitude -56.1 and latitude -12.2.
  - A square area of approximately 200 km (east-west) × 325 km (north-south) is defined using geographic coordinates
- Convert bounding box into a GeoPandas GeoDataFrame.
- Plot it on a map using hvplot



3 Access LULC Data from MPC

- Access data for the study area using the Microsoft Planetary Computer's STAC API and psytac-client
- Search for land cover datasets (from the "io-lulc-9-class" collection) → select desired timeline (i.e. 2017-2020) → download
  - "io-lulc-9-class" = Output data obtained from training UNet with Sentinel-2 satelite images
- Load data to xarray using odc.stac.load()
  - Change resolution from 10m to ~30m, or 0.0027 degrees (faster to load)
  - Coordinate reference system EPSG:4326

```python
# Step 3: Query Microsoft Planetary Computer data using psytac-client

from pystac_client import Client
from planetary_computer import sign
import odc.stac

# Use MPC's STAC API
catalog = Client.open("https://planetarycomputer.microsoft.com/api/stac/v1")

# Define study area
x, y = (-56.1, -12.2)
km2deg = 1.0 / 111
xd = 200 * km2deg
yd = 325 * km2deg
bbox = [x - xd, y - yd, x + xd, y + yd]

# Search
query = catalog.search(
    collections=["io-lulc-9-class"],
    bbox=bbox,
    datetime="2017-01-01/2022-12-31",
    limit=100
)

items = list(query.get_items())
print(f"Found {len(items)} items")

# Sign STAC items with temporary tokens
signed_items = [sign(item) for item in items]
# Load data
lcxr = odc.stac.load(
    signed_items,
    bands=["data"],
    chunks={},
    fail_on_error=False,
    crs="EPSG:4326",
    resolution=0.0027
)
```

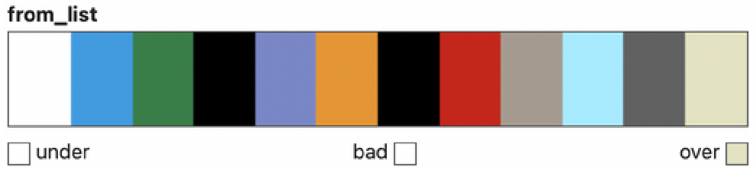
Accessing the Global LULC Dataset from Microsoft Planetary Computer

Exploring Land Cover Change Using Sentinel-2 Data and Python

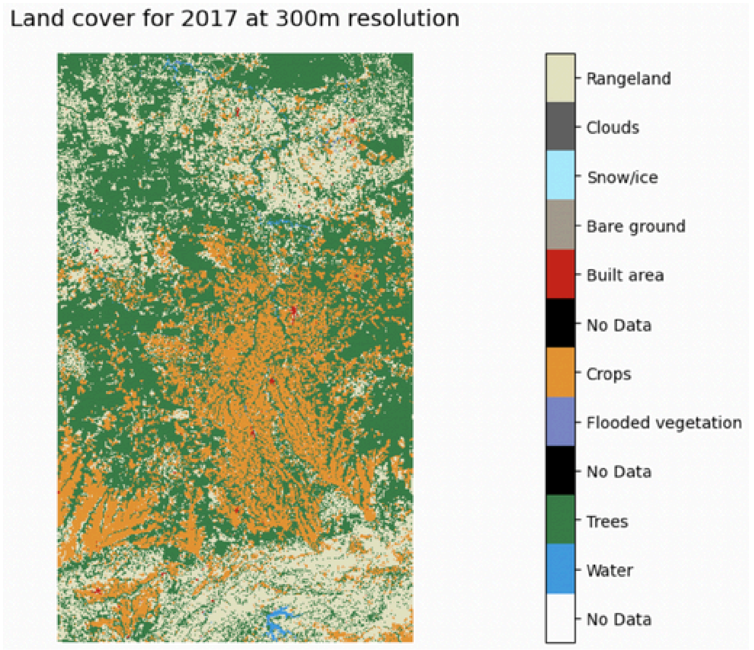
4 Visualize the Data

- Create a colormap. 1 color per land cover category

```
{'No Data': 0, 'Water': 1, 'Trees': 2, 'Flooded vegetation': 4, 'Crops': 5, 'Built area': 7,
 'Bare ground': 8, 'Snow/ice': 9, 'Clouds': 10, 'Rangeland': 11}
```



- Select one timestamp and visualise the map statically using matplotlib



- Visualise all years (optionally, can use holoviews for interactive map with a slider)

5 Analysis

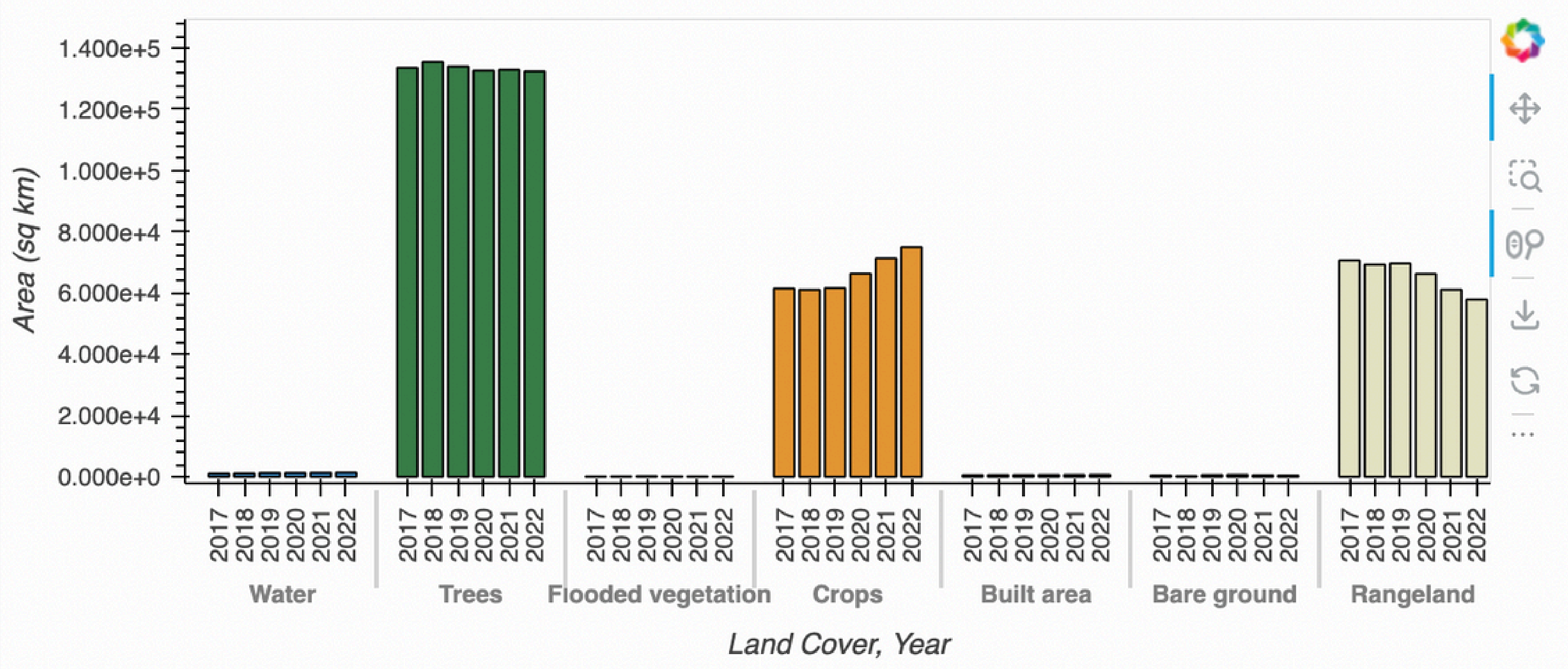
- Aggregate land cover change
- Pixel-by-pixel change
- Analysing zonal change

Next discussion!

Analyzing Land Cover Change Over Time

Aggregating Land Use Patterns with Municipal Boundaries

Type of Analysis	Key Findings

(1) Aggregate land cover change
• Counted total pixel area per land cover class across multiple years (2017-2022).
• Visualized changes using bar plots to show how land cover categories grew or shrank over time.
	• Trees consistently occupy the largest area, but their extent slightly decreases over the years.
	• Crops show a steady increase → agricultural expansion
	• Water, flooded vegetation, built areas, and bare ground while fluctuating, remains very low throughout the years.
	• Clouds and snow/ice do not appear → had 0 weightage
	• No grass and scrubs, 3 possibilities:
	  ○ Actual absence
	  ○ Resolution limitation → misclassfication (blended with other category)
	  ○ Combined into "rangeland"
	
	
	• Shows only overall trends, but doesn't tell what land types were changing into what.

Comparing Aggregate vs. Pixel-Level Land Cover Change Analysis

From Broad Trends to Detailed Transitions

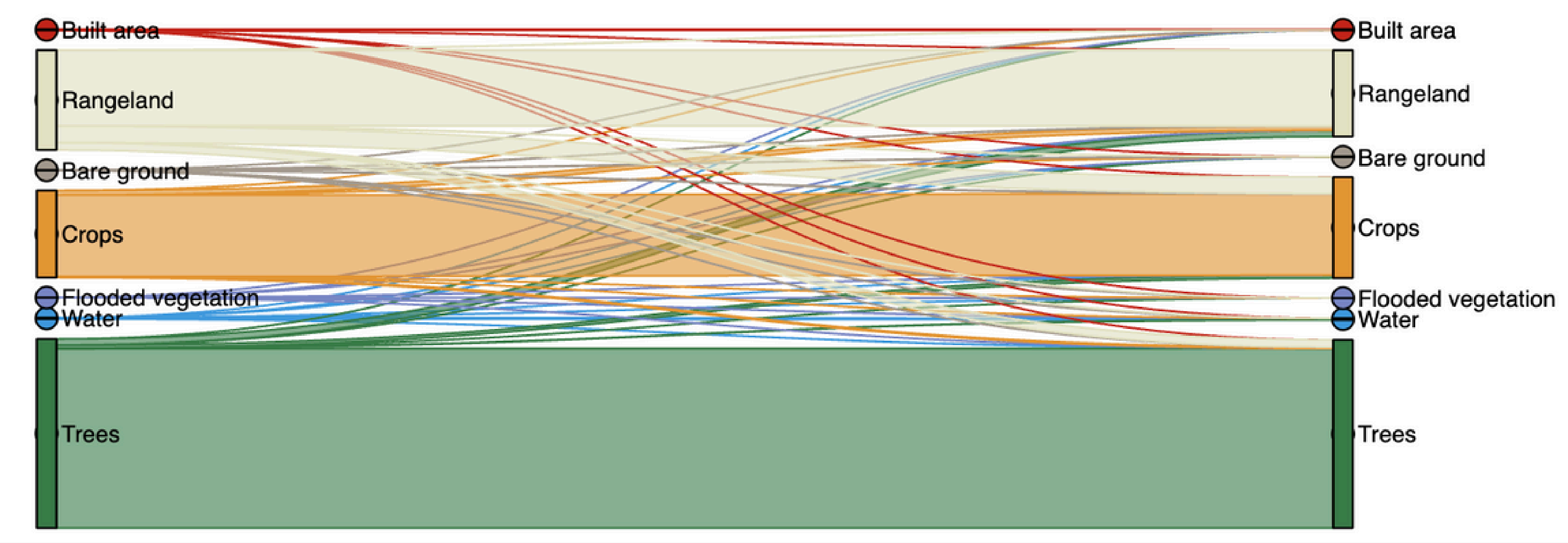
Type of Analysis	Key Findings

(2) Pixel-by-pixel change

- Created a confusion matrix to track how each pixel changed category.
- Used the matrix to generate a Sankey diagram — visualizing transitions between classes (identifying which categories were traded off)
- Provides a more nuanced picture of land dynamics (e.g. seasonal farming, deforestation, pasture expansion).

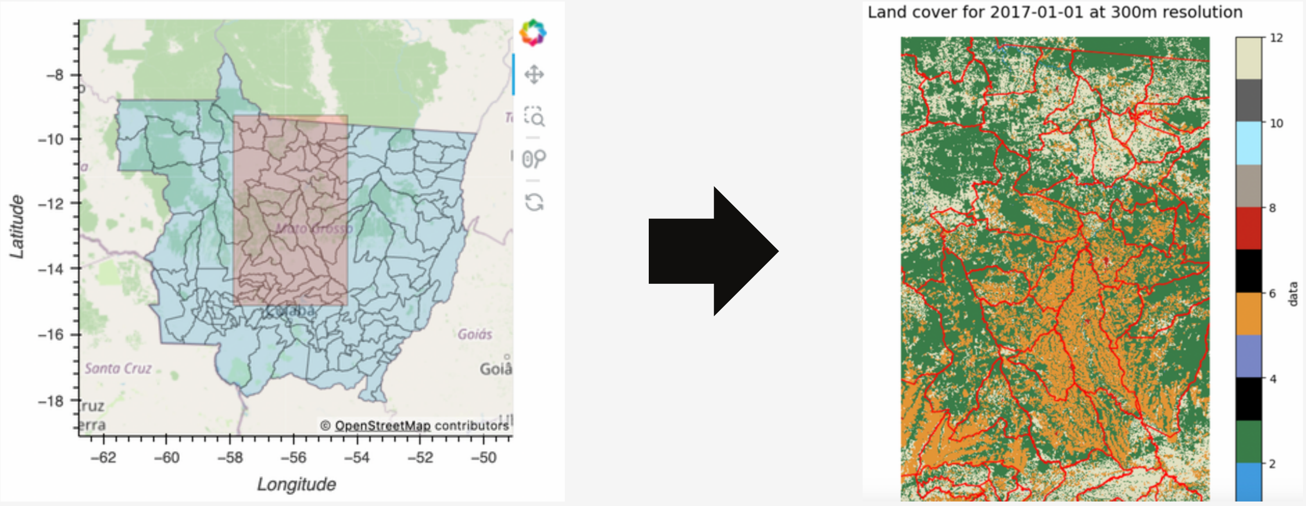
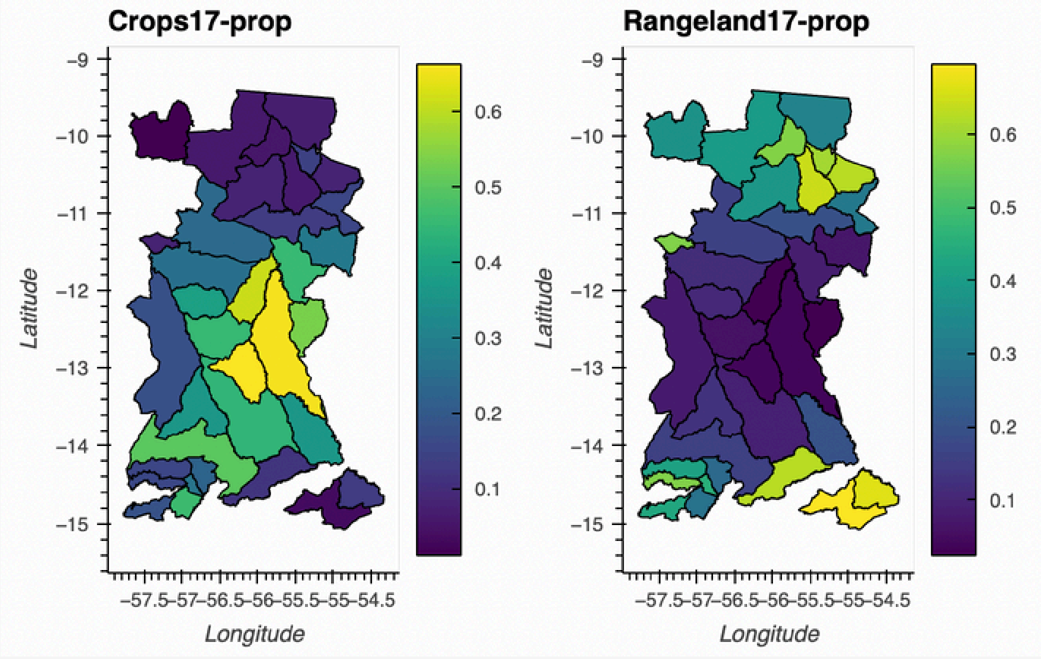
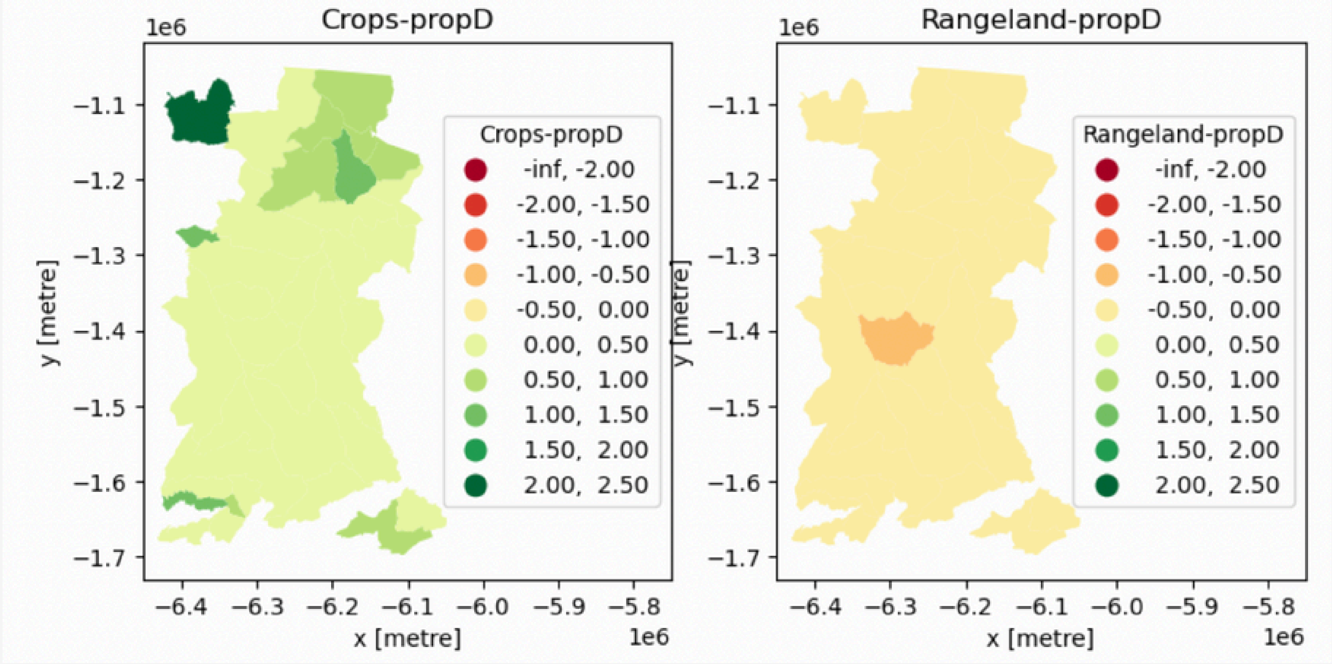
- The most frequent change was from Rangeland → Crops
- 2nd: Rangeland → Trees
- 3rd : Trees → Rangeland
- 4th: Crops → Rangeland
- 5th : Trees → Crops

Land Cover	Water	Trees	Flooded vegetation	Crops	Built area	Bare ground	Rangeland
Land Cover							
Water	10185	473	66	74	10	5	601
Trees	2577	1410068	245	23500	180	504	48050
Flooded vegetation	214	100	678	7	0	15	143
Crops	167	9397	3	638037	476	1746	33554
Built area	5	60	1	125	5227	5	160
Bare ground	15	179	42	2221	15	1251	1136
Rangeland	1110	57494	323	129038	638	1691	595239



Zonal Analysis of Land Cover Change

Understanding how land use trends vary across municipalities for targeted insights

Type of Analysis	Key Findings
(3) Analysing zonal change • Zooms in on administrative zones—specifically Brazilian municipalities—to understand how land cover is changing within each of them • Quantifies and visualises how much each municipality is changing	• From the plots below we should be able to start observing the land cover composition of each municipality, and how they might differ. For example, which are the municipalities with relatively high (or low) crop cover compared to rangeland cover? • Zonal analysis is ideal when you want to tie land use trends to political or socio-economic units like cities or regions—offering actionable insights for local planning.  • Then by calculating the proportion of each land cover class within each municipality, we can observe a pattern in a particular year (left) or the proportional change over time (right). For example: The cropland in Claudia went from 27.9% in 2017 to 33% in 2022, then the change is +18.3%.  

Strengths and Improvements: UNet Land Cover Model & Analysis Notebookxx

Evaluating model precision, data handling, and analytical depth for better land use insights

The UNet Model

Strength

- High spatial resolution (10m) → significant improvement over previous products like MODIS (500m), ESA CCI (300m), or Copernicus (100m).
- Well-designed training dataset (robust)
- Weights for class imbalance
- Use of Cloud

Improve-ment

- For land classes/categories that are more difficult to differentiate (i.e. grass vs scrubs or crops vs flooded veg.), incorporate NDVI or EVI curves to capture seasonal patterns in vegetation behavior
- Add Sentinel-1 radar or Landsat to improve classification in cloudy areas
- Add more hierarchy in the classes/categories (i.e. natural forest and plantation forest within 'trees' or industrial and residential for 'built-ups') for finer class distinction, allowing more actionable insights

The Notebook

Strength

- Includes both temporal analysis (2017-2022) and spatial (pixel-by-pixel, zonal statistics)
- Code is well documented (replicable)
- Effective visualisations (both static with matplotlib and interactive with hvplot)

Improve-ment

- Provide the environment used to run the script
- Automate repetitive code with function
- Include accuracy or uncertainty handling → the notebook use LULC data as input with 85% accuracy (disclose this). So instead of saying "Crops in 2017 = 11.2%", we say "Crops in 2017 = 11.2% ± 1.68%"
- Merge in socio-economic data (i.e. agricultural output, deforestation pressure, land policies) at the municipal level
- Increase resolution for better accuracy (current is 300m)

References

- *Abercrombie, S.P. and Friedl, M.A., 2015. Improving the consistency of multitemporal land cover maps using a hidden Markov model. IEEE Transactions on Geoscience and Remote Sensing, pp.703-713.*
- *Brown, C.F., Brumby, S.P., Guzder-Williams, B., Birch, T., Hyde, S.B., Mazzariello, J., Czerwinski, W., Pasquarella, V.J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R. and Tait, A.M., in review. Dynamic World: Near real-time global 10m land use land cover mapping. Nature Scientific Data.*
- *Brown, J., Tollerud, H., Barber, C., Zhou, Q., Dwyer, J.L., Vogelmann, J., Loveland, T., Woodcock, C., Stehman, S.V., Zhu, Z. and Pengra, B., 2019. Lessons learned implementing an operational continuous US national land change monitoring capability: The LCMAP approach. Remote Sensing of Environment.*
- *Buchhorn, M., Lesiv, M., Tsendbazar, N.E., Herold, M., Bertels, L. and Smets, B., 2020. Copernicus Global Land Cover Layers—Collection 2. Remote Sensing.*
- *Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N., Wikramanayake, E., Hahn, N., Palminteri, S., Hedao, P., Noss, R. et al., 2017. An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. BioScience.*
- *Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J.C., Mathis, M. and Brumby, S.P., 2021. Global land use / land cover with Sentinel 2 and deep learning. In: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, pp.4704-4707. DOI: 10.1109/IGARSS47720.2021.9553499.*
- *McGlinchy, J., Muller, B., Joseph, M. and Diaz, J., 2019. Application of UNet fully convolutional neural network to impervious surface segmentation in urban environment from high resolution satellite imagery. In: IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, IEEE, pp.3915-3918.*
- *Mousivand, A. and Arsanjani, J.J., 2019. Insights on the historical and emerging global land cover changes: The case of ESA-CCI-LC datasets. Applied Geography, pp.82-92.*
- *Olofsson, P., Foody, G., Stehman, S. and Woodcock, C., 2013. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. Remote Sensing of Environment, 129.*
- *Ronneberger, O., Fischer, P. and Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention, Springer, Cham, pp.234-241.*
- *Song, X.P., Hansen, M.C., Stehman, S.V., Potapov, P.V., Tyukavina, A., Vermote, E.F. and Townshend, J.R., 2018. Global land change from 1982 to 2016. Nature, pp.639-643.*
- *Wager, S., Wang, S. and Liang, P., 2013. Dropout Training as Adaptive Regularization. In: Advances in Neural Information Processing Systems (NIPS).*
- *Wickham, J., Stehman, S.V. and Homer, C.G., 2018. Spatial patterns of the United States National Land Cover Dataset (NLCD) land-cover change thematic accuracy (2001-2011). International Journal of Remote Sensing, pp.1729-1743.*
- *Zhang, P., Ke, Y., Zhang, Z., Wang, M., Li, P. and Zhang, S., 2018. Urban land use and land cover classification using novel deep learning models based on high spatial resolution satellite imagery. Sensors, 3717.*