decent exposure

deforestation and physical assets

Rob Norris, Martina Kaplanová

Can we use location and type of industrial assets to predict deforestation?

Overview of project

- Product: tool for deforestation assessment of an asset portfolio

- Data:
  - Industrial assets
  - Deforestation

- Tech Stack:
  - Geospatial: GeoPandas, Rasterio, rioxarray, Cartopy
  - Modelling: scikit-learn, XGBoost
  - App: Streamlit

Overview of task and stakeholders

our goal
vs
Climate & Co. goal

- EU regulation:
  - Declare how business practices relate to deforestation
  - Includes financial institutions
  - **How to assess "riskiness" of a portfolio?**

- Climate & Company:
  - Work with Swedish Pension Fund
  - Divest deforestation-related assets

- Our interest:
  - Geospatial
  - Climate-related
  - ML exercise

climate & company

Geospatial ESG (Environment, Social, Governance)

The location/s of an **asset** or a company's asset and their suppliers' assets are geolocated. Known as **asset data**, once defined these locations or areas can be compared or modelled with **observational data** – datasets that provide insight.

Within the environmental space, these might provide insights into variables such as a factory's heat profile as a proxy for power usage, methane emissions, or direct impacts to the natural world such as by considering overlays with protected areas, **deforestation**, habitat fragmentation, endangered species, habitat connectivity, biodiversity, etc.

Data

Data I: Deforestation

TIFF (Tag Image File Format): stores raster graphics and image information.

GeoTIFF: allows georeferencing information to be embedded within a TIFF file.

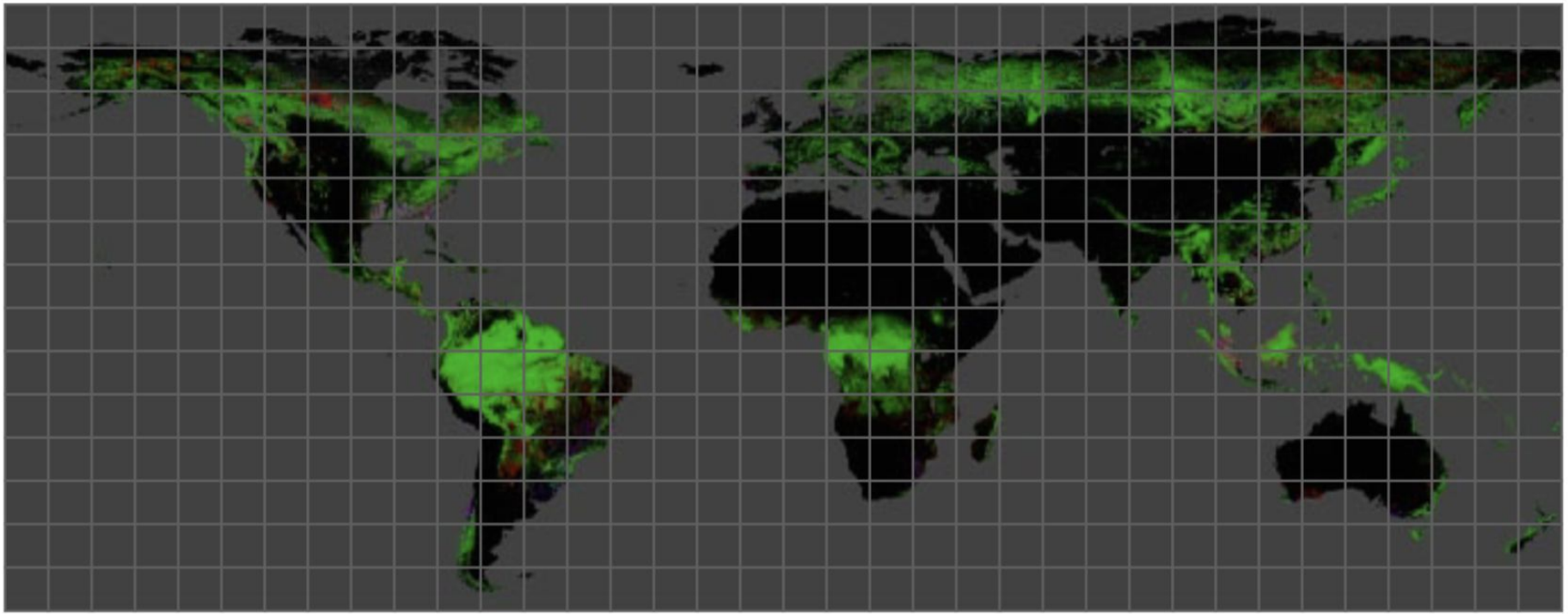
The potential additional information includes:

- map projection
- coordinate systems
- …

Data I: Deforestation

- **Global Forest Change 2000-2022**:
  - time-series of Landsat images
  - global forest extent and change

- Information included (**GeoTIFF** layers of 10 x 10 degrees):
  - **treecover2000:** Tree canopy cover for year 2000
  - **lossyear:** Year of gross forest cover loss event



Data II: Asset Data

---

3 sources suggested by Climate & Co.:

- **Global Energy Monitor (GEM)**: energy-related assets
- **Climate Trace (CLT)**: mining assets
- **Spatial Finance Initiative (SFI)**: heavy industry assets

EDA

EDA I: Deforestation

- **Global Forest Change 2000-2022:**
  - 60+GB of data, globally
  - Each GeoTIFF file covers 10 x 10 degrees in 40K x 40K pixels

- Initial plan: time-series
  - ~2 hours processing time on MacBook Pro (Retina, 15-inch, Mid 2015)

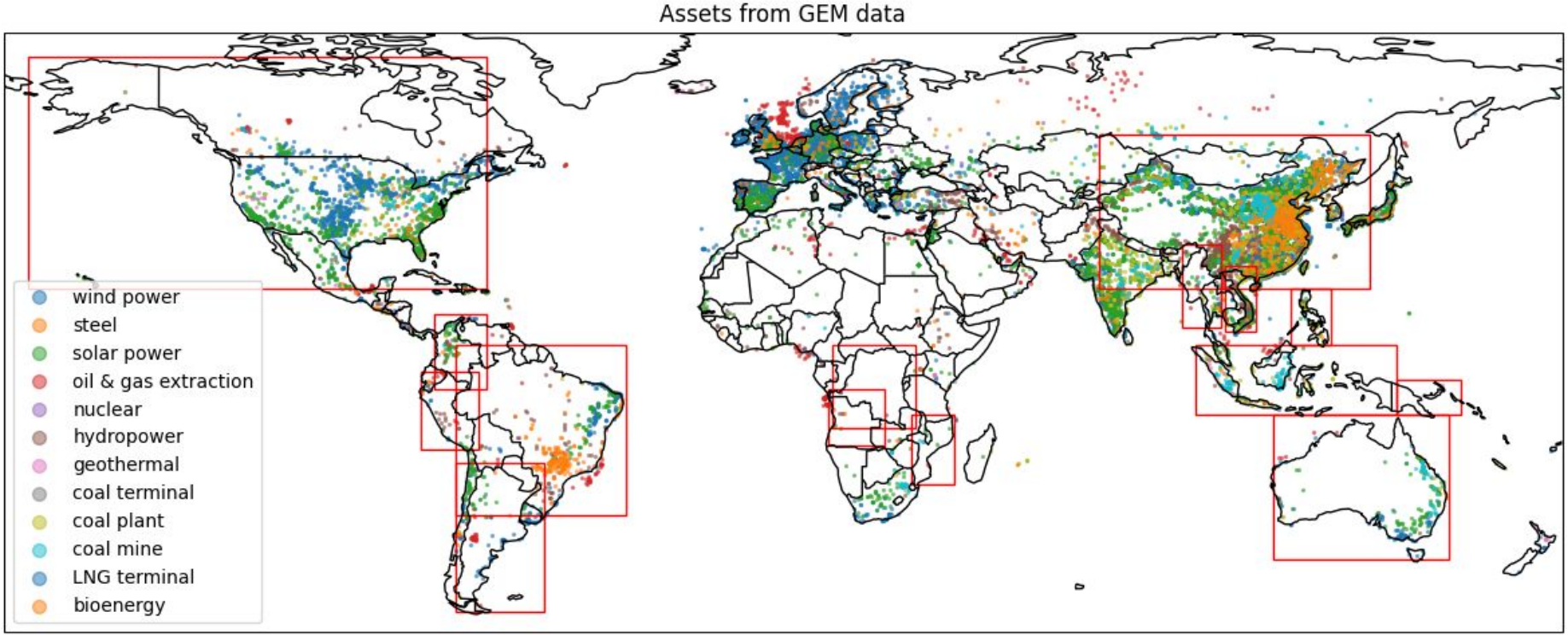
- Final plan: Regression
  - Top 15 countries with the most deforestation, plus USA and China.
  - ~30 mins download, and ~5 mins processing time on MacBook Pro (Retina, 15-inch, Mid 2015)

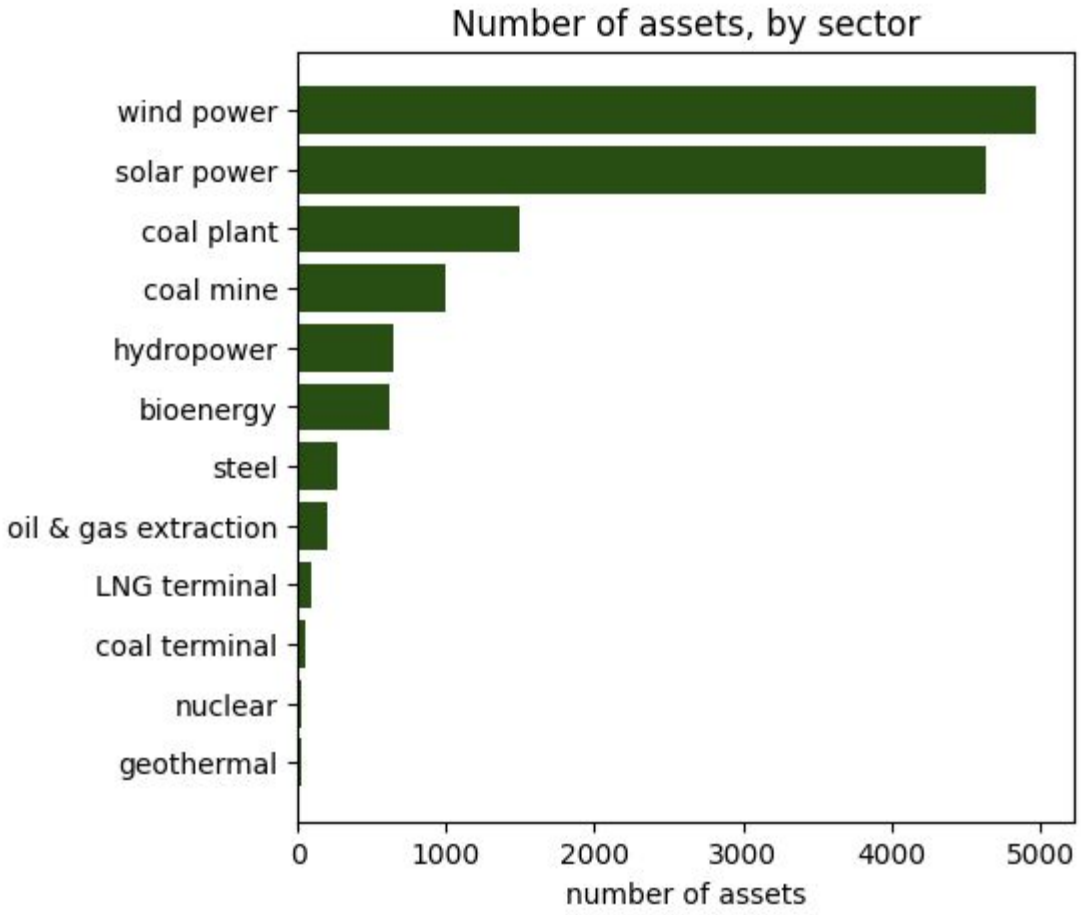
EDA II: Asset Data

- In total: ~24k assets

- Focus on at-risk areas of deforestation: 16,029 assets

- Information included in GEM data:
  - asset ID and asset owner
  - latitude and longitude
  - sector
  - year of start of operation
  - size/subunits of industrial asset


Assets from GEM data
wind power
steel
solar power
oil & gas extraction
nuclear
hydropower
geothermal
coal terminal
coal plant
coal mine
LNG terminal
bioenergy


Number of assets, by sector
wind power
solar power
coal plant
coal mine
hydropower
bioenergy
steel
oil & gas extraction
LNG terminal
coal terminal
nuclear
geothermal
0
1000
2000
3000
4000
5000
number of assets

EDA III: Combined data

Subset to areas with:

1) with non-zero tree cover in 2000
2) with non-zero deforestation between 2000 and 2022

Summary stats:

- 25+% assets: no deforestation in a 3-year window
- 75+% assets: <1% of area deforestation

Product:
tool for deforestation assessment of an asset portfolio

tool for deforestation assessment of an asset portfolio

- **Start:** list of assets in the portfolio

- **Goal:** deforestation by asset
  - in an X sqkm area _around_ the assets
  - in a Y years around the start of operation of the asset

- **Idea:** identifies the clearing the forest in the vicinity of the asset

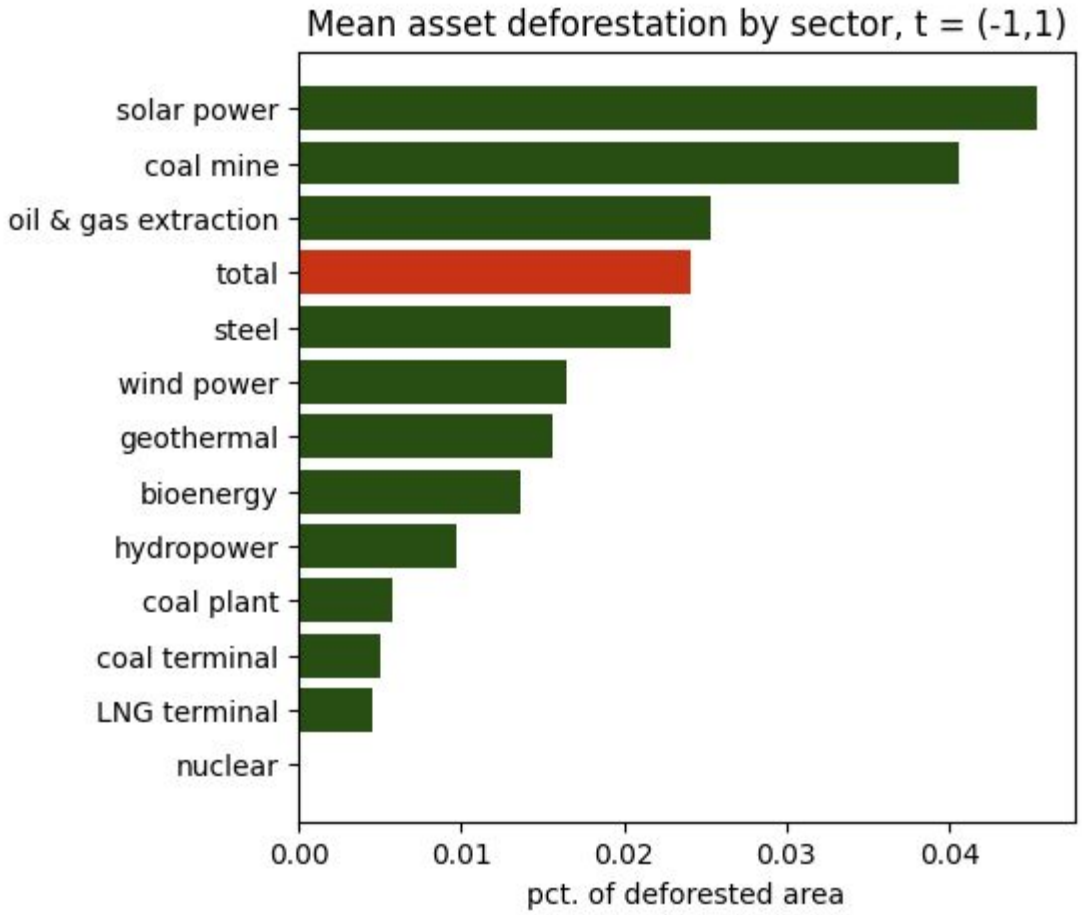
- **Value:** heuristic of assets' **exposure** to deforestation

Machine Learning Exercise

Does introduction of a new industrial asset contribute to deforestation in the area?

(1) Does this differ by sector? → YES

(2) Can we predict the deforestation with confidence? → NO :(


Mean asset deforestation by sector, t = (-1,1)
solar power
coal mine
oil & gas extraction
total
steel
wind power
geothermal
bioenergy
hydropower
coal plant
coal terminal
LNG terminal
nuclear
0.00
0.01
0.02
0.03
0.04
pct. of deforested area

Method description: asset info to predict canopy loss

---

1. How to choose a right outcome variable?
2. How to choose the right timing?

→ **TRADE OFF: number of observations vs intensity of deforestation**

- **Regression problem**: example on linear regression

$$\textit{canopy change} = a + b * [\textit{sector information}] + c * [\textit{start date}] + d * [\textit{asset size info}] + \textit{error}$$

Results

Tried:
- Different models
- Feature engineering
- Outcome engineering
- Hyperparameter tuning

Baseline: the prediction is the mean

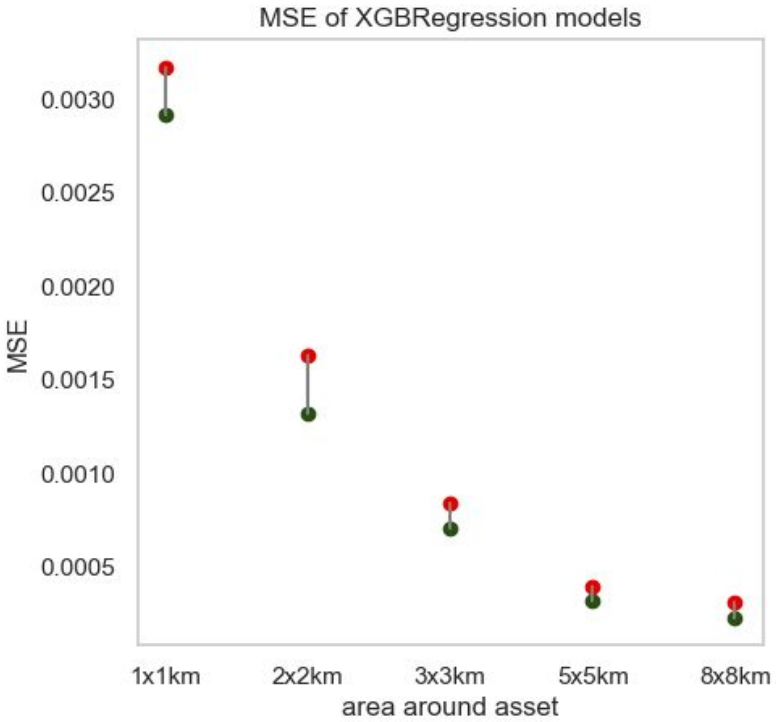
	1x1 km	**8x8 km**
MSE	0.00486	0.00035
R2	<0.00001	-0.00092

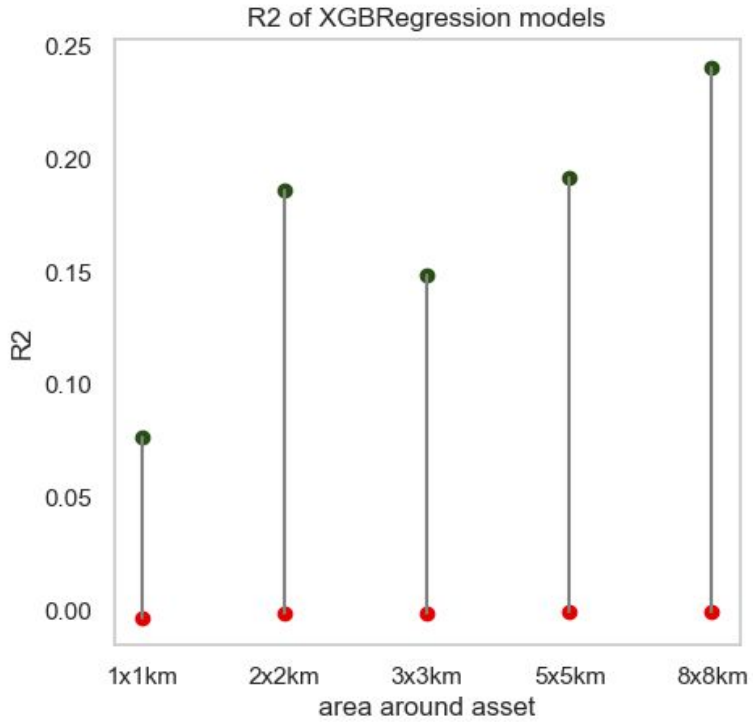
Best model: **XGBRegression** with regression trees

	1x1 km	**8x8 km**
MSE	0.00431	0.00027
R2	0.11387	0.21565

Results

Tried:
- Different models
- Feature engineering
- Outcome engineering
- Hyperparameter tuning





Further suggestion

- cover more areas of deforestation
- gather data on other types of assets from other datasets
- include more data on asset surrounding (geography, distance to infrastructure, distance to urban areas)
- alt. outcome: distance to closest area of deforestation

Challenges

(1) **Data limitations**

- Value chain of the assets vs the operations of the asset itself

→ **accessible data does not mean relevant or usable data**

(2) **Time series or not?**

→ **planning helps, but dead ends are sometimes inevitable!**

(3) **Own feels: we do *only* (poorly predicting) regression**

- A little anticlimactic after the course

→ **we have done a *lot* of processing work**

→ **some things are not easy to model without better data/more time :)**

Thank you for your attention!