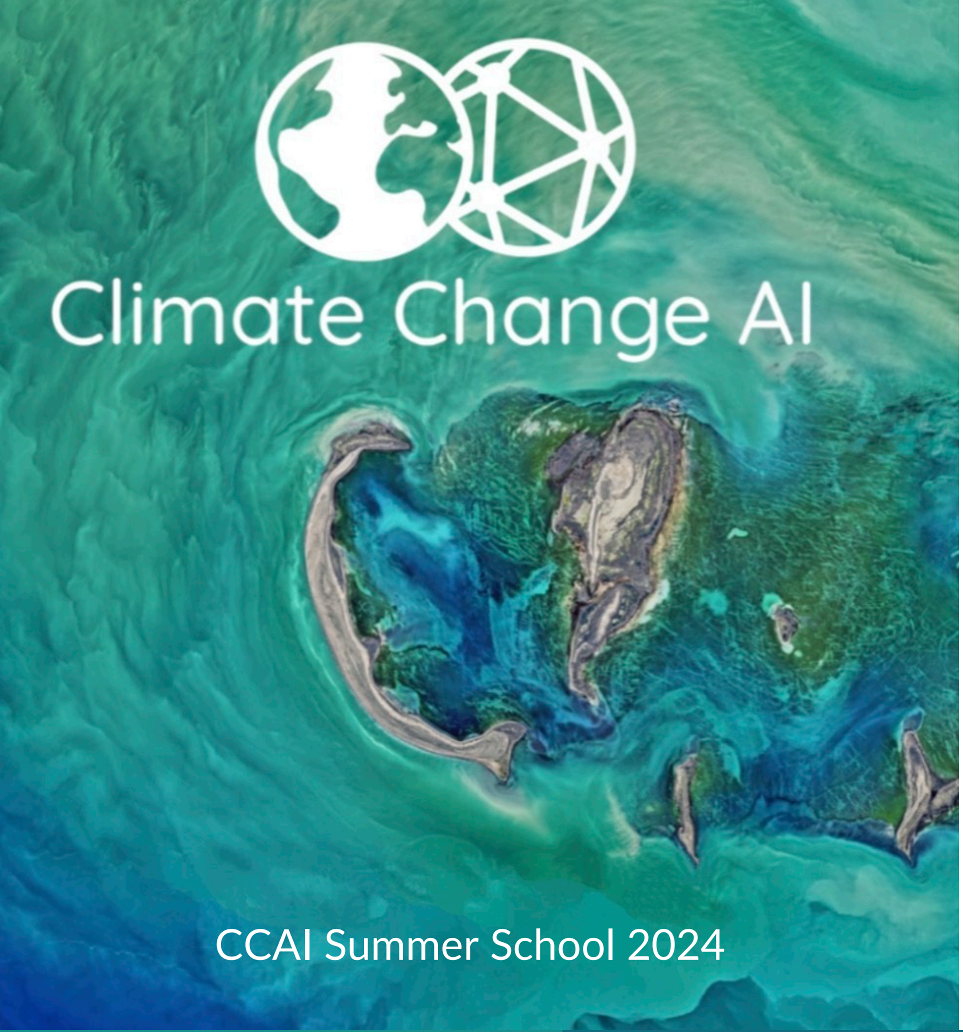


Machine learning for Emissions Accounting and Monitoring

Dr Kasia Tokarska de los Santos

CarbonPool

Who am I?

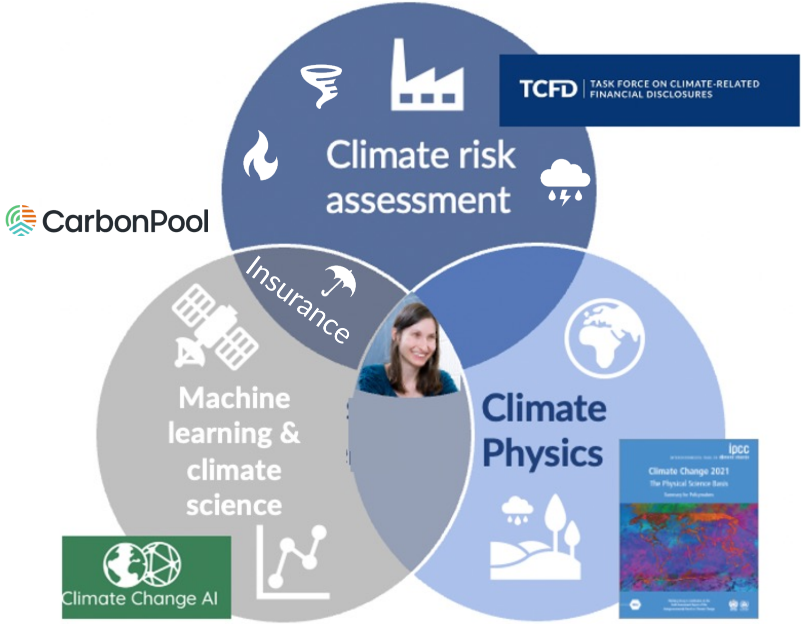
PhD in Climate modelling (University of Victoria, Canada)

Postdoctoral research (ETH Zurich & Uni of Edinburgh)
- IPCC AR6 contributing author
- Working with CMIP5 and CMIP6 models

Industry experience
- Climate risk modelling, climate impacts and corporate disclosures (TCFD, TNFD)
- Carbon removal insurance, nature-based removals (CarbonPool)

Climate Change AI
* Bridging ML and climate science communities

Current position:
Principal Climate Modeller at CarbonPool
(carbon offset insurance)



Today's Outline

Part 1: Motivation for emission accounting

- Introduction of physical and transition risk
- Introduction to GHG Emissions in the industry
- Example of Scope 3 emissions accounting

Part 2: Different approaches to emission estimation

- Traditional estimates (bottom-up)
- Data-driven estimates
- Satellite data (Top-down estimates)

Part 3: Applications of machine learning for emissions monitoring

- Monitoring of nature-based carbon removals and offsets:
- ML applications in monitoring biomass changes and deforestation (link to the CCAI lecture on MRV for forestry)

Part 4: Challenges and Future Directions

Part 1: Motivation for emission accounting

In this section you will learn about:

- Introduction of physical and transition risk
- Introduction to GHG Emissions in the industry
- Example of Scope 3 emissions accounting

Physical Risks

Wildfires

Heatwaves

Hurricanes

Drought

Sea level rise

Flooding

Reducing emissions lowers physical risks

Increasing severity and frequency of different events, such as:

- Storms, droughts, fires, flooding, extreme heat, etc.
- Uneven impacts
- Feedbacks (e.g., permafrost thaw)

Physical Risks

Wildfires

Heatwaves

Hurricanes

Drought

Sea level rise

Flooding

Reducing emissions lowers physical risks

Transition Risks

Policy

Regulation

Liability

Market

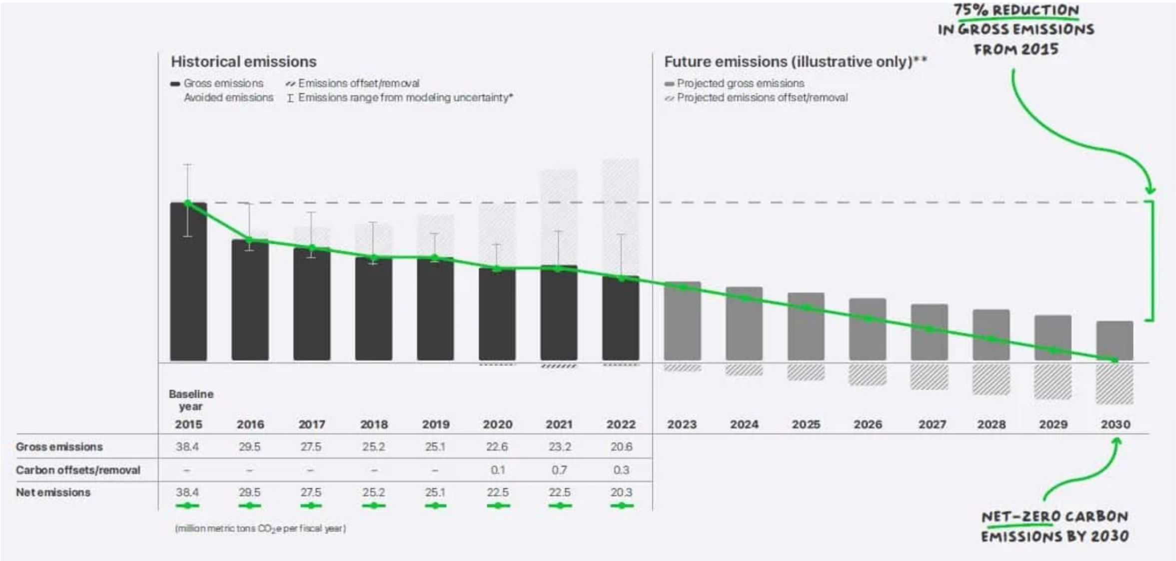
Reputation

Technology

Reducing emissions lowers transition risks

Meeting net-zero targets by 2030…. or later

- Countries and individual companies have ***net-zero commitments.***
- Reaching net-zero requires the ***reduction*** of most emissions.
- Emissions that are difficult to abate need to be ***removed***.
- **Emissions accounting** is necessary at both global/country levels and individual company levels to determine excess emissions relative to net-zero targets.



Source: Apple Net Zero emissions road map.
https://carboncredits.com/apple-reveals-first-ever-carbon-neutral-watch-aims-to-offset-25-product-emissions-with-carbon-credits/

Greenhouse gas emissions by sector

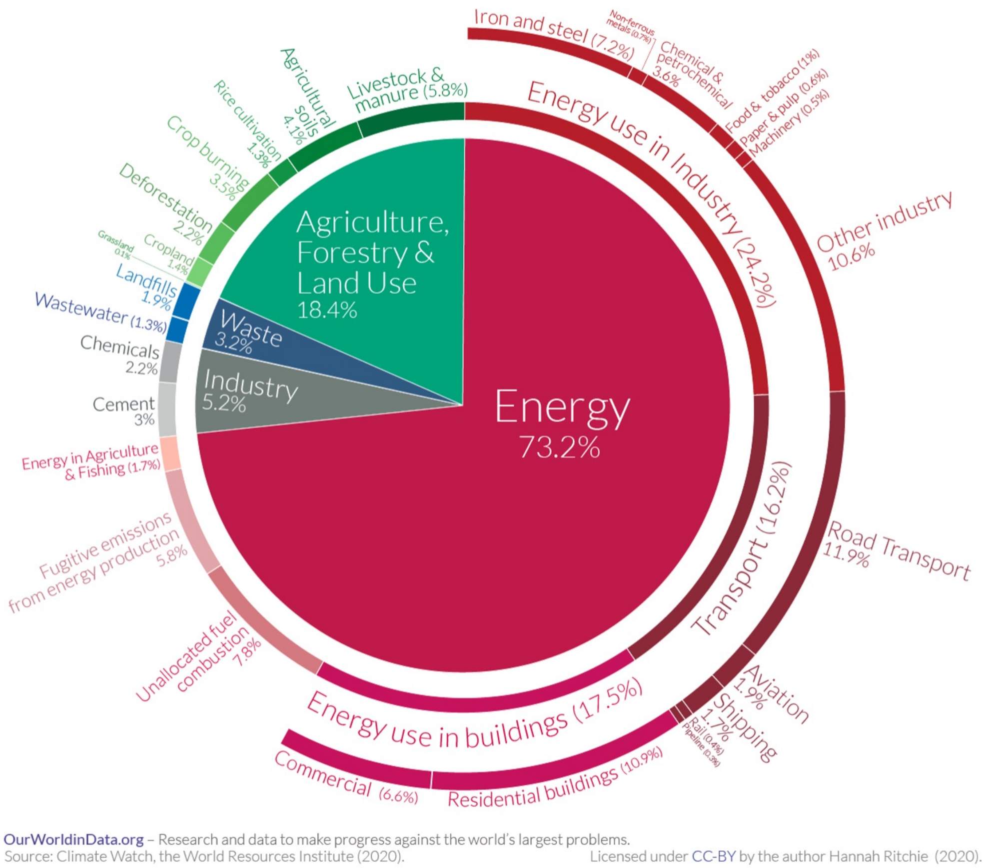


Figure source: Our World in Data, 2016

Scope 1, 2, 3 Emissions

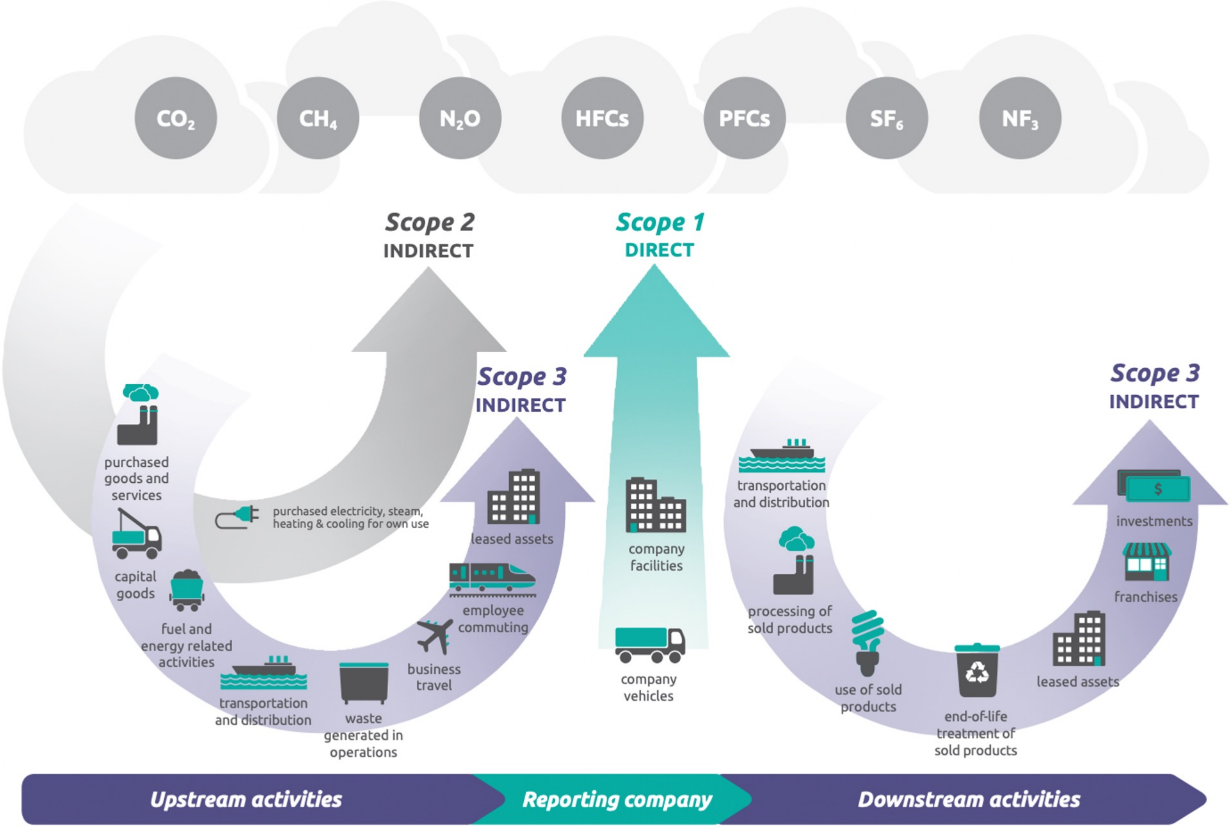


Figure source: GHG protocol. Figure 1.1.

Quick Poll #1

Which scope contributes most of the emissions?

- ❑ Scope 1
- ❑ Scope 2
- ❑ Scope 3

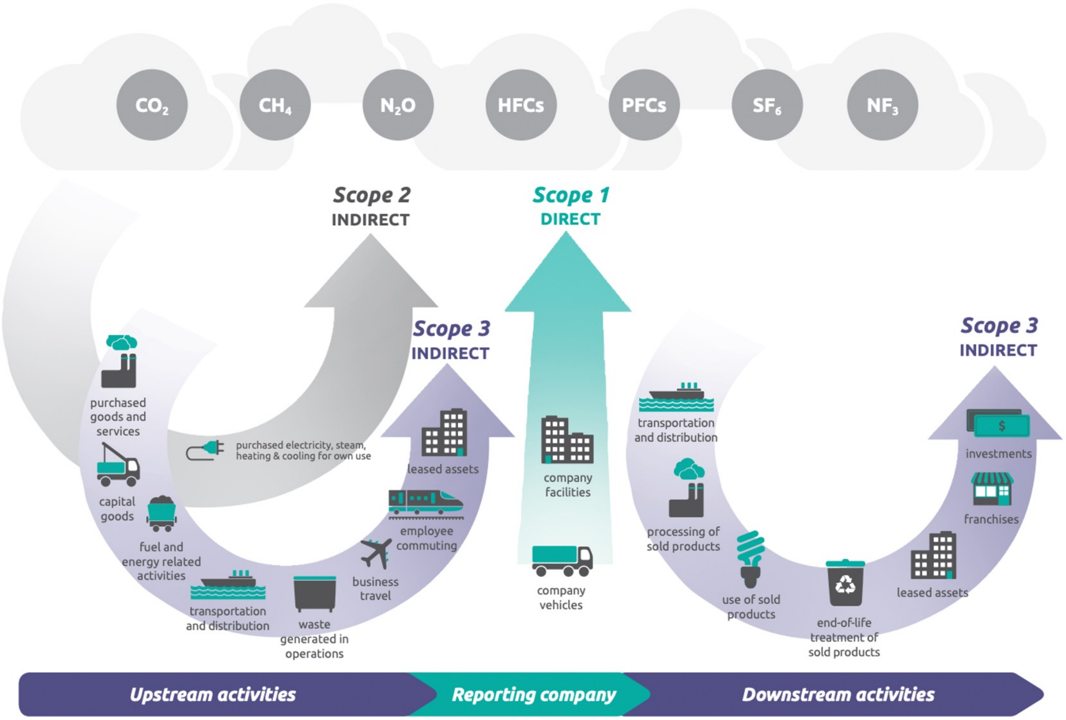


Figure source: GHG protocol. Figure 1.1.

Scope 1, 2, 3 Emissions

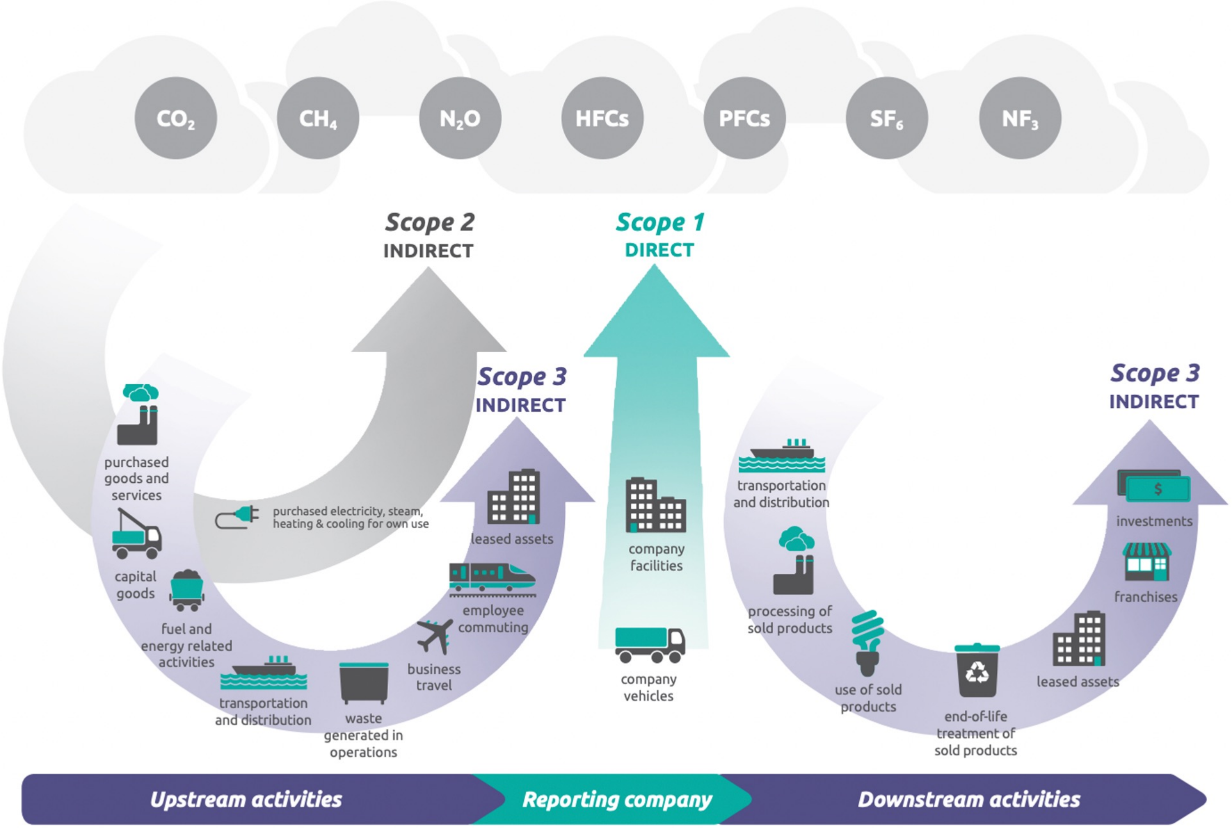


Figure source: GHG protocol. Figure 1.1.

Scope 1, 2, 3 Emissions

Emissions type	Scope	Definition	Examples
Direct emissions	Scope 1	Emissions from operations that are owned or controlled by the reporting company	Emissions from combustion in owned or controlled boilers, furnaces, vehicles, etc.; emissions from chemical production in owned or controlled process equipment
Indirect emissions	Scope 2	Emissions from the generation of purchased or acquired electricity, steam, heating, or cooling consumed by the reporting company	Use of purchased electricity, steam, heating, or cooling
	Scope 3	All indirect emissions (not included in scope 2) that occur in the value chain of the reporting company, including both upstream and downstream emissions	Production of purchased products, transportation of purchased products, or use of sold products

Figure source: GHG protocol.

Example of GHG accounting

Imagine you are running an ice-cream company "*EarthlySweets*" distributing ice cream to different supermarkets

Scope 1	+	Scope 2	+	Scope 3

Example of GHG accounting

Imagine you are running an ice-cream company "*EarthlySweets*" distributing ice cream to different supermarkets

Scope 1	Scope 2	Scope 3
Running five fleet vehicles (distribution and transportation of ice cream) Heating office spaces using on-site boilers Emissions from the refrigeration equipment used to keep the ice cream cold **Total: 26,600 kg CO_{2e} /year**		

Example of GHG accounting

Imagine you are running an ice-cream company "*EarthlySweets*" distributing ice cream to different supermarkets

Scope 1

Running five fleet vehicles (distribution and transportation of ice cream)

Heating office spaces using on-site boilers

Emissions from the refrigeration equipment used to keep the ice cream cold

Total:
26,600 kg CO_{2e} /year

+

Scope 2

Electricity use:
20,000 kWh/year

Includes:

- lighting,
- air conditioning of office spaces,
- energy used for keeping the ice-cream cool (ice-cream storage

Total:
8,000 kg of CO_{2e}

+

Scope 3

Quick Poll #2

What would be an example of Scope 3 emissions for the Earthly Sweets ice-cream company?

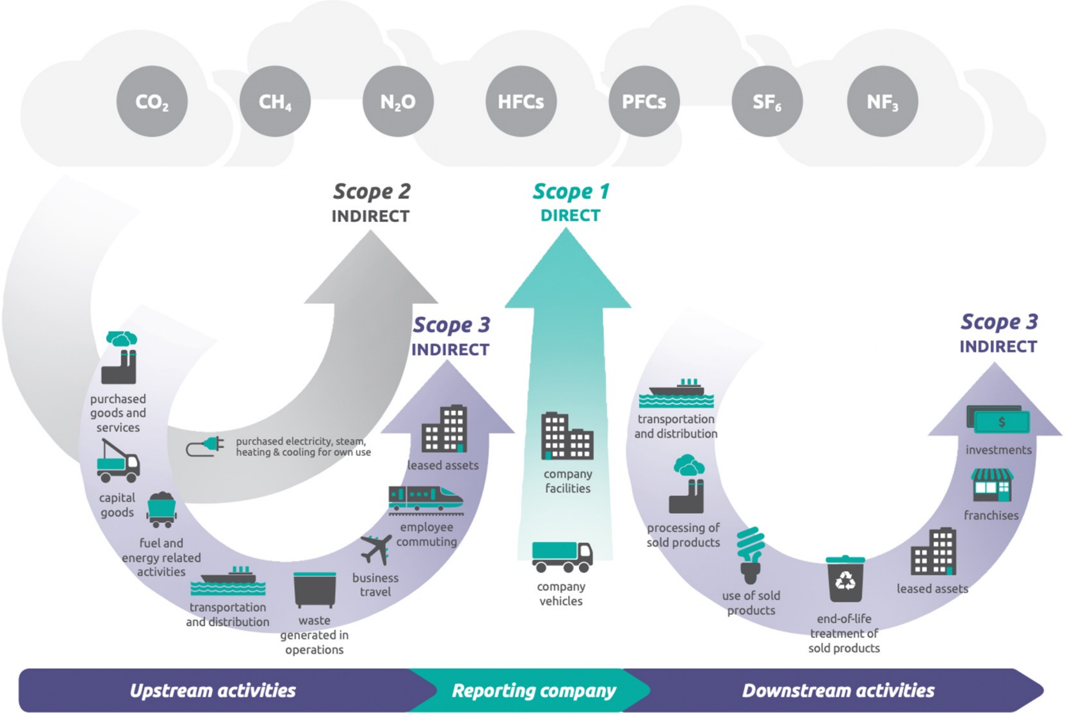


Figure source: GHG protocol. Figure 1.1.

Example of GHG accounting

Imagine you are running an ice-cream company "*EarthlySweets*" distributing ice cream to different supermarkets

Scope 1

Running five fleet vehicles (distribution and transportation of ice cream)

Heating office spaces using on-site boilers

Emissions from the refrigeration equipment used to keep the ice cream cold

Total:
26,600 kg CO_{2e} /year

+

Scope 2

Electricity use:
20,000 kWh/year

Includes:
- lighting,
- air conditioning of office spaces,
- energy used for keeping the ice-cream cool (ice-cream storage

Total:
8,000 kg of CO_{2e}

+

Scope 3

Business travel (e.g., flights to ice cream conferences)

Emissions from employee commuting to office

Emissions from the production and transportation of the raw ingredients used to make the ice cream.

Total:
1150,333 kg of CO_{2e}.

Scope 1 emissions of one company are Scope 3 emissions of another company

	Extraction, processing and transport	Power generation	Transmission & distribution	End user consumption
	Emissions associated with extraction	Emissions from combustion of fuels in power generation	Power losses & consumption of power by utility (10% of total generated power)	Consumption of power by end user (90% of total generated power)
Mining / extraction company	**Scope 1** (5 tCO_2e)	**Scope 3 (use of sold products)** (100 tCO_2e)	-	-
Power generator	**Scope 3 (fuel- and energy-related activities)** (5 tCO_2e)	**Scope 1** (100 tCO_2e)	-	-
Utility	**Scope 3 (fuel- and energy-related activities)** (10% * 5 tCO_2e = 0.5 tCO_2e)	(reported as scope 2)	**Scope 2** (10% * 100 tCO_2e = 10 tCO_2e)	**Scope 3 (fuel- and energy-related activities)** (4.5 tCO_2e + 90 tCO_2e = 94.5 tCO_2e)
End user	**Scope 3 (fuel- and energy-related activities)** (90% * 5 tCO_2e = 4.5 tCO_2e)	(reported as scope 2)	**Scope 3 (fuel- and energy-related activities)** (0.5 tCO_2e + 10 tCO_2e = 10.5 tCO_2e)	**Scope 2** (90% * 100 tCO_2e = 90 tCO_2e)

Source: Supplement to the Reference Guide for the GRESB Infrastructure Asset Performance Component Guidance on Scope 3 Greenhouse Gas Emissions Reporting

Scope 3 emissions tend to be most under-reported

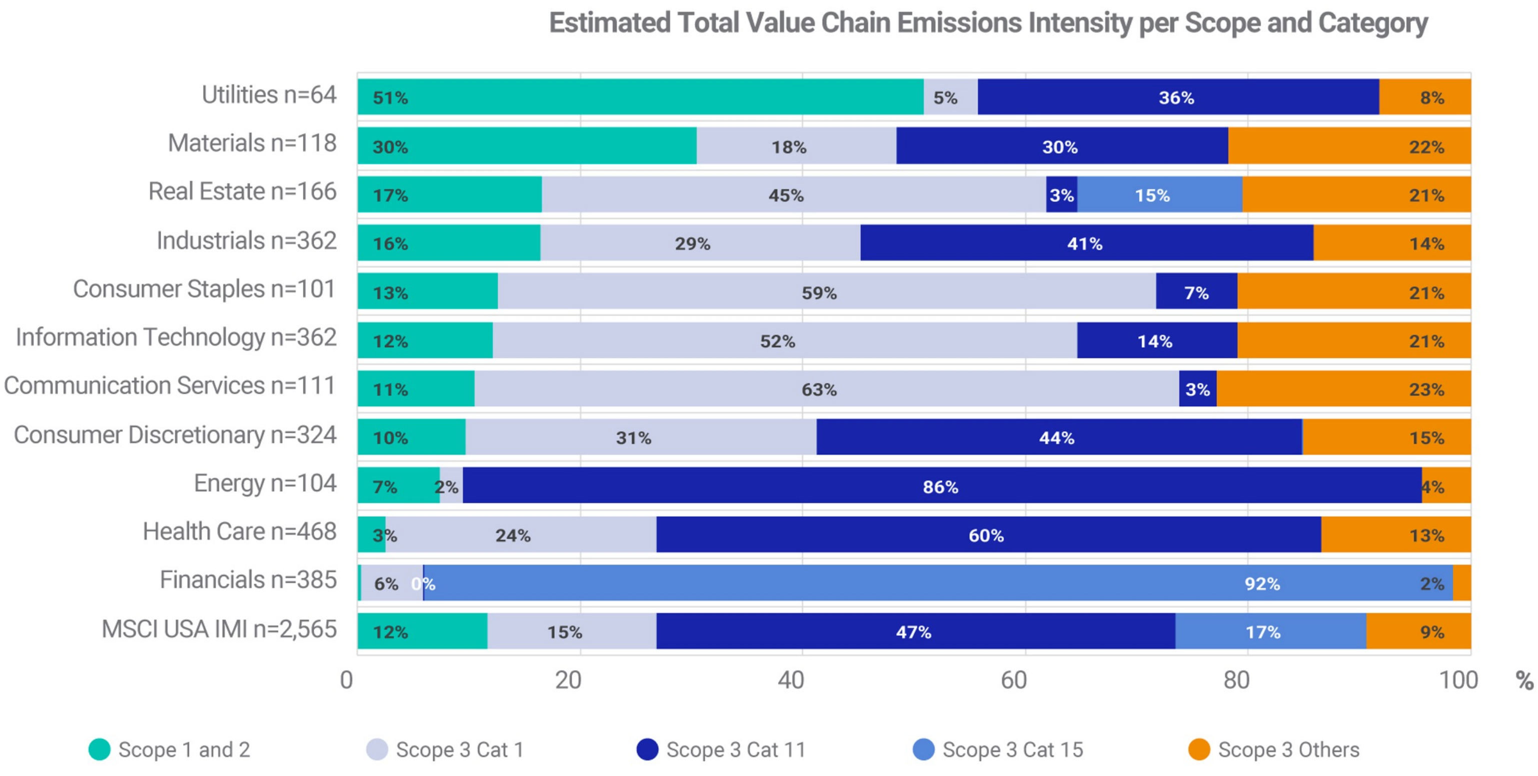


Figure source: MSCI

Part 2: Different approaches to emission estimation

In this section you will learn about:

- Traditional estimates (bottom-up)
- Data-driven estimates
- Satellite data (Top-down estimates)

Machine Learning applications for tackling climate change

Machine learning (ML):

Group of techniques that automatically extract patterns from (large amounts of) data

Strengths

- Doing simple tasks quickly and automatically
- Finding patterns in big datasets
- Optimizing complex systems

Weaknesses

- Sensitive to bad or biased data
- Poor at generalizing if data changes
- Finds correlation, not causation

Source: Climate Change AI

ML for emission accounting

The idea is to make use of already reported emissions from industries in various sectors (e.g., CDP Database) and make inference on emissions for unknown company X in a given sector.

Food company A: Scope 1 (12 GTCO2e) Scope 2 (10 GtCO2e) Scope 3 (130 GtCO2e)

Food company B: Scope 1 Scope 2 Scope 3

ML for emission accounting

The idea is to make use of already reported emissions from industries in various sectors (e.g., CDP Database) and make inference on emissions for unknown company X in a given sector.

Food company A: Scope 1 (12 GTCO2e) Scope 2 (10 GtCO2e) Scope 3 (130 GtCO2e)

Food company B: Scope 1 Scope 2 Scope 3

Quick Poll #3

Is this approach of emissions mapping from one sector to another likely to be accurate?

- ❑ Yes
- ❑ No
- ❑ Sometimes

ML for emission accounting

The idea is to make use of already reported emissions from industries in various sectors (e.g., CDP Database) and make inference on emissions for unknown company X in a given sector.

Large international company in California

Food company A: Scope 1 (12 GTCO2e) Scope 2 (10 GtCO2e) Scope 3 (130 GtCO2e)

Food company B: Scope 1 Scope 2 Scope 3

Small family-owned store in Finland

ML for emission accounting

The idea is to make use of already reported emissions from industries in various sectors (e.g., CDP Database) and make inference on emissions for unknown company X in a given sector.

For more accurate predictions, some methods **make use of additional attributes** such as: country of the headquarters, number of employees, company revenue/size, etc.

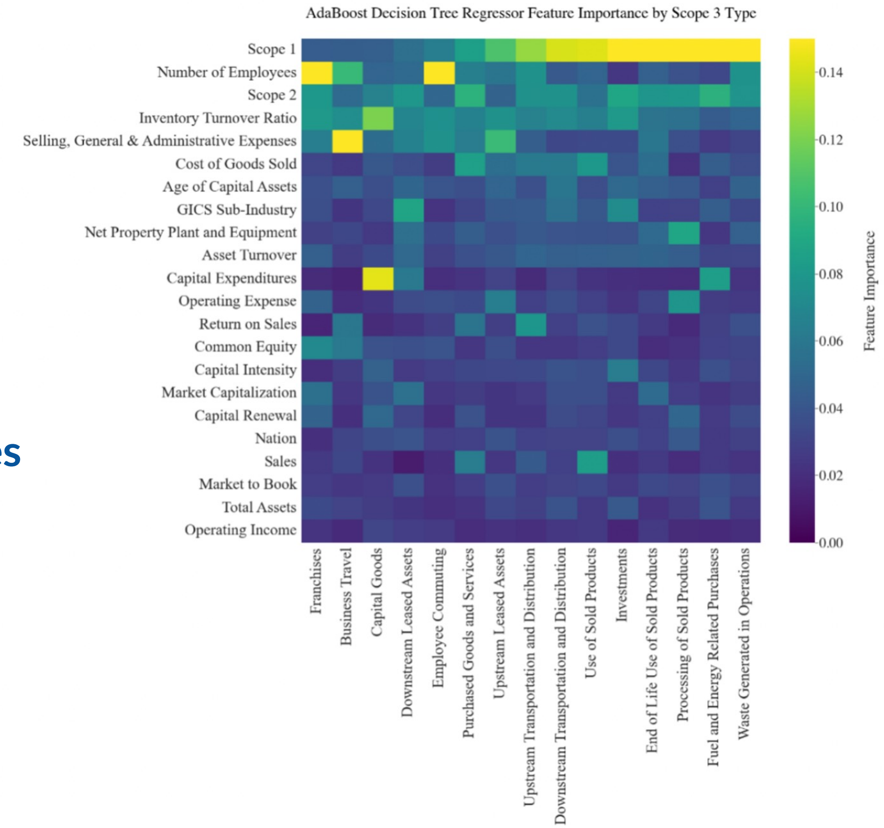
Food company A: Scope 1 (12 GTCO2e) Scope 2 (10 GtCO2e) Scope 3 (130 GtCO2e)

Food company B: Scope 1 Scope 2 Scope 3

Caution: **Data may be biased** towards certain types of companies and some sectors that report emissions.

Scope 3: Boosting methods

Features important for predicting each category of Scope 3 emissions may differ



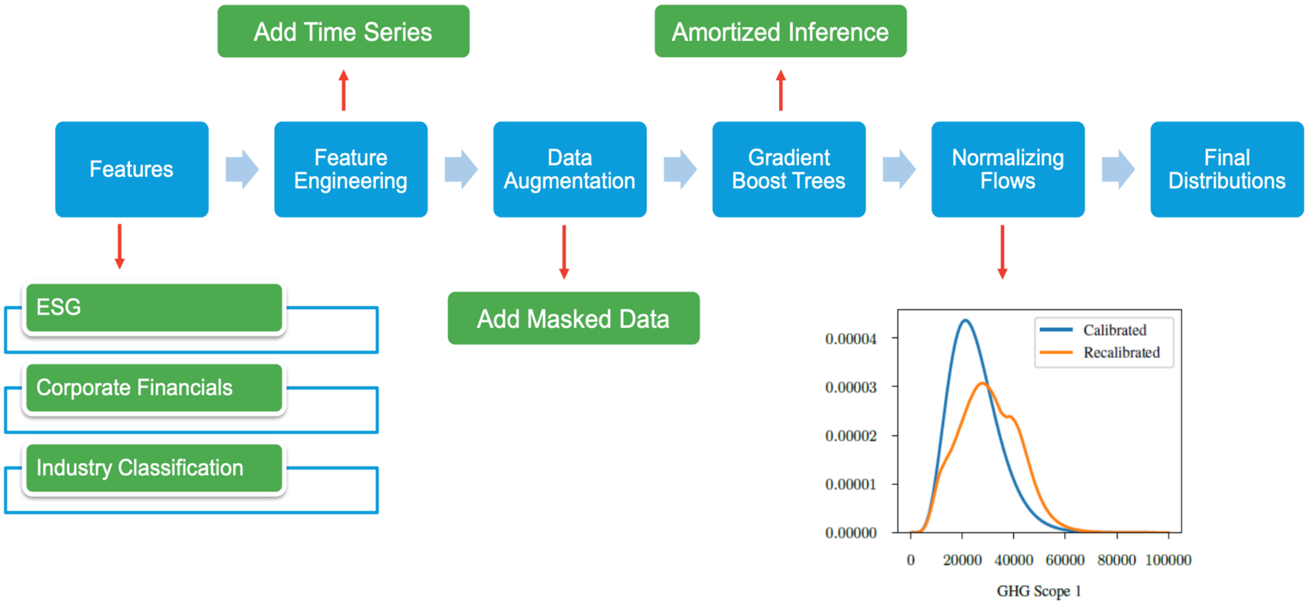
Questions:

Does it make sense to use different features for each individual Scope 3 category?

What are the pros and cons of doing that?

Source: Serafeim, George, and Gladys Vélez Caicedo. "Machine Learning Models for Prediction of Scope 3 Carbon Emissions." Harvard Business School Working Paper, No. 22-080, June 2022.

Scope 3: Inference - based methods

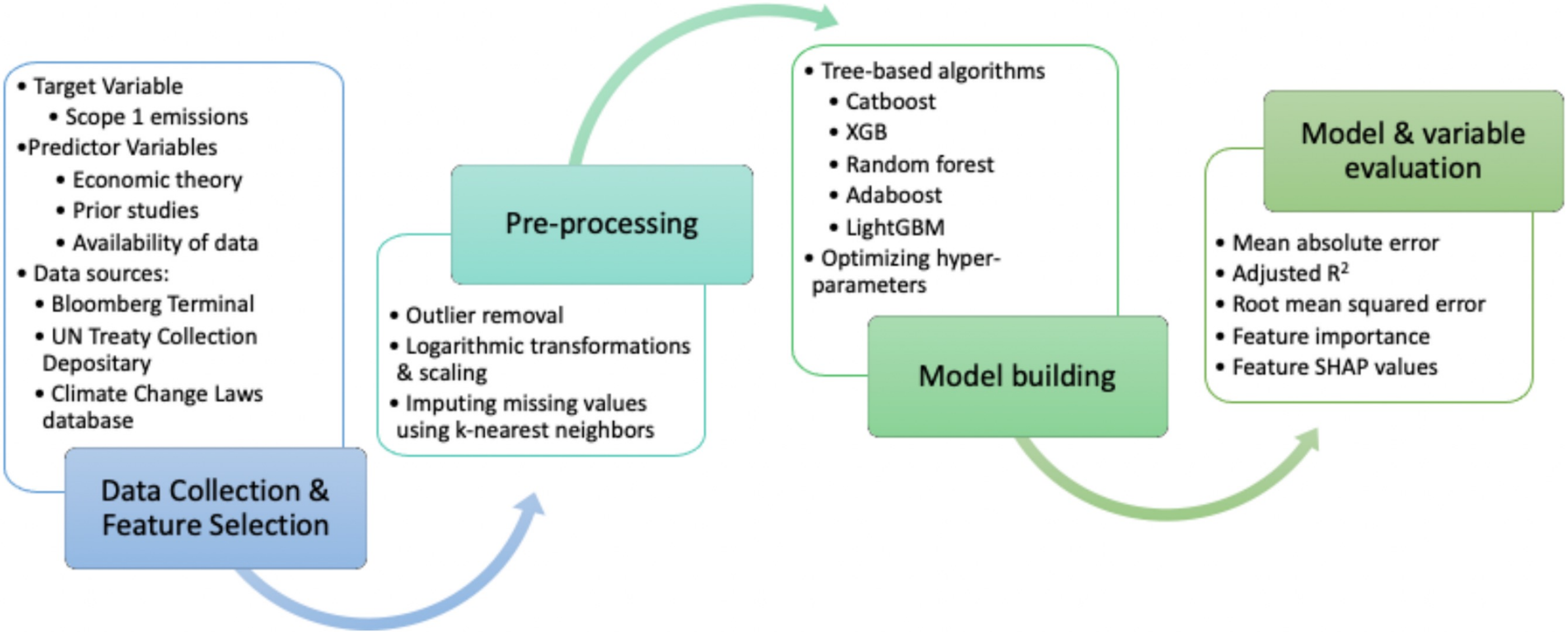


Source: Han et al. 2021. Estimation of Corporate Greenhouse Gas Emissions via Machine Learning. https://www.climatechange.ai/papers/icml2021/4

Common challenges:

- **Imbalanced datasets**
- **Missing features**
- **Mismatched feature patterns**
- **Need for uncertainty quantification**

Scope 1: Tree based ML methods



Source: Hadziosmanovic et al. 2021. Estimating Corporate Scope 1 Emissions Using Tree-Based Machine Learning Methods

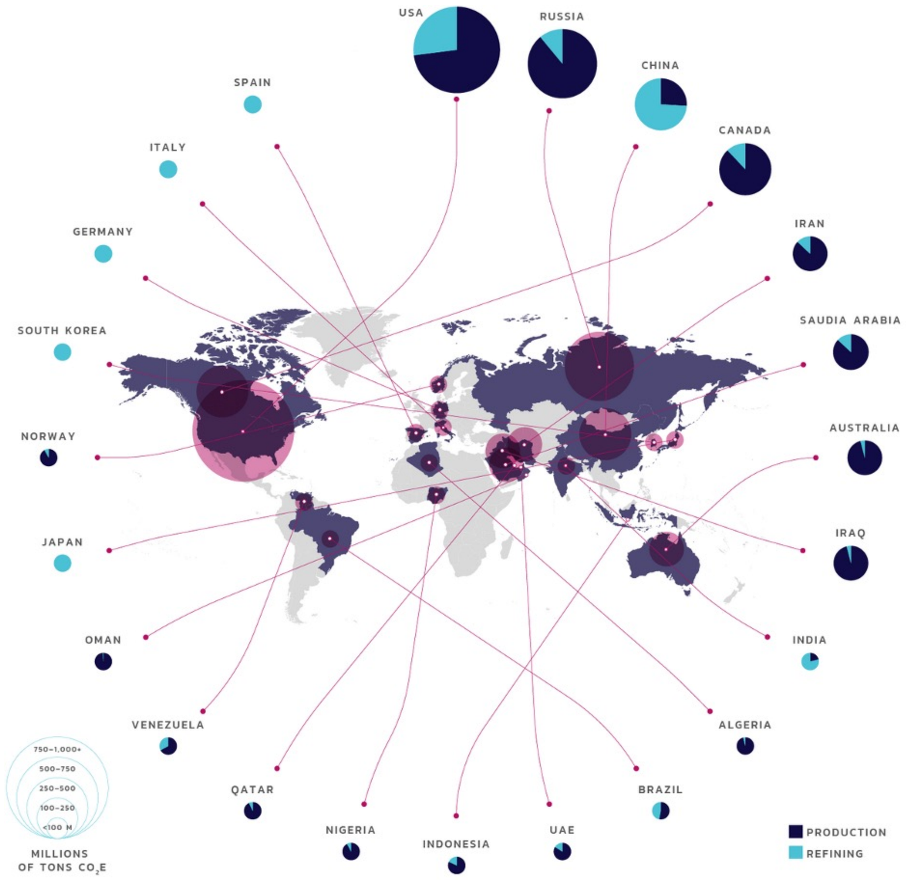
Climate TRACE

Climate TRACE aims to **track real-time greenhouse gas emissions** with high accuracy.

Leverages ML to analyze datasets from satellite data, **filling gaps in current emissions** data and providing near real-time updates.

Improves upon self-reported emissions, reducing risks of underestimation and inaccuracies.

Collaborates with various stakeholders, including nonprofits, tech companies, and academic institutions, to create an integrated and transparent emissions tracking system.



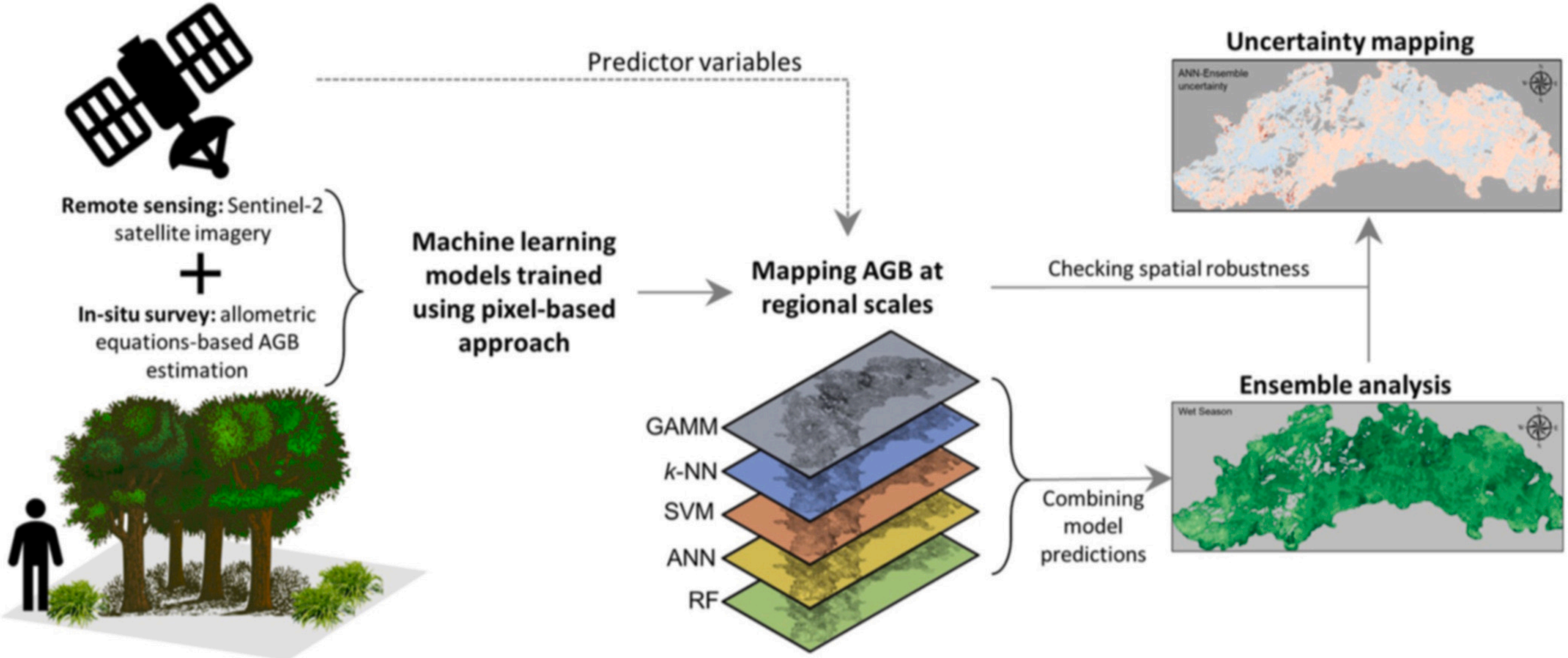
Source: https://datadrivenlab.org/climate/climate-trace-offers-foundations-for-live-emissions-tracking/

Part 3: Applications of machine learning for emissions monitoring

In this section you will learn about:

- Monitoring of nature-based carbon removals and offsets
- ML applications in monitoring biomass changes and deforestation (link to the CCAI lecture on MRV for forestry)

Biomass data products use ML



Source: Singh et al. 2022. Remote sensing-based biomass estimation of dry deciduous tropical forest using machine learning and ensemble analysis

Quick Poll

What are the sources of uncertainty in predicting biomass from space?

Figure source: GHG protocol. Figure 1.1.

Example: Biomass data products use ML to generate granular global datasets (1/2)

Ground sample distance (pixel size): 1m
Coverage: Global terrestrial surface.
Metric: Vegetation canopy height (m) for all vegetation >1m in height.

Source: Earth Blox dataset review: Meta Global Canopy Height (1m)
https://www.earthblox.io/blog/earth-blox-dataset-review-meta-global-canopy-height-1m 2/11

Example: Biomass data products use ML to generate granular global datasets (2/2)

Pros:

✅ It is the only tree-level canopy height model available for the whole globe.
✅ You can certainly see individual trees.
✅ Since it is canopy height (not biomass) you can use your own locally specific
allometric equations to convert height to biomass, and therefore be more
confident in the result.
✅ Underlying the method is very detailed global imagery (0.5m pixels), so at worst, it's a good tree/non-tree map (with the buildings removed).
✅ You can use it for tasks such as ecosystem fragmentation (for biodiversity scores) and estimating forest carbon stock

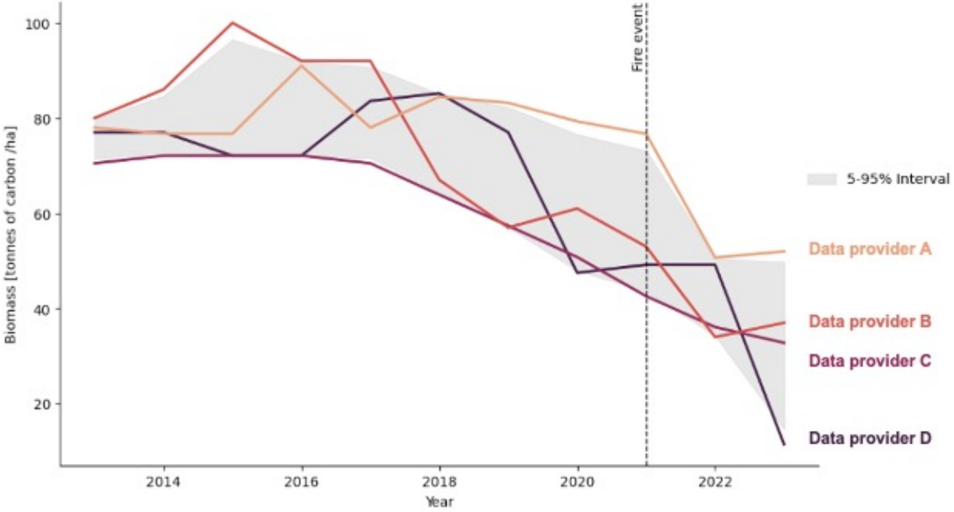
Cons:

❌ As with all global maps, it can have a good global performance but **may perform poorly in any specific location.**
❌ Areas with persistent cloud cover may have artefacts (expressed as areas of no canopy). Unhelpfully, these areas are not flagged in the data.
❌ Tiling artefacts means that you should corroborate this data with other
sources when looking at specific project areas.
❌ **Lack of temporal consistency** (training data over a short period)
❌ As with many similar products, this is just a snapshot, so can't be used to map change (e.g., not a time-series_

Source: Earth Blox dataset review: Meta Global Canopy Height (1m)
https://www.earthblox.io/blog/earth-blox-dataset-review-meta-global-canopy-height-1m 2/11

Biomass data products use ML, but there is still room for improvement…

A zoomed-in visual check of Klamath East fire damage. Imagery source: Google Earth Pro. The figure shows the difference between pre-fire and post-fire forest.



Different ways of assessing changes in aboveground carbon storage for the Klamath East project in California.

Datasets used (anonymized): Biomass index based on Sentinel-2 NDVI and EVI vegetation indices and SAR data, Kanop biomass data, Chloris biomass data, Planet biomass data.

Note: Data and analysis done by CarbonPool.

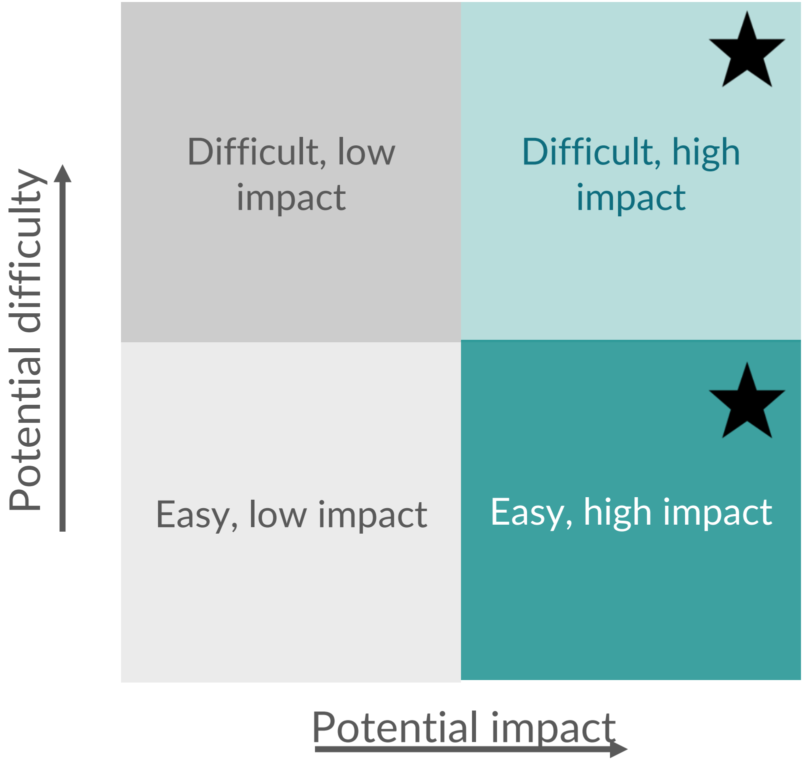
Part 4: Challenges and future directions

Opportunities and challenges of using ML for emissions accounting

- **What are the main benefits of the presented methods (and others)?**
- **What are the main drawbacks of these methods?**
- **Is there something missing that would need to be accounted for?**
- **Can we trust inferences of emissions using ML?**
- **What are the alternatives?**

Recommendations for selecting ML models to predict emissions

What are additional dimensions that need to be considered in evaluating the methods not shown here?



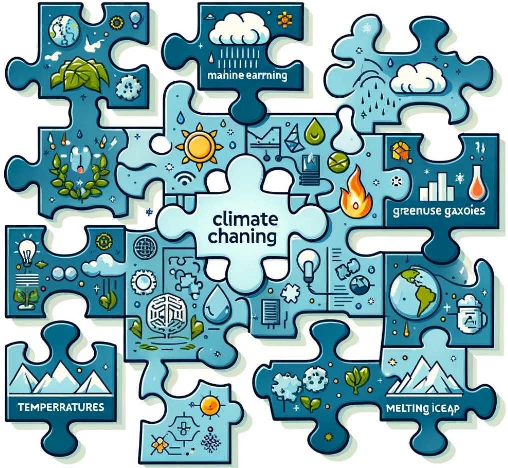
ML is one piece of the puzzle

ML is a powerful tool, **not a silver bullet**

Not necessarily relevant to every problem.

Where ML is relevant, collaboration is key:

- Targeting meaningful problems where ML adds value
- Avoiding oversimplification or overcomplication
- Recognizing potential risks



nature machine intelligence

Explore content | About the journal | Publish with us | Subscribe

nature > nature machine intelligence > perspectives > article

Perspective | Published: 17 August 2021

Towards neural Earth system modelling by integrating artificial intelligence in Earth system science

Christopher Irrgang, Niklas Boers, Maike Sonnewald, Elizabeth A. Barnes, Christopher Kadow, Joanna Staneva & Jan Saynisch-Wagner

Nature Machine Intelligence **3**, 667–674 (2021) | Cite this article

PHILOSOPHICAL TRANSACTIONS A

royalsocietypublishing.org/journal/rsta

Opportunities and challenges for machine learning in weather and climate modelling: hard, medium and soft AI

Matthew Chantry[1], Hannah Christensen[1], Peter Dueben[2] and Tim Palmer[1]

[1]Atmospheric, Oceanic and Planetary Physics, University of Oxford, Oxford, UK

[2]European Centre for Medium Range Weather Forecasts, Reading, UK

MC, 0000-0002-1132-0961; PD, 0000-0002-4610-3326

Introduction

Check for updates

Cite this article: Chantry M, Christensen H, Dueben P, Palmer T. 2021 Opportunities and challenges for machine learning in weather and climate modelling: hard, medium and soft AI. *Phil. Trans. R. Soc. A* **379**: 20200083. https://doi.org/10.1098/rsta.2020.0083

Accepted: 4 August 2020

Emission data landscape - resources

Where do I find emission data?

CDP database: https://www.cdp.net/en/data

GHG Protocol (Methodology): https://ghgprotocol.org/scope-3-calculation-guidance-2

Science based targets: https://sciencebasedtargets.org/

Interesting papers discussed in this section:

Serafeim, George, and Gladys Vélez Caicedo. "Machine Learning Models for Prediction of Scope 3 Carbon Emissions." Harvard Business School Working Paper, No. 22-080, June 2022. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4149874

Han et al. 2021. Estimation of Corporate Greenhouse Gas Emissions via Machine Learning https://www.climatechange.ai/papers/icml2021/4

Hadziosmanovic et al. 2021. Estimating Corporate Scope 1 Emissions Using Tree-Based Machine Learning Methods https://s3.us-east-1.amazonaws.com/climate-change-ai/papers/neurips2022/56/paper.pdf

And many more!

Q & A