Predicting global CO_2 fluxes using machine learning

Laia Amorós, Janne Hakkarainen, Iolanda Ialongo, Monika Szeląg
laia.amoros@fmi.fi
Finnish Meteorological Institute
26.5.2022

Introduction

- **Carbon dioxide** (CO_2) is one of main contributors to climate change, and most countries have agreed on reducing their CO_2 emissions in the coming years.
- Understanding atmospheric CO_2 fluxes, i.e. **emissions and sinks** of CO_2, is a fundamental problem in climate science that can help monitoring CO_2 global emissions.
- There exist several models to estimate global CO_2 fluxes. These are based on inverse modelling using ground-based CO_2 measurements, and more recently also satellite-based measurements. These also use wind fields, which are very time-consuming to compute.
- **We will use a ML approach to estimate global CO_2 fluxes using satellite data**.

CO_2 flux models: regression values

Global CO_2 fluxes can be estimated using inverse modelling.

- **CarbonTracker** CT2019b CO_2 fluxes, developed at NOAA Global Monitoring Laboratory, available from 2000 to 2018.
- Copernicus Atmosphere Monitoring Service (**CAMS**) CO_2 fluxes, available from 1979 to 2020.

These will be used as regression values.

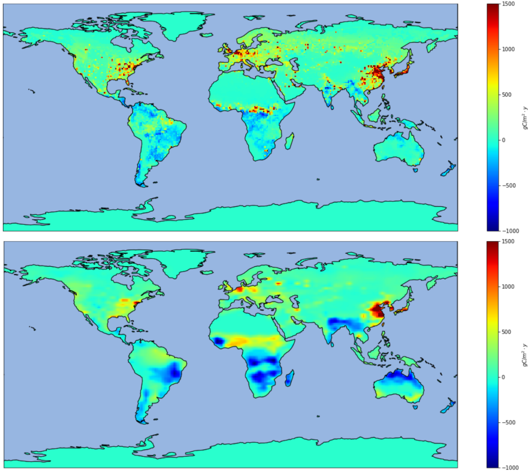


Figure: (1) CT2019b (2) CAMS surface, January 2015

Satellite data: features

We can use satellite measurements to track the amount of different gases present in the atmosphere. In particular we use the following monthly averages from 2015 to 2018:

- **XCO_2 anomalies** based on XCO_2 measurements [1] from OCO-2, with a spatial resolution of 1.29×2.25 km^2 across a narrow swath. Available since July 2014.
- **CO** from MOPITT, with a resolution of $1^\circ \times 1^\circ$. Available since March 2000.
- **NO_2** from OMI, with a resolution of $0.25^\circ \times 0.25^\circ$. Available since October 2004.
- **SIF** [2] from OCO-2, at a 1.29×2.25 km^2. Available since September 2014.

These are re-gridded at $1^\circ \times 1^\circ$ to be used as features, as well as hemisphere (north or south) and month.

[1] XCO_2 = space-borne column-averaged CO_2 dry air mole fraction.
[2] SIF = solar induced fluorescence.

Satellite data: features

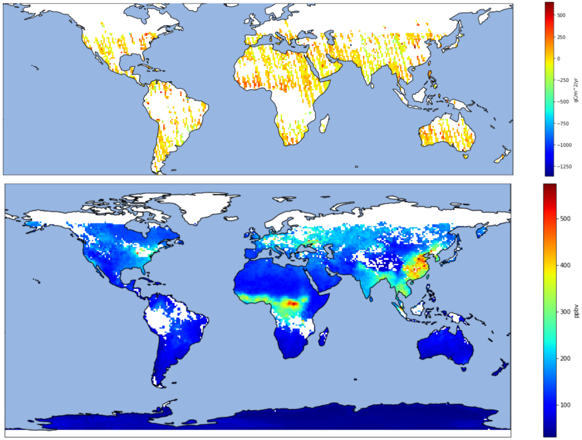


Figure: (1) XCO_2 anomalies and (2) CO, January 2015.

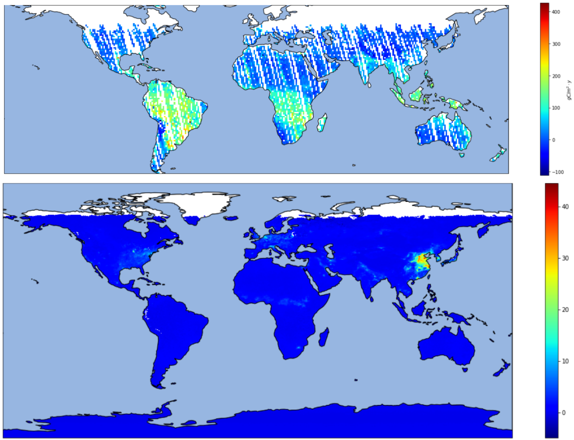


Figure: (3) SIF and (4) NO_2, January 2015.

Machine learning models and training

We consider 2 different machine learning (ML) models based on decision trees:

(1) **Random Forest algorithm**, parallel learning.

(2) **Extreme Gradient Boosting** algorithm (or XGBoost), iterative learning.

We differentiate CO_2 fluxes:

- For CarbonTracker: biospheric, fire, fossil and total CO_2 fluxes.
- For CAMS: biospheric (which includes fire), fossil and total CO_2 fluxes.

As **training set** we use the following data from 2015 to 2017:

as **features**: monthly averages of $\{CO_2, CO, NO_2, SIF\}$, hemisphere and month;

as **regression values**: CarbonTracker CT2019 fluxes or CAMS fluxes.

As **test set** we use the same data for 2018.

Predictions of our ML models: XGBoost

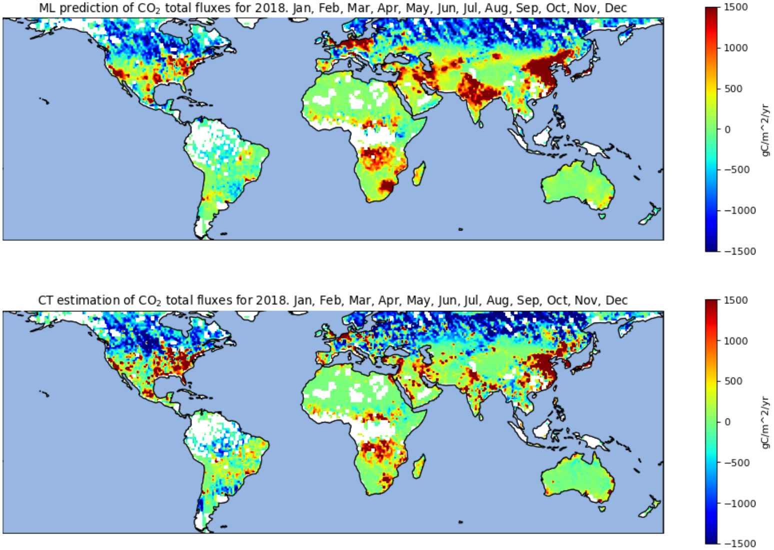
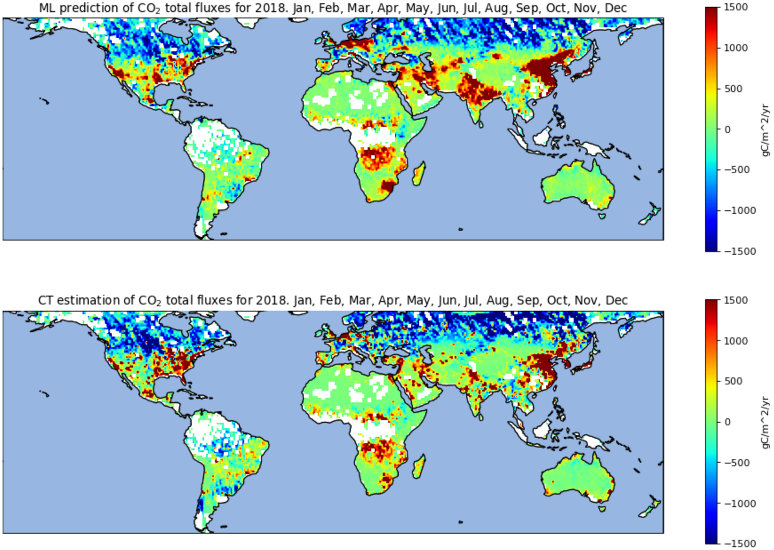


Figure: XGBoost vs CarbonTracker, 2018

Predictions of our ML models: Random forest



ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

Figure: Random forest vs CarbonTracker, 2018

Predictions of our ML models: XGBoost

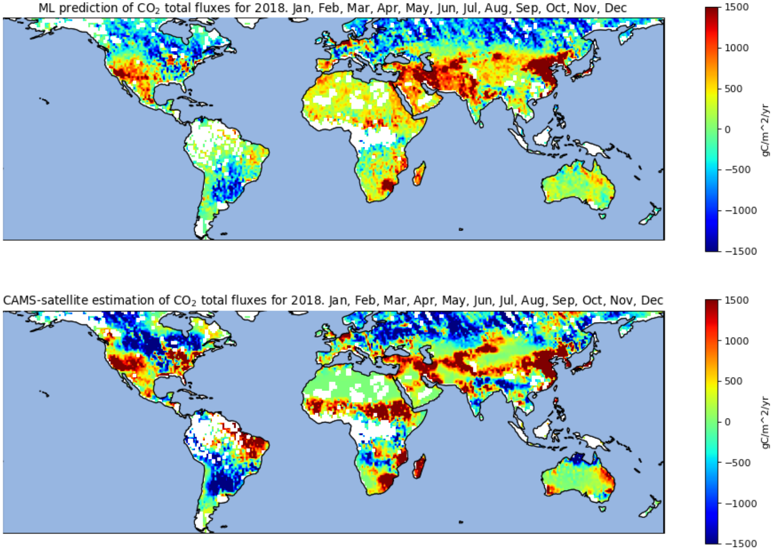


Figure: XGBoost vs CAMS, 2018

ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

Performance of our ML models

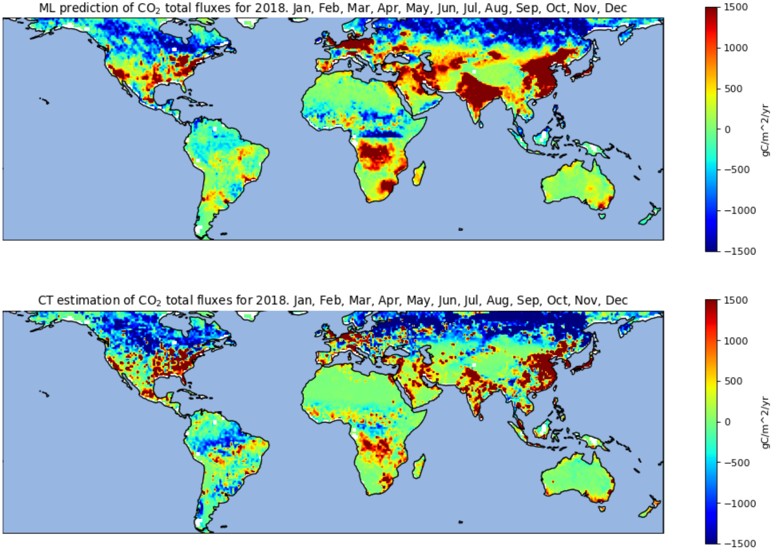
Features	Total fluxes			Fossil fluxes			Biospheric fluxes		
	CT	CAMS su	CAMS sa	CT	CAMS su	CAMS sa	CT	CAMS su	CAMS sa
NO_2	26.2	15.9	16.8	45.1	44.7	44.7	6	5.3	5.3
CO_2	2.4	2.9	8.5	4.9	2.9	2.9	1.7	2.7	8.7
CO	4.3	5.9	7.7	8.8	4.7	4.7	4.9	6.4	8.9
SIF	14.8	14.4	21.1	8.7	7.4	7.4	19.3	16.9	26.9
North	27.3	35.4	26.1	21.8	32	32	40.1	41.9	26.9
Month	23.5	23.1	16.5	8	6.1	6.1	26.2	24.2	19.4
Year	1.6	2.3	3.3	2.7	2.3	2.3	1.5	2.6	4
Lat	–	–	–	–	–	–	–	–	–
Lon	–	–	–	–	–	–	–	–	–
rmse	291.5	276.1	304.3	193.5	89.7	89.7	220.3	262.8	293.1
mae	153.4	187	210.8	65	37.6	37.6	117.5	177.4	202.5

Table 1: XGBoost model performance

Features	Total fluxes			Fossil fluxes			Biospheric fluxes		
	CT	CAMS su	CAMS sa	CT	CAMS su	CAMS sa	CT	CAMS su	CAMS sa
NO_2	40.2	27.2	26	54.7	65.7	65.7	16.8	16.1	16.5
CO_2	9.6	10.4	15.2	10.1	6.4	6.4	9.5	11.1	14.9
CO	12.4	15.2	7.4	13.3	9.5	9.5	14.2	16.8	18.7
SIF	19.6	22.4	25.6	12.7	10.1	10.1	30.9	26.7	31.8
North	3.6	6	3	2.4	3.7	3.7	6.9	7.6	2.8
Month	12.6	16	10.5	5	3.5	3.5	19.6	18.6	12.1
Year	2	2.7	3	1.8	1.2	1.2	2.2	3.1	3.2
Lat	–	–	–	–	–	–	–	–	–
Lon	–	–	–	–	–	–	–	–	–
rmse	297.8	281.7	315.1	196.3	90.9	90.9	226.6	267.5	299.3
mae	155.5	189.6	217.3	64.9	37.8	37.8	119.7	179.9	206.1

Table 2: Random forest model performance

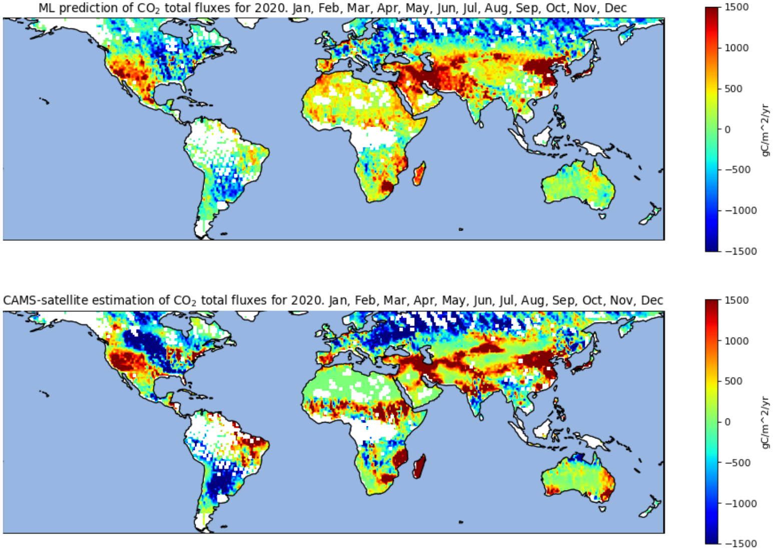
Predictions of our ML models: XGBoost without CO_2



ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

Figure: XGBoost vs CarbonTracker, 2018. Without CO_2 data

Predictions of our ML models: Covid effect



ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

Figure: XGBoost vs CAMS, 2018

Thanks! Questions?