Gaining Insights Using AI: where physics and biology meet data science

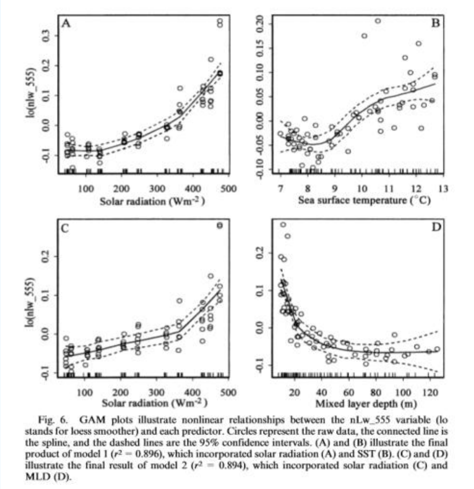
Dr Samantha Lavender

Pixalytics

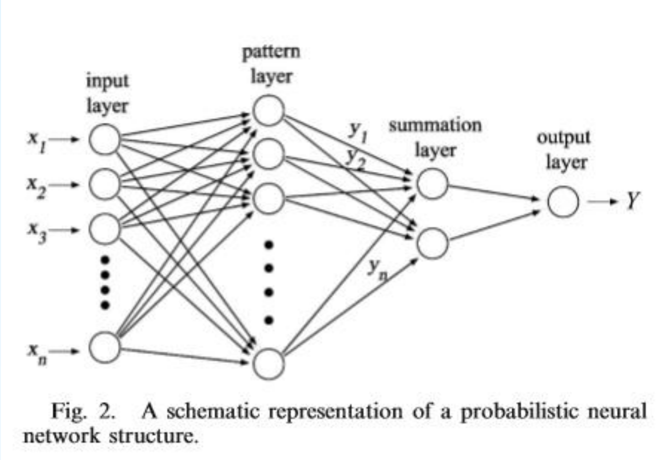
A seismic shift in compute, data and queries

- Large Language Models (LLMs) have democratised asking geospatial questions, built on cloud computing and trained using data from the internet/humans
- It is easier to run/train a Neural Network than to develop an algorithm/model, and you don't have to understand the biology/physics involved
- Questions we should be asking:
  - How do we believe the answers?
  - Where do we as experts best contribute?
  - Where does fundamental knowledge play a role?

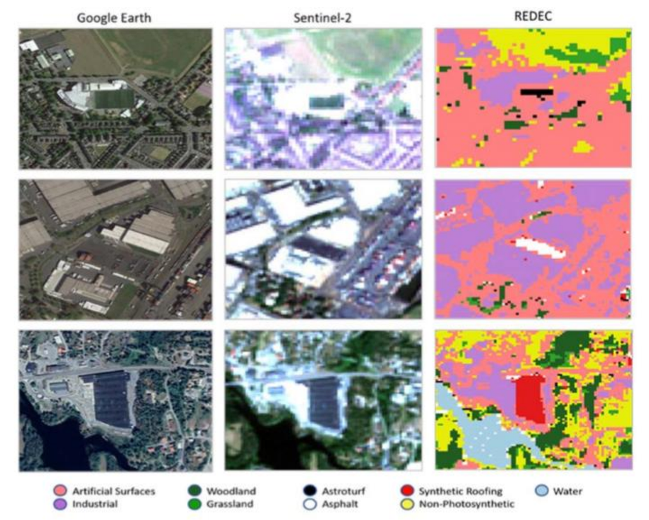
A personal journey



Generalized additive model (2006)

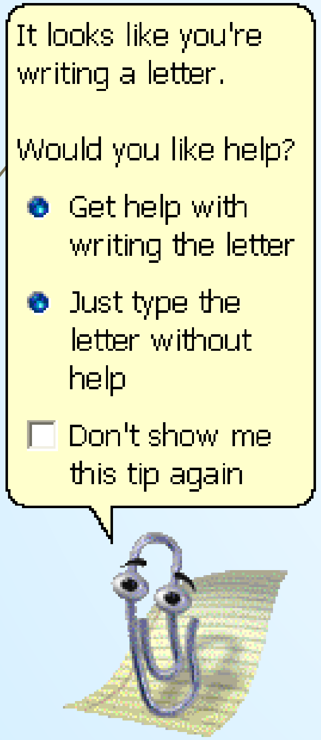


Neural Network (2008)



Random Forests (2019-), Neural Networks (2021-)

Expansion to broader techniques for data analysis, trying to resist Large Language Models (LLMs) for everyday tasks

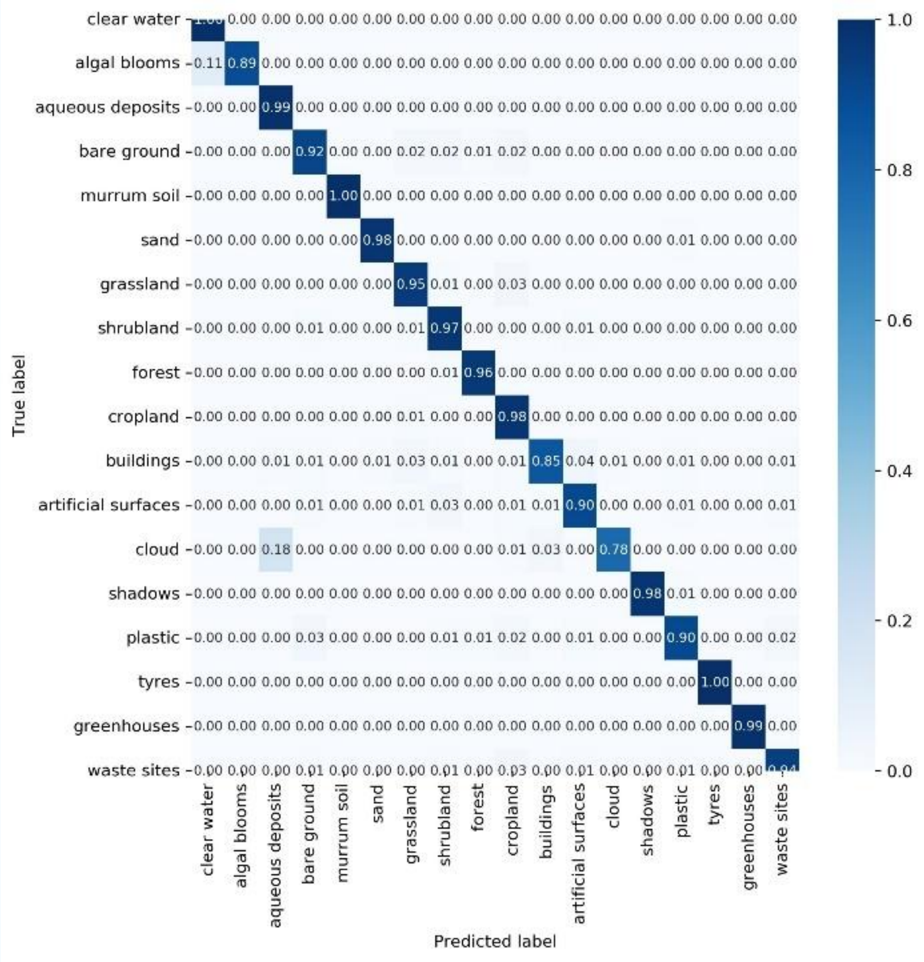


https://en.wikipedia.org/wiki/Office_Assistant

A lot of work on training data, supported by iterative training

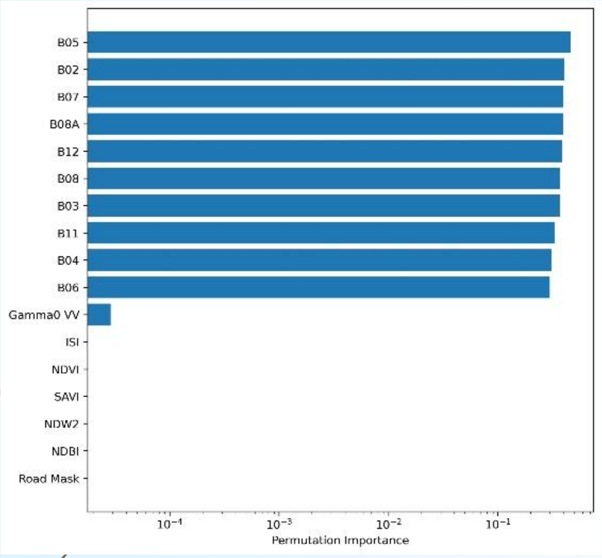
Table 2-1. Locations of marine plastic for training and verification

Country	Water Body	Plastic Visible in Google Earth?	Used for training	Notes
Kenya	On-land dump	Yes	Yes	Dandora dumping ground https://theintercept.com/2020/04/19/africa-plastic-waste-kenya-ethiopia
Indonesia (Bali)	Kuta Beach	Yes	Yes	After a storm in February 2019 https://www.youtube.com/watch?v=PRBfiQUyvwU
Indonesia (Java)	Muncar Port	Yes (street view)	Yes	https://www.huffingtonpost.co.uk/entry/indonesia-plastic-waste-pollution-solutions_n_5cabc096e4b02e7a705c317c?ri18n=true
Philippines	Manilla	Yes (street view)	Yes	https://www.farandwide.com/s/worst-dirtiest-beaches-world-c4c5c451e15e40a4
United Kingdom (Scotland)	Plastic waste storage sites on land	Yes	Yes	Training data used for Page *et al.* (2020), limited access due to sensitivity of illegal waste storage sites
Indonesia	Brantas (river)	No	Example processed, Figure 2-1	https://borgenproject.org/tag/plastic-pollution-in-indonesia/
Indonesia	Solo (river)	No		
Indonesia	Serayu (river)	No		
Indonesia	Progo (river)	No		
China (Hong Kong)	Clear Water Bay	Maybe	Not as this stage	https://www.scmp.com/sport/hong-kong/article/2097251/hong-kong-water-sportspersons-called-clean-plastic-pollution-world
India	Gulf of Khambhat	Yes	Not as this stage	Looks like there is some plastic by the shore and just off the coast
Philippines	Camotes Sea, Cebu City	Yes	Not as this stage	In some images there is plastic build up around the edges of the river. https://cebudailynews.inquirer.net/196623/fencing-the-waters
South Africa	Durban harbour	N/A	Not as this stage	Flooding over the Easter weekend. Photo example in Biermann *et al.* 2020

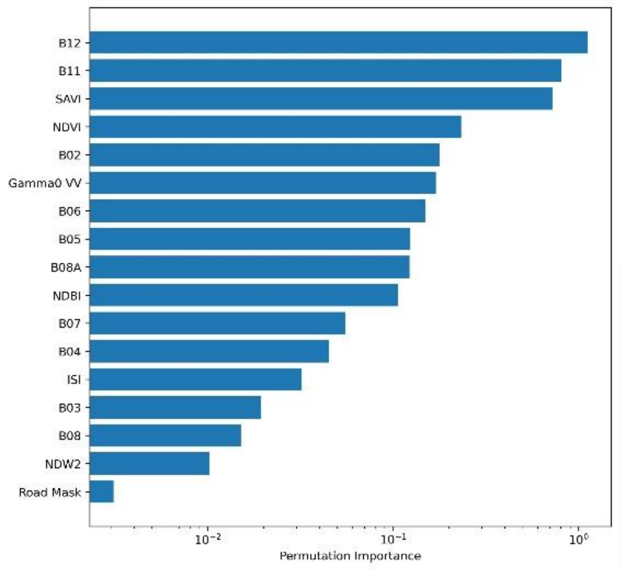


Model training uses KerasTuner, statistics were both overall accuracy and plastic accuracy

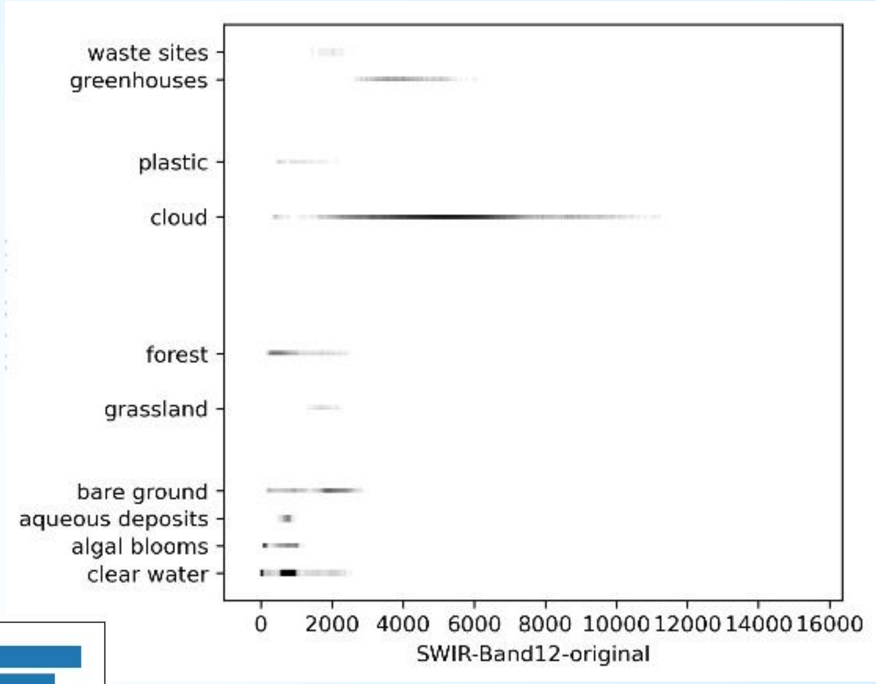
Testing different inputs layers

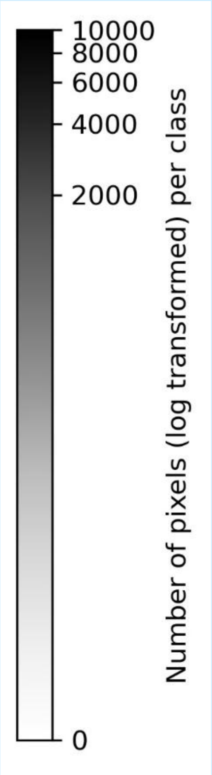


Linear Regression



Non-Linear: Random Forest





Distribution for the Training Data

Tracking down and removing anomalous data

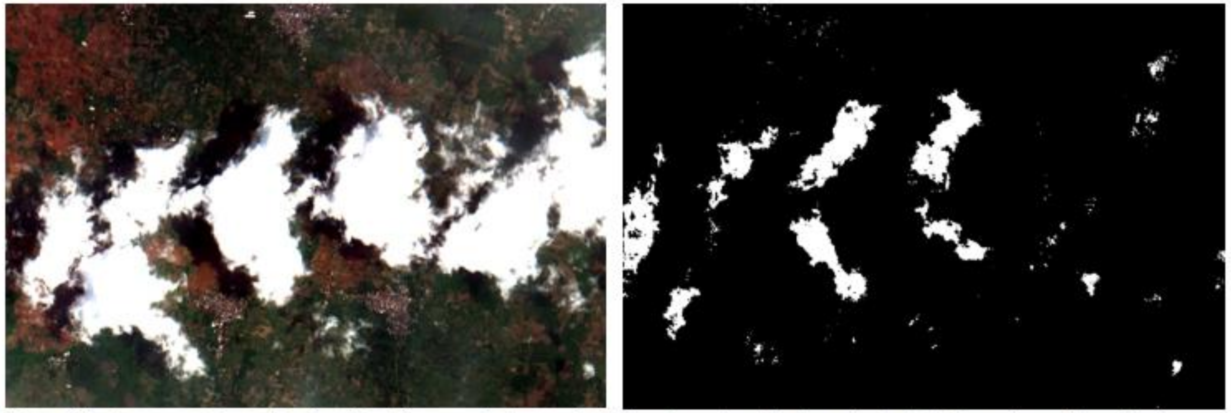


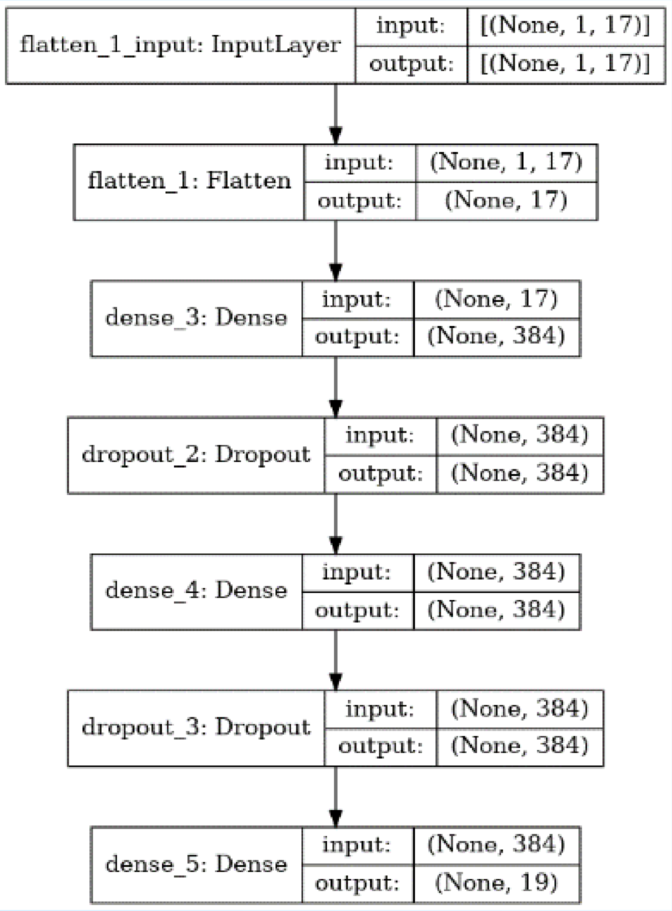
Figure 2-1: A comparison of a Sentinel-2 pseudo-true colour image subset to the pixels that have been identified using the new shadow index.

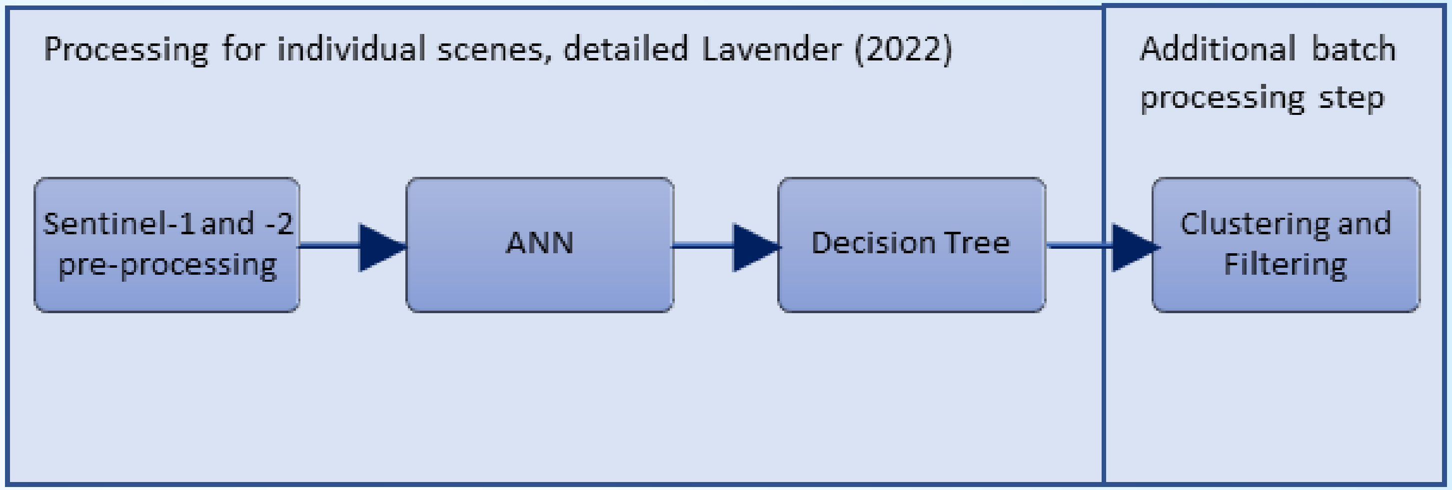
Improved Shadow Index

Fast moving objects can create rainbows: moving clouds as well as airplanes and road vehicles

Layer No.	Input
1	MSI Band 2 (490 nm)
2	MSI Band 3 (560 nm)
3	MSI Band 4 (665 nm)
4	MSI Band 5 (705 nm)
5	MSI Band 6 (740 nm)
6	MSI Band 7 (783 nm)
7	MSI Band 8 (842 nm)
8	MSI Band 8A (865 nm)
9	MSI Band 11 (1 610 nm)
10	MSI Band 12 (1 190 nm)
11	Normalised Difference Vegetation Index (NDVI)
12	Soil Adjusted Vegetation Index (SAVI)
13	Normalised Difference Water Index (NDWI2)
14	Road Mask based on the Sentinel-2 truck detection approach of H. Fisser (https://bit.ly/3ywcSG9)
15	Normalised Difference Build-Up Index (NDBI)
16	Improved Shadow Index (ISI)
17	Gamma0 VV (Sentinel-1)

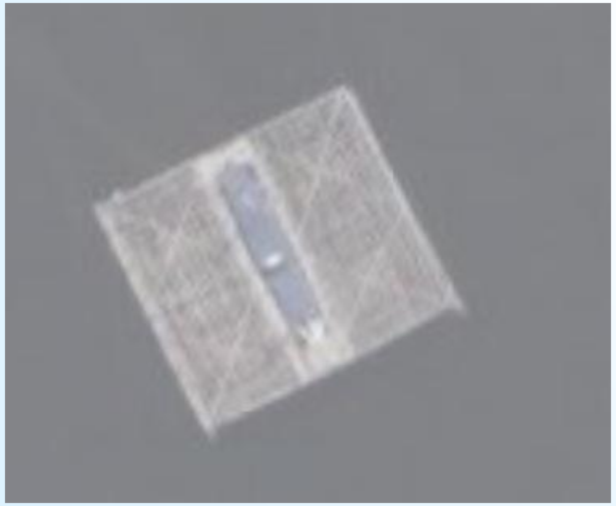
The model design

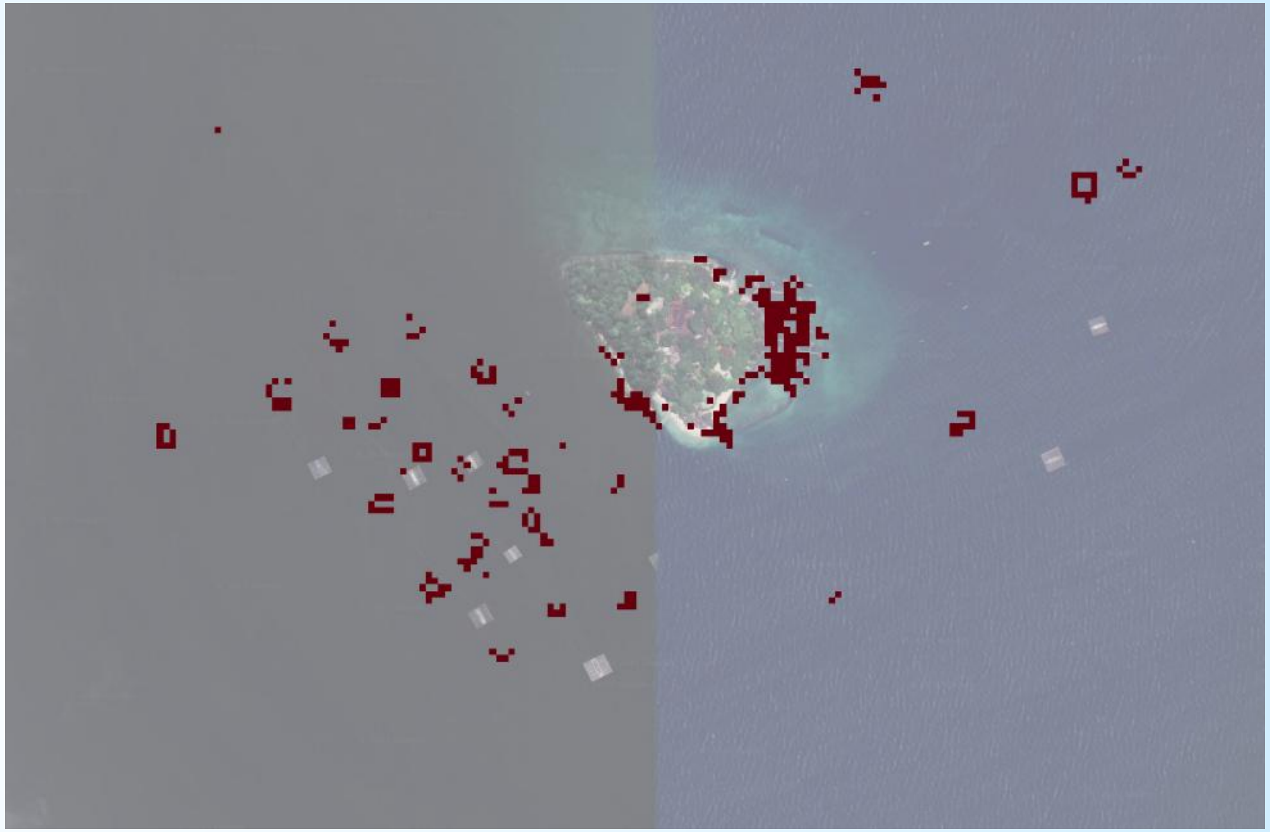




- **Flatten:** converting multi-dimensional input data into a one-dimensional vector, making it suitable for the subsequent dense layer
- **Dense:** fully connected layers where each neuron receives input from all neurons in the previous layer, supporting complex feature learning.
- **Dropout:** randomly shut a fraction of the layer's neurons, helps to prevent overtraining.

Detection of aquaculture activities

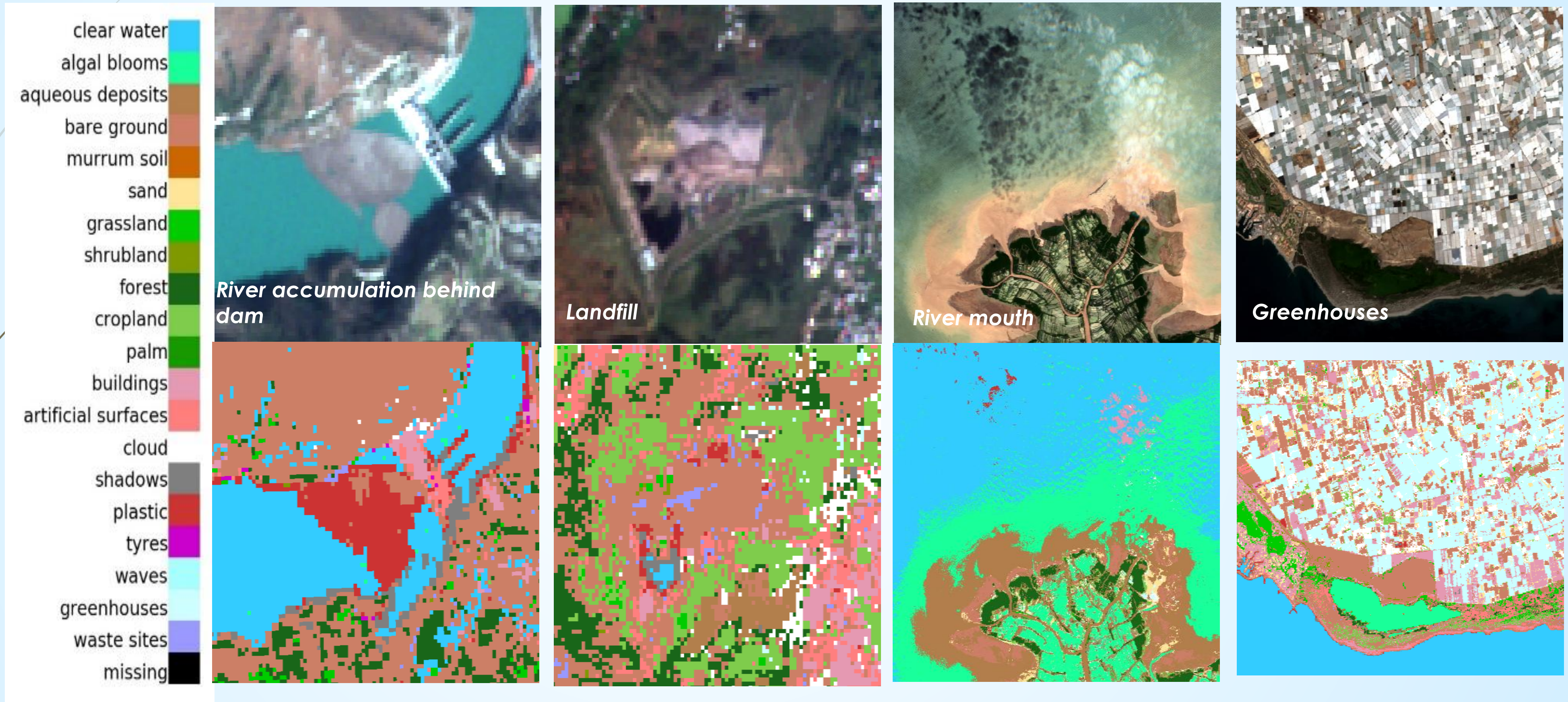




Plastic detection overlaid on Google Earth

There are multiple detections off Pulau Opak Besar Timur that appear to be aquaculture cages, with the strong detection on the right of the island being an eco-tourism resort.

Expanding the detection to multiple sources of waste in the environment

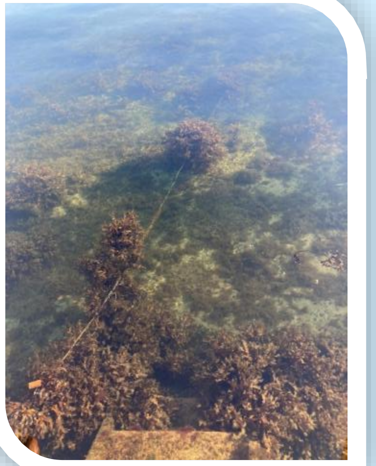


Lavender (2022): https://doi.org/10.3390/rs14194772

Details of the training dataset has been uploaded to Ocean Scan: https://www.oceanscan.org/dataset/a8375686-76c5-42d4-936b-2ad8bd59118e

Mapping Eelgrass Meadows in Swedish Coastal Areas – Earth Observation in support of Blue Carbon Research (SwedCoast-BlueCarb)

- The aim of the SwedCoast-BlueCarb project (https://www.spaceclimateobservatory.org/swedcoast-bluecarb) is to map blue carbon (eelgrass/seagrass and macroalgae) and bottom substrates by combining satellites, field data, and specialist knowledge.
- Blue Carbon habitats provide spawning grounds for fish and are crucial for healthy and self-sustaining fisheries, but are being impacted by our changing climate.
- The Baltic Sea is optically dominated by coloured dissolved organic matter, which makes the use of remote sensing to map underwater vegetation challenging.



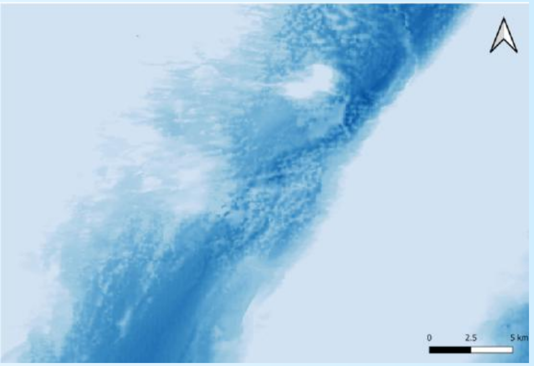
Fucus vesiculosus and *Zostera marina*, courtesy of Lena Kautsky and Maria Asplund

Rymdstyrelsen Swedish National Space Agency
UK SPACE AGENCY
SPACE4CLIMATE Uniting UK Space-enabled Climate Expertise and Services
SCO SPACE FOR CLIMATE OBSERVATORY
Stockholm University

Earth Observation (EO) Usage

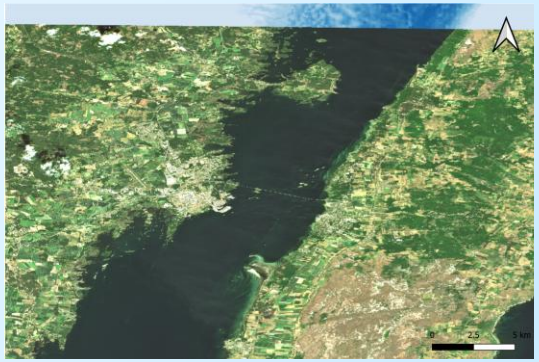
- **Operational tool:** Processing and visualisation of the EO-derived product through an interactive platform
- **Source data:**
  - **Copernicus Sentinel-2:** Baseline mapping
  - **Maxar WorldView-2:** Assessment of commercial high-resolution data
  - **Copernicus Sentinel-3:** Deriving water quality information
  - **Digital Terrain Model (DTM) / Bathymetry Data:**
    - EMODnet Bathymetry 2022 [~115 m spatial resolution] - Mean depth full coverage with water depth to Lowest Astronomical Tide (https://doi.org/10.12770/ff3aff8a-cff1-44a3-a2c8-1910bf109f85)
    - Delta DTM [~30 m spatial resolution] - vertical biases of DEM removed (https://doi.org/10.1038/s41597-024-03091-9)
  - **Input to atmospheric correction**:
    - MERRA-2 for aerosols (https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/)

DTM/Bathymetry

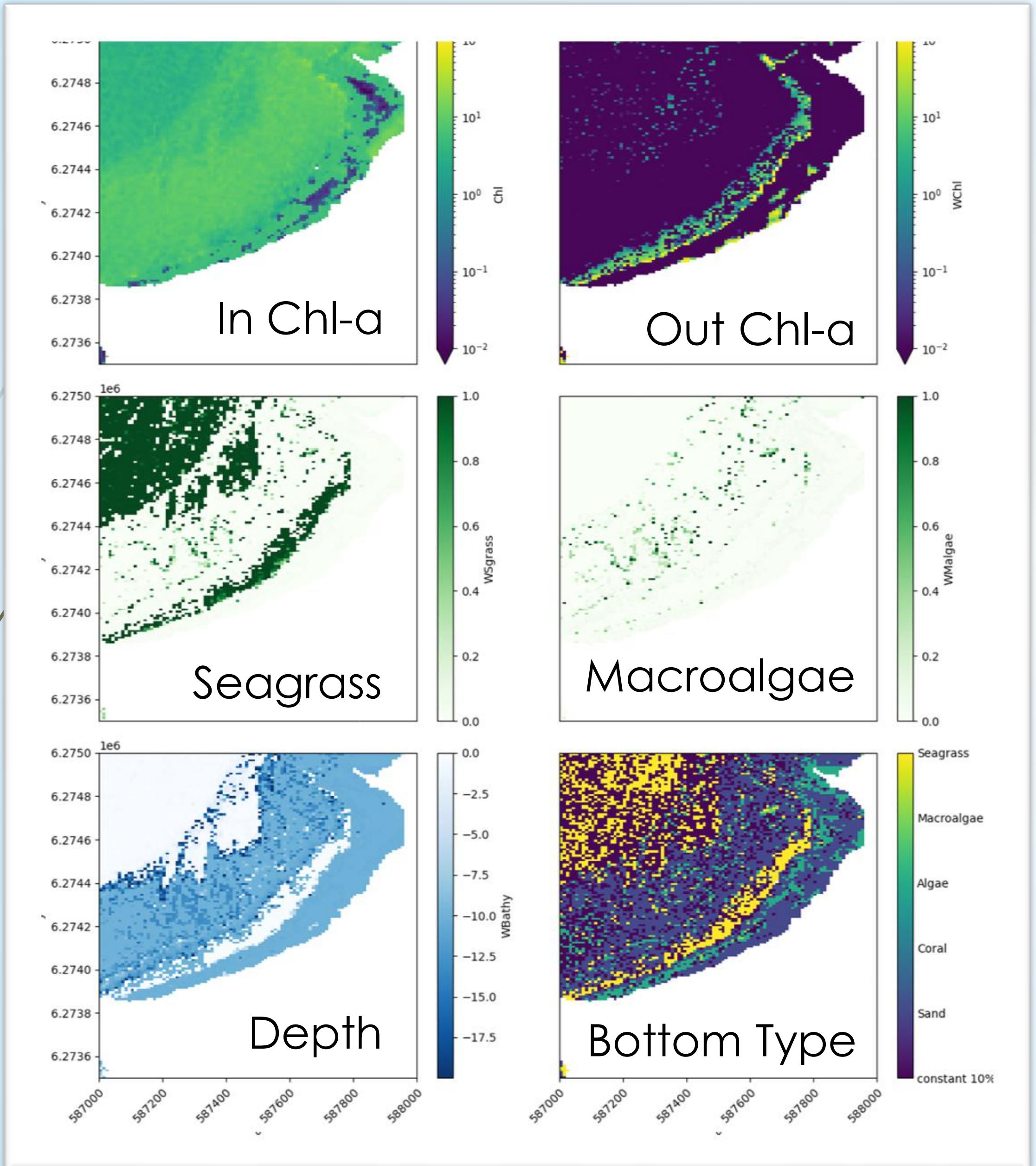


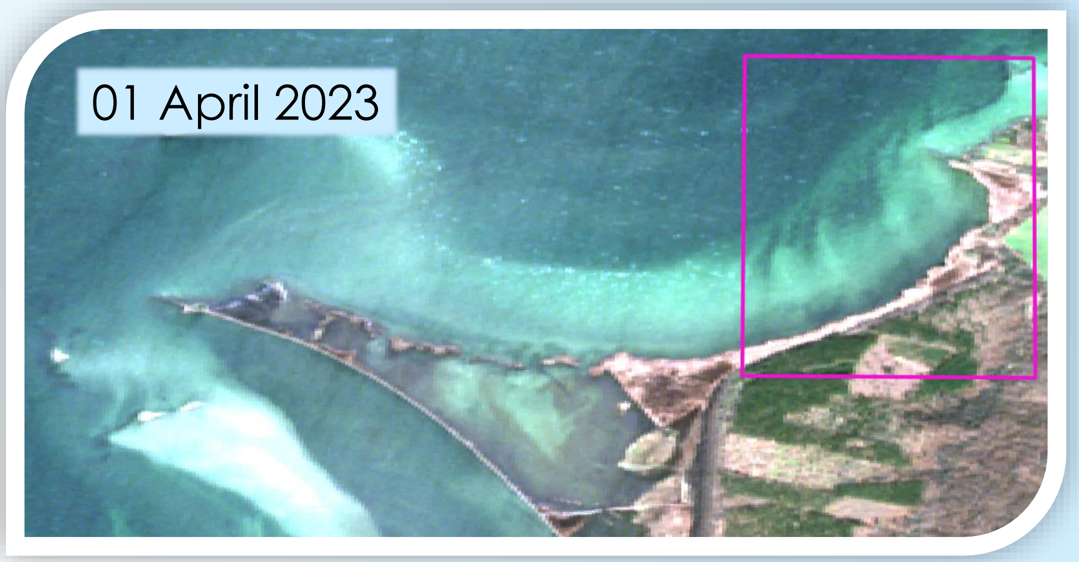
Sentinel-2 March 2023

Sentinel-2 June 2023



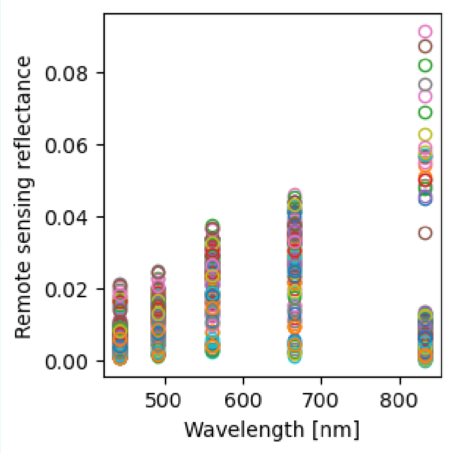
Preliminary Results



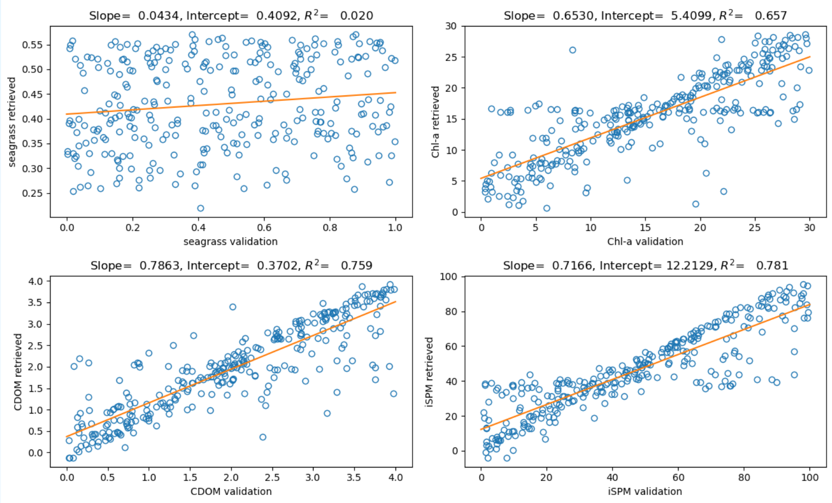


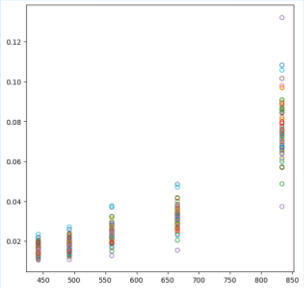
- Mask for processing is currently based on input Chl-a retrieval being valid (Goyens algorithm within ACOLITE)
- Modified König WaterQuality python package that uses the WASI (Gege) model for retrievals
- Runs reasonably well, but is too slow for operational use

Testing a Gaussian Process Emulator

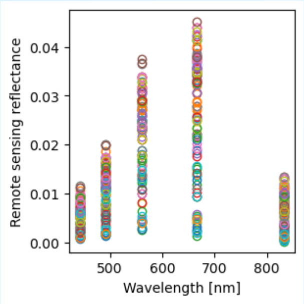


Full range of spectra generated

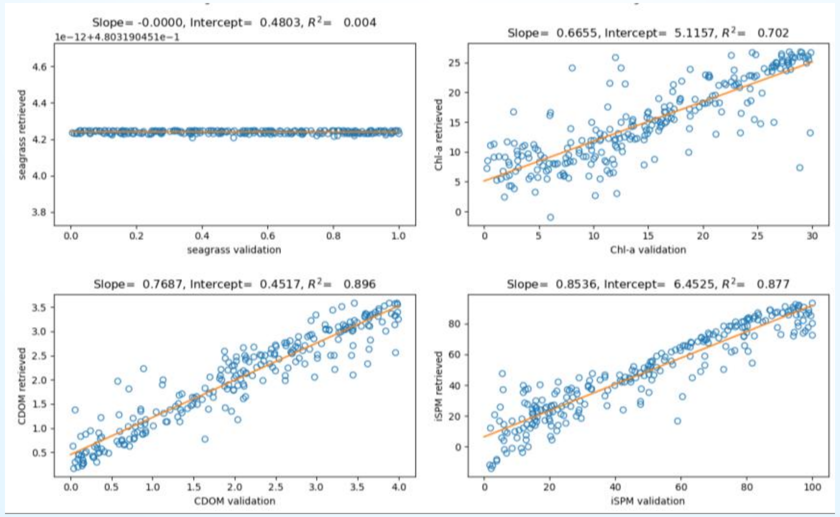


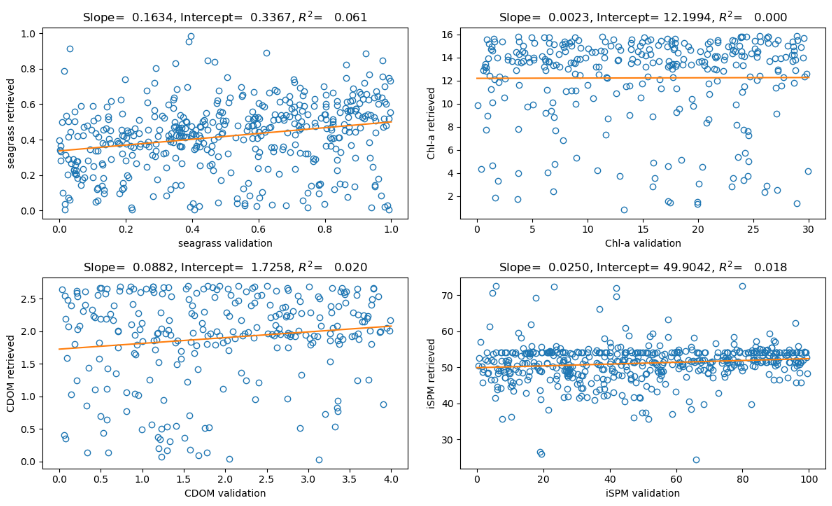


Shallow water subset



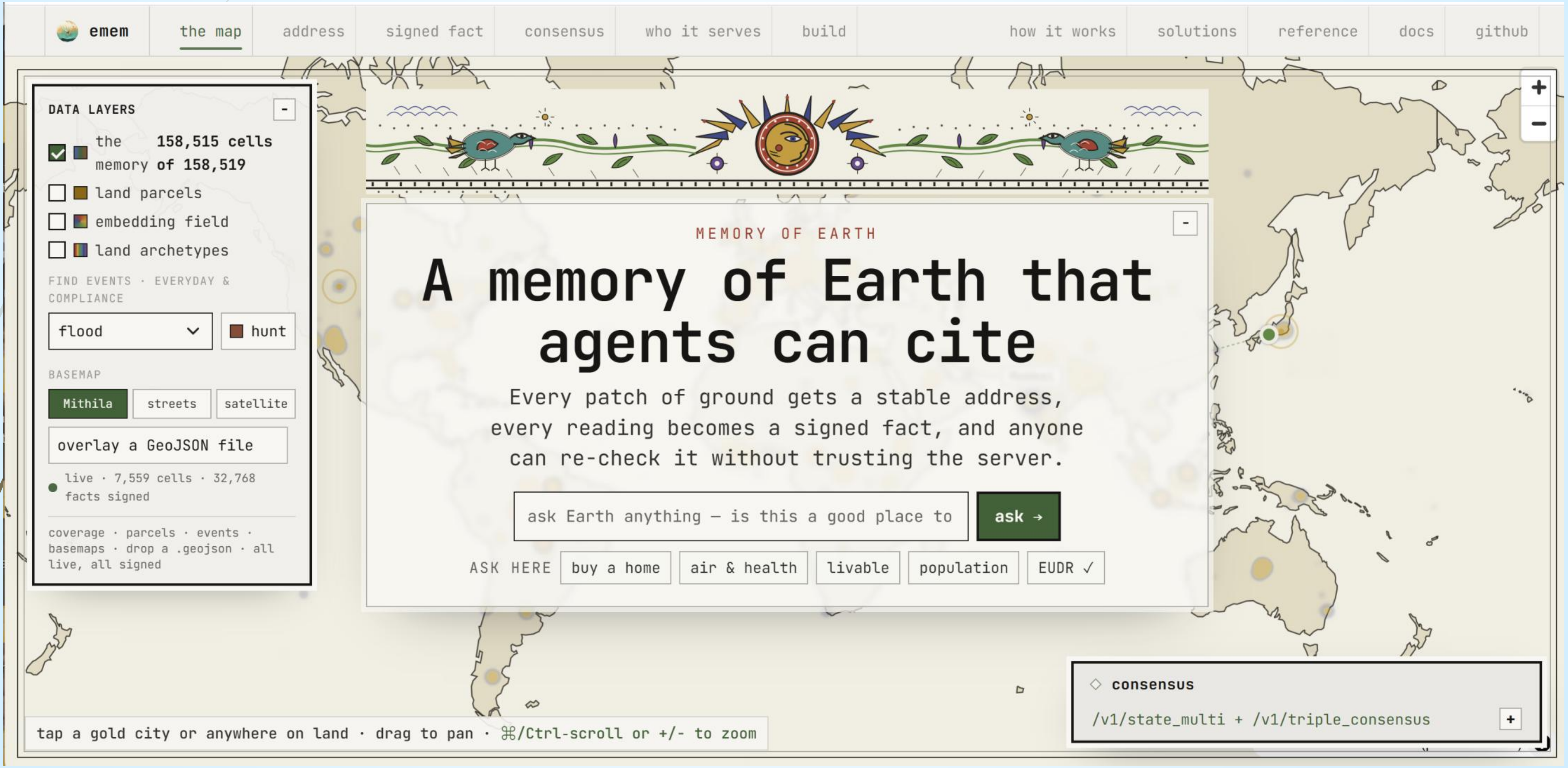
Deep water subset





The power of data and compute, with embedded trust (an example)

https://emem.dev/



How we got here and where do we go?

- **Symbolic:** Writing out all the rules, which is very hard
- **Expert systems:** Data analysis
- **Machine Learning:** Learning from patterns in the data

- More data, bigger models, more compute to get generative AI
- **Negative Consequences:** impact on water/energy/greenhouse gases

- **Positive Actions:**
  - Analyse data/produce predictions we've not been able to before
  - Balance resource usage/output accuracy:
    - Prune large models to make them smaller
    - Apply quantisation, i.e. round up the numerical accuracy
  - Societal solutions
    - Democratisation - datasets sourced from all over the world
    - Accountability

Turing Institute Lectures, https://www.turing.ac.uk/events/event-type/lecture