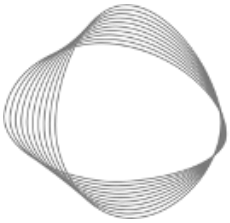
Power sector: Electricity Generation Emissions Methodology



Jeremy Freeman[1], Ali Rouzbeh Kargar[1], Isabella Söldner-Rembold[2], André Ferreira[2], Heather D. Couture[3], Jeyavinoth Jeyaratnam[1], Joseph O'Connor[2], Jordan Lewis[1], Madison Hobbs[1], Hannes König[1], Colin McCormick[1], Nick Amuchastegui[1], Tiffany Nakano[4], Charmaine Dalisay[4], Aaron Davitt[1], Lee Gans[1], Christy Lewis[1], Gabriela Volpato[1], Matthew Gray[2], and Gavin McCormick[1]

1) WattTime, 2) Transition Zero, 3) Pixel Scientia Labs, 4) Global Energy Monitor

1. Introduction

The energy sector is responsible for a significant portion of carbon dioxide emissions emitted, approximately 75% of total CO_2 emissions globally (Elavarassan et al., 2020). Electricity generation (power) represents a significant fraction of this amount, at times accounting for more than 40% of global GHG emissions (Nassar et al., 2017). With the passing of the Paris climate agreement, which sets goals to reduce CO_2 emissions and limit global warming, countries are required to provide reliable and timely information on their contribution to global GHG emissions (Kuhlman, et al., 2019; Elavarassan et al., 2020; Cusworth et al., 2021).

Currently, electricity generation emissions monitoring and reporting includes continuous emissions monitoring systems (CEMS) and bottom-up approaches (Liu et al., 2020). CEMS measures emissions at individual power plants and generally provides reliable and accurate emission measurements. However, CEMS are costly and deployed in limited regions globally, such as the United States (Cusworth et al., 2021). A more common approach to estimate GHG emissions is bottom-up self-reporting, which quantifies emissions at power plants using fuel consumption (activity data) and fuel quality and properties (emission factors (Vaughn et al., 2018; Liu et al., 2020). Yet, this approach tends to have uncertainties in fuel properties which leads to uncertainty in power plant emissions (Kuhlman, et al., 2019; Cusworth et al., 2021). Additionally, self-reported emissions tend to have varying levels of detail, time resolution, and reporting delay. Together, both approaches to monitor and report emissions create a fragmented system with uncertainty in tracking global individual power plant emissions and their contribution to global GHG emissions.

Currently, there is no comprehensive, third-party source of measured GHG emissions data for the global power sector. The Climate TRACE power sector aims to complement current monitoring and reporting approaches with emissions estimates derived from remote sensing. The power sector data includes power generation emission estimates from thermal power plants with fuel sources including coal, natural gas, fuel oil, diesel, biomass, and waste. We use a combination of existing

country- and region-level data, satellite data, and machine learning models to generate a comprehensive set of country-level annual emissions estimates, which we have published for the period 2015-2021, alongside asset-level annual emissions estimates for the 573 highest emitting plants in our dataset, which we published for the years 2019-2021.

2. Materials and Methods

2.1 Overview of Approach

We combined multiple complementary approaches to generate asset-level emissions estimates for a large number of fossil fuel power plants. For a given plant, we first estimated its total generation with a combination of the following approaches:

1. Satellite-derived estimates of plant generation using machine learning. This was possible only for certain classes of plants that we expect to display visible signals of generation in satellite images, described later in section 2.3.1.
2. Reported ground-truth generation data, described later in section 2.2.4. This was used only for plants in our machine learning (ML) training set. Note that while this data exists for a number of plants outside of our current training set, significant work is required to link it to our dataset and validate its completeness.
3. Region- and fuel-specific average capacity factors derived from country-level data to estimate generation for any remaining plants, described later in section 2.3.8.

Finally, we used region-, fuel-, and prime-mover-specific average carbon intensities to convert asset-level generation estimates to emissions estimates, and aggregate asset-level estimates to the country level. This process is explained in more detail in section 2.3.9.

The majority of this document will detail the methodology of approach 1, in which we used a satellite and machine learning pipeline, trained on plants with detailed ground truth generation data, to predict the generation and emissions of plants without such data. Our machine learning approach consists of two steps:

1. Our image- or sounding-level models predict the activity of a power plant from a single satellite image. We train separate models using PlanetScope, Sentinel-2A/B and Landsat-8 satellite imagery, matched to high-time-resolution generation data from individual power plants in the United States, Europe, and Australia.
2. Our generation ensemble model predicts the average 30-day capacity factor of a plant from the outputs of the above sounding-level models.

2.2 Datasets Employed

2.2.1 Global Fossil Power Plant Inventory

To accurately reflect emissions from the entire power sector, it is important to have a (mostly) complete, harmonized inventory of global power plants that are currently operational.
To use a plant in our machine learning modeling we required the following inventory data:

1. An accurate plant location for our satellite imagery
2. Attributes of the power plant including type, fuel, cooling technology, and air pollution control equipment to identify if the plant is suitable for our models
3. Capacity to identify the plant is of sufficient size to be modeled and to calculate the generation from the modeled capacity factor. Used in conjunction with unit operating dates to find plant capacity at any given date.
4. Fuel and prime mover type to estimate the emission factor.

For plants not suitable for machine learning modeling we required the capacity, fuel type, prime mover type and less accurate location data (province/country level).

Unfortunately, all existing power plant inventories have shortcomings, including missing, outdated, conflicting or incomplete information. Therefore, we developed our own harmonized global inventory of fossil fuel power plants. Each dataset contains different, complementary data that were merged together and standardized. Table 1 below describes how we use each of the datasets, including which data we republish.

Table 1 Datasets employed to create a harmonized Global Fossil Power Plant Inventory

Dataset Name	Plant Metadata Used	Published (use varying by asset/country)
US Energy Information Administration EIA-860, EIA-860m	Plant Name, Unit Fuel Type, Location, Unit Capacity, Unit Operating Dates, Unit Cooling type, Unit Pollution Control Tech SO_2	Asset Level Data, Aggregated into Country Level Data
World Resources Institute (WRI) Global Power Plant Database (GPPD)	Plant Name, Plant Fuel Type, Location, Plant Capacity, Plant Operating Dates	Asset Level Data, Aggregated into Country Level Data
S&P Global/Platts World Electric Power Plant (WEPP) database	Unit Fuel Type, Unit Capacity, Unit Operating Dates, Unit Cooling type, Unit Pollution Control Tech SO_2	The asset level dataset is proprietary and is used internally without being republished.
Global Energy Monitor (GEM) Global Coal Plant Tracker (GCPT) and Global Gas Plant Tracker (GGPT)	Plant Name, Unit Fuel Type, Location, Unit Capacity, Unit Operating Dates	Asset Level Data, Aggregated into Country Level Data
Other Sources (e.g., press releases, newspaper articles, company websites)	All	Asset Level Data, Aggregated into Country Level Data

Only the US Energy Information Administration (EIA) dataset provides all relevant data points for every plant, but only includes information for US-based plants. For the US, therefore, EIA data was primarily used. For the rest of the world, a combination of the other datasets was used.

To harmonize our datasets and get all the required information for every plant. unit and plant level mapping between datasets was performed. Global Energy Monitor provides unit and plant level mappings to Word Electric Power Plant (WEPP), while Global Power Plant Database (GPPD) contains plant level mappings to WEPP. For those plants missing linkages, internal work was done to match them.

GPPD, WEPP, and Global Energy Monitor have overlapping information such as the capacity of many plants. Plants with discrepancies for overlapping values were investigated and validated via primary sources such as newspaper articles, press announcements etc.

Additionally, some datasets were more up-to-date than others. Global Energy Monitor, for example, contains recently built plants not found in other datasets. Comparison and validation of the base datasets were done to ensure the most up-to-date plant information was included in our final dataset.

2.2.2 Plant Validation and Infrastructure Mapping

To validate and augment our plant-level data, OpenStreetMap (OSM), a free, public geographic database was used. Firstly, we manually cross-referenced and corrected the geolocation of power plants in our harmonized dataset. Secondly, OSM allows us to annotate ("tag") certain physical features of power plants (Figure 1). We used tags to label parts of the plant from which we expect to see vapor plumes, namely flue stacks and cooling towers. These annotations can be used to force our machine learning models to observe only the most pertinent parts of the plant, improving their performance. For every power plant on which we run machine learning training or inference, we completed the following manual tasks using OSM:

1. Confirmed that there is a power plant at the provided coordinates and that it is the correct plant by checking plant information and visible technology (e.g., cooling equipment, coal piles etc.) on the ground matches our information about the plant;
2. Annotated all flue stacks (chimneys); and
3. Annotated all cooling towers, mechanical cooling equipment and air cooling equipment.

We used in-house annotations for specialized tags that are not relevant for OSM, including labeling of flue stacks with flue gas desulfurization technology. More information on our activities on OSM can be found on the Climate TRACE OSM wiki page (https://wiki.openstreetmap.org/wiki/Organised_Editing/Activities/Climate_TRACE).

Figure 1 White Bluff power station as shown on OpenStreetMap (top) and in OpenStreetMap edit mode (bottom) showing how aerial imagery was used to annotate locations of flue stacks (transparent white circle) and cooling towers (red circles).

2.2.3 Weather data

Our machine learning models were trained to observe visible vapor plumes to predict power plant activity. However, it was observed that visible vapor plume formation was reduced at high ambient temperature and low relative humidity. In particular, our models based on signals from flue gas desulfurization performed poorly at high temperatures and low humidity. For example, above 25 degrees Celsius, on/off classification performance drops to the level of random guessing. We filtered such examples out of training and inference, as described in Section 2.3.2. For this reason,

historical weather data from 2015-2021 was scraped for all of our plants from World Weather Online, available at https://www.worldweatheronline.com**.**

2.2.4 Plant-Level Electricity Generation Data

To train our machine learning models, we used multiple sources of ground truth for high time-resolution (hourly to sub-hourly) plant-level generation in MWh for plants in regions where this was available. While many datasets were available that provide low time-resolution generation data - days to months - or generation aggregated across a large number of power plants, these were not helpful for use in our machine learning model training set. Our ground truth datasets include the US EPA Air Markets Program Data (AMPD), European Network of Transmission System Operators for Electricity (ENTSO-E), and Australia National Electricity Market for the US, Europe, and Australia, respectively. These datasets provide us with actual generation at hourly or sub-hourly intervals at several thousand power plants, from 2015 to the present.

Considerable work was required before using these data sources in our models. Similarly, to what's described in 2.2.1, each asset time series included in the various power generation datasets must be matched to the power plants in our database. It is crucial that this work be done carefully to ensure that the generation data that will be treated as "ground truth" is complete and not mis-attributed to the wrong plant. Mappings between ground truth plants and the power plant inventory have resulted in plants with ground truth data in 23 countries.

Then, in power plants that have clearly visible activity-related signals, the power generation values can be validated, whether through hand-labeling studies or by inspecting samples where our models have particularly confident errors. It's important to be careful with discarding false negatives, i.e., cases where the ground truth data claims that the power plant is active, but the models think that it's off, as these might represent a faint signal or an otherwise difficult context that's beyond the limits of the modeling approach. When it's a false positive, e.g., ground truth data points to no power generation but there are obvious plumes coming out of the cooling towers and flue stacks, it's more likely to be an actual mislabeling issue. Thus, whenever a ground truth data point has a matching asset in our database and its value adds up to what is observed in the satellite images, that sample is considered valid and used in the model training and outputs. Otherwise, it's ignored and taken note of in an exclusions database.

Beyond the role as targets for training models, this high resolution ground truth data is directly assigned to the yearly power generation of a power plant, in the cases when it's available and has been validated. In the Climate TRACE asset data reporting schema, assets with ground truth capacity factors are denoted as "t" in the "other1" field.

Various additional datasets are used for country-level estimates of generation, carbon intensity, and for baseline generation data. These include the Environmental Investigation Agency, Ember

Electricity Generation Data and IEA Electricity Generation Data, all providing global information.

2.2.5 Satellite Data and Processing

Remote sensing imagery from the PlanetScope constellation, Sentinel-2A/B and Landsat-8 satellites were employed in a machine learning modeling approach to infer a power plant's operational status through the identification of emitted water vapor and smoke plumes. Below, a description of each satellite and imagery processing steps are provided.

Planet Lab's PlanetScope satellite constellation consists of approximately 130 individual satellites, called "Doves", with the first launch of this constellation in 2014 (Planet, 2022). The PlanetScope constellation provides daily revisits with an equator crossing time between 7:30 – 11:30am (Planet, 2022). Each PlanetScope satellite images the earth's surface in the blue, green, red and near-infrared (NIR) wavelengths (~450 nm – ~880 nm), with the exception of the "SuperDove" instrument which includes additional wavelengths (Dos Reis et al., 2020; Moon et al., 2022). The PlanetScope Analytic Radiance (TOAR) PSScene4Band was downloaded via Planet Labs API, providing a spatial resolution of ~3m for power plant plume identification (Figure 2).

The European Space Agency's (ESA) Sentinel-2 mission comprises two satellites- Sentinel-2A, launched in 2015, and Sentinel-2B, launched in 2017 (Main-Knorn et al., 2017). Each Sentinel-2 satellite has a 10-day revisit time with a 5-day combined revisit and an equatorial crossing time of 10:30am (Shikwambana et al., 2019). Both satellites are equipped with a multispectral (MSI) instrument which provides 13 spectral band measurements, blue to shortwave infrared (SWIR) wavelengths (~442 nm to ~2202 nm) reflected radiance and, depending on the band, provides measurements at 10m to 60m spatial resolution (Main-Knorn et al., 2017; Shikwambana et al., 2019; Figure 2). To identify power plant plumes, Sentinel-2A/B Level-1C Top of Atmosphere (TOA) products are freely available and were downloaded via Google Earth Engine (GEE). More information on the Sentinel-2 mission and data products can be found at the ESA Sentinel-2 mission website (sentinel.esa.int/web/sentinel/missions/sentinel-2).

The Landsat-8 mission is jointly managed by NASA and U.S. Geological Survey (USGS; Marchese et al., 2019). Landsat-8 was launched in 2013 and has a 16 day revisit with an equatorial crossing time of 10am (+/- 15 minutes). Landsat-8 is equipped with two instruments- Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS). Together, these instruments provide 11 spectral band measurements, blue to thermal infrared (TIR) wavelengths (~430 nm to ~1250 nm) and, depending on the band, provide 30m and 100m spatial resolutions (Mia et al., 2017; Figure 2). To identify power plant plumes, the freely available Landsat-8 collection 1, Tier 1 TOA was downloaded from GEE. More information on the Landsat-8 mission and data products can be found at the USGS website (https://www.usgs.gov/landsat-missions/landsat-8).

Figure 2 The Ninghai power station as seen from satellite Landsat 8, 30m spatial resolution (left), Sentinel 2, 10m spatial resolution (center) and PlanetScope, 3m spatial resolution (right). These images only represent the visible bands (red, green and blue).

For all satellite datasets, a region of interest (ROI) was produced for each power plant by setting an outer boundary that envelops the plant itself, all associated facilities, and any other affected areas of interest. Filters were applied to ensure image quality and low cloud cover. With these parameters, 1.2 million images captured in years 2015-2021 were downloaded for more than 3600 plants in 127 countries.

2.3 Modeling Pipeline

In order to identify power plant activity, a machine learning modeling pipeline was built to perform three tasks using satellite imagery:

1. To estimate if a plant is running or not (on/off), given a satellite image of that plant at a certain point in time (sounding-level classification).
2. To estimate to what extent that plant was being utilized, given a satellite image of that plant at a certain point in time (sounding-level regression).
3. To aggregate the predictions from models (1) and (2) into integrated estimates of utilization over the last 30 days (generation model).

2.3.1 Plant Selection

Our sounding-level models were predicated on the existence of signals indicating the activity of a plant in a satellite image. The technology used at a power plant dictates the type and strength of signal available to our models. At the time of this work, we were able to model signals from two type of plants:

1. Plants using natural draft towers (NDT) for cooling. These have a large structure that allows vapor plumes formed during cooling to be clearly seen.
2. Plants using wet flue gas desulfurization (FGD). After desulfurization, flue gases become saturated with water, increasing the visibility of plumes from the flue stack.

Separate models were built for these two sets of plants. We applied further criteria to these plants for modeling:

1. Require at least 80% of units by capacity to have the specified technology (FGD or NDT) and that this is true for all dates covered by the training set (2016 to present). Training only; plants with any FGD or NDT capacity were included in inference.
2. Require that at least one NDT tower or FGD-enabled flue stack has been annotated in OpenStreetMap or in our in-house annotations database.
3. Exclude plants with generating capacity less than 500 MW for training; inference is run for plants with capacity as low as 50 MW.
4. Exclude plants with known data quality issues.

2.3.2 Satellite Image Selection

The following satellite images were excluded from our models:

1. `Images with substantial cloud cover (≥ 20% of the image)`. We only exclude images with relatively high cloud cover to avoid falsely excluding images containing large plumes, which are easily misclassified as clouds.
2. Images in which our flue stack and cooling tower annotations are not fully captured in the satellite image.
3. Images with known data quality issues.

In addition, for FGD models only, images were excluded in which ambient weather conditions were unfavorable for plume visibility (see section 2.2.3). At high temperatures and/or low relative humidity, the water vapor in the flue stack does not readily condense, plume visibility is reduced, and our models perform poorly. For these models we used an empirically-derived cutoff rule for plume visibility, illustrated in Figure 3:

1. Exclude images in which the ambient temperature is greater than 25 degrees Celsius
2. Subtract the ambient temperature, in Celsius, from the relative humidity percentage. If this value is smaller than 15%, exclude the image.

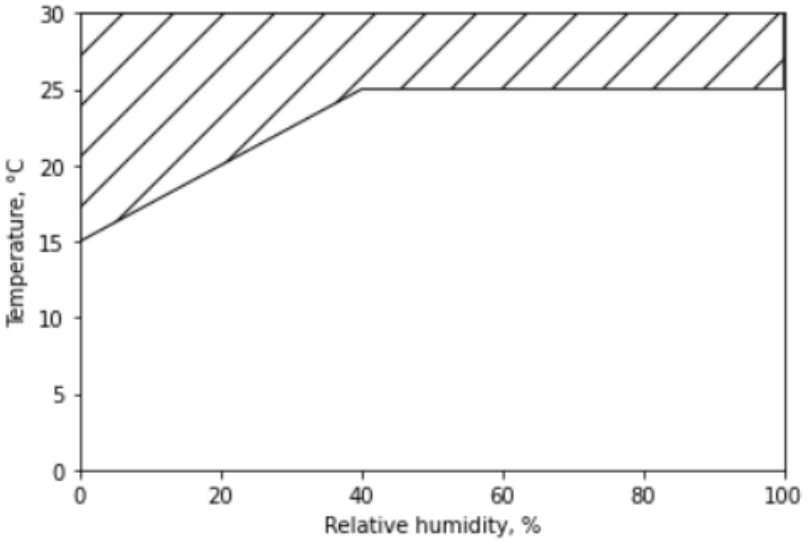


Figure 3 Visual representation of the empirically-derived weather-based cuts used in our FGD models. The top hatched region is excluded, while the below region is included. At high temperature and low humidity, vapor plume visibility is reduced.

2.3.3 Ground Truth Labels for Model Training

At training time, we required a dataset of satellite images that are linked to plant-level generation data. We used satellite image timestamps to match each image to the nearest record of plant-level generation data, described in Section 2.2.4. For regression models, we labeled each image with the "capacity factor" in that image – the generation of the plant divided by its capacity at the given timestamp. For classification, we label plants with > 5% capacity factor as "on" and everything else as "off". This nonzero threshold is a rule of thumb that we have found to be useful, since we see a handful of plants reporting very low levels of generation that can functionally be considered "off".

2.3.4 Sounding-Level Models and Features

To estimate power plant operating status and capacity factor from satellite imagery, our models were split into two main classes.

1. Gradient-boosted decision trees using the XGBoost library.
2. Multi-instance convolutional neural networks (CNNs) using the TensorFlow framework.

Figure 4 illustrates the structure of our XGBoost and multi-instance CNN models.

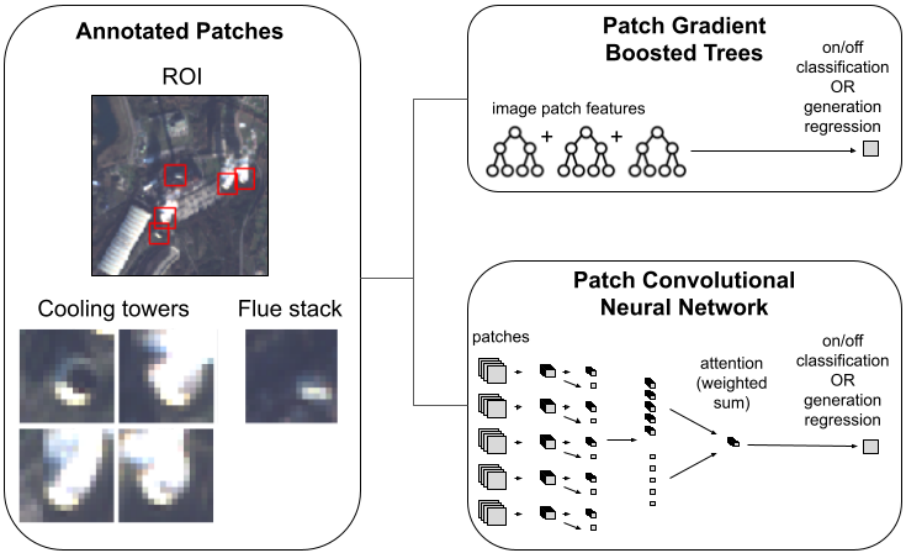


Figure 4 Diagram of the machine learning models used to estimate power plant operating status and capacity factor from satellite imagery. It showcases how relevant parts of power plants are cropped from satellite imagery, in this case forming patches of cooling towers and flue stacks, which were then used by machine learning models to classify the power plant's operating status as on or off and/or to regress on the capacity factor (i.e., what fraction of the power plant's potential is being used to generate power). These models can either be based on convolutional neural networks (CNNs), which combine the patches through an attention layer, or gradient boosted trees, that aggregate statistical features of the images of each infrastructure type. Image taken from Couture et al. (2020a).

XGBoost models require engineered features. We utilized our augmented OSM dataset of flue stack and cooling tower locations to cut out small patches around each location. We then calculated the mean activation of each satellite band for each patch, along with other statistics. This resulted in a vector of statistics of length equal to the number of patches in the image. We then aggregate this vector to the image level, with mean, min and max operations.

The multi-instance CNNs utilizes the annotated plant features directly. Here, we passed the model a vector of image patches, which are processed by a shared convolutional layer before being aggregated using an attention layer into a single feature vector before the final dense layers.

Each of these approaches can be equally applied to our classification and regression problem, and to NDT or FGD models. In practice, we build at least one of each of these models for all distinct combinations of problem class (classification, regression), and plant type (NDT, FGD). We then independently tune the hyperparameters of each of these models. Model evaluation is performed using plant-wise cross-validation, described later in Section 2.3.6.

2.3.5 Generation Ensemble Model

The sounding-level models described in Section 2.3.4 give us an instantaneous estimate of the power plant activity in a single image. Collecting these instantaneous estimates creates an irregular time series of classification and regression estimates for each plant (Figure 5). In order to estimate emissions of a plant over a given period of time, we built a second-stage model responsible for aggregating these sounding-model time series into features and predicting a rolling 30-day capacity factor for each plant. This capacity factor can then be multiplied by the recorded capacity of the plant and an emissions factor, as described later in Section 2.3.9, to estimate emissions.

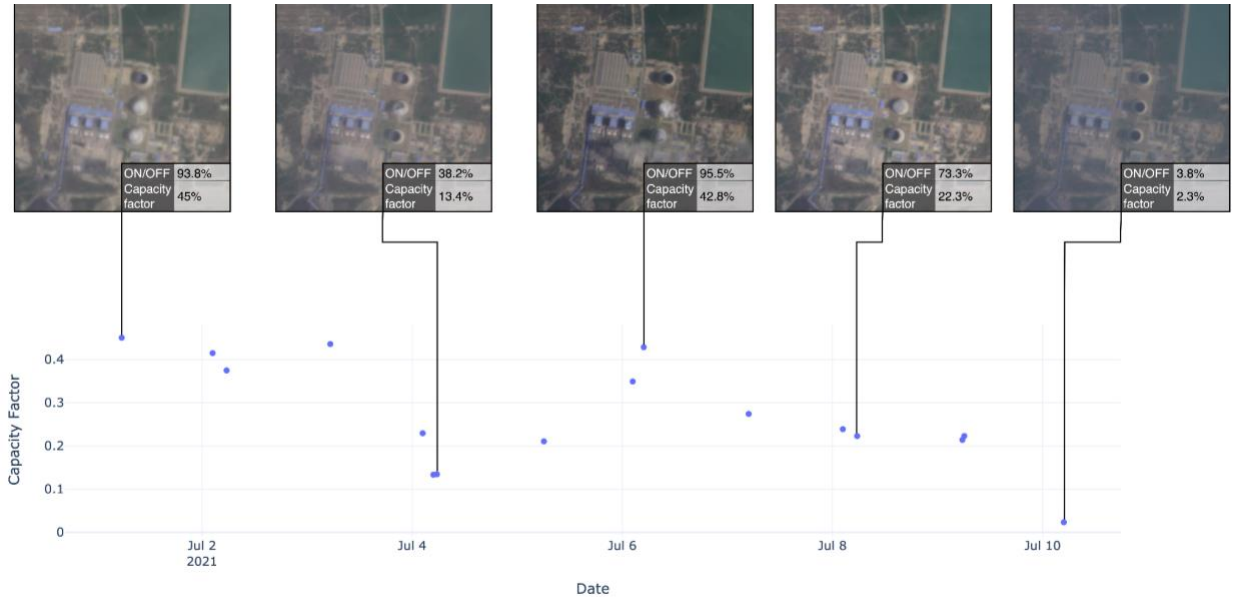


Figure 5 Depiction of the capacity factor time series, alongside on/off classification, for the power plant Talwandi Sabo Power Project in India, using PlanetScope satellite imagery and a CNN-based model.

A simple approach to estimate the total 30-day capacity factor of a plant would be to average its sounding-level capacity factor regression predictions over the last 30 days. We instead used a set of features based on this approach and trained a regularized linear model to combine these into a final 30-day capacity-factor prediction. We train two of these linear models, one for FGD plants and one for NDT plants.

To calculate features for each plant at each point in time, we calculated the min, mean and max 30-day prediction for each sounding-level model, resulting in three features per model type and satellite. Then, in addition to our 30-day feature windows, we added 60-day and 180-day windows. These longer windows allowed our models to observe the longer-term behavior of the plant when making a prediction. This resulted in a set of 27 features for our FGD generation model and 42 for our NDT model.

Finally, we trained an L1-regularized linear model to predict 30-day utilization from these features. Since this model was trained on the outputs of another machine learning model, we must take extra care to avoid data leakage. We trained the model using out-of-sample cross-validation predictions from the sounding level models and evaluated it using the same set of cross-validation folds. Figure 6 shows a sample of cross-validation predictions for a single plant from our NDT generation model.

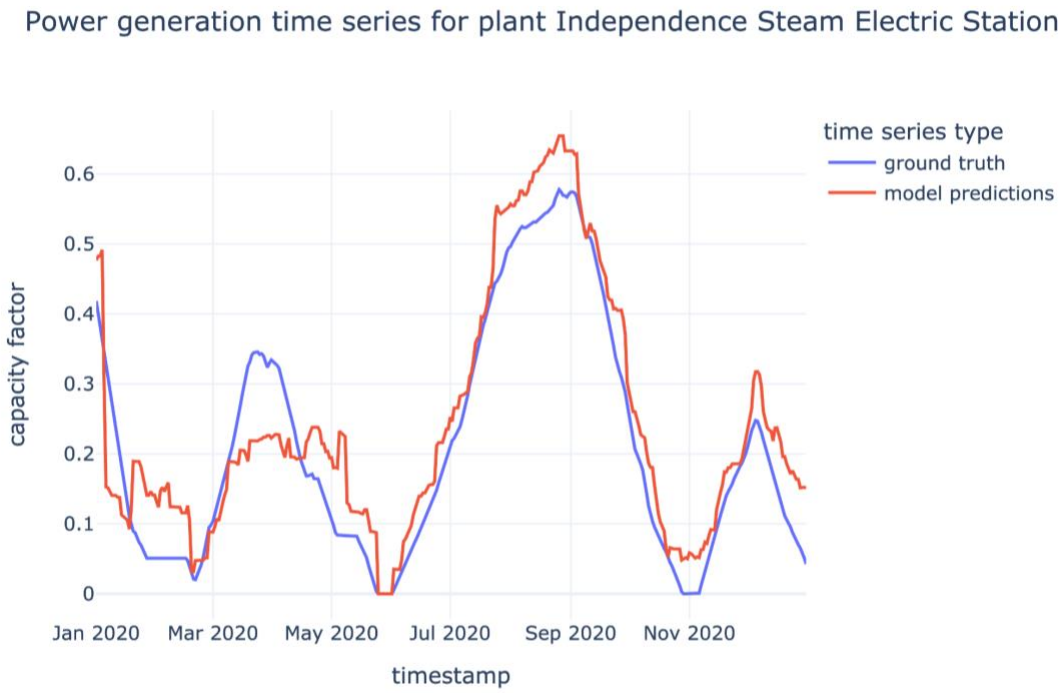


Figure 6 Example of predicted and true 30-day rolling capacity factor time series for Independence Steam Electric Station, a natural-draft-cooled coal-fired power plant. The predictions were made out-of-sample in our cross-validation loop.

2.3.6 Cross-Validation

Since we aimed to estimate the emissions of power plants for which no reported generation data exists, it was essential to validate that our models generalize to plants that were not present during training. In order to do this, we ensured that all of the images of a single plant were contained within the same fold, i.e., a plant cannot have images in both the training set and the test set.

This approach can continue to allow information leakage between folds if not performed carefully. Since multiple power plants are often located very close to one another, it is possible that the plumes from Plant A, in the training set, were visible in images of Plant B, in the test set and vice-versa. To reduce the risk of this leakage, we built a regular 1x1 grid in latitude-longitude space and kept plants in the same grid cell together during cross-validation. This process is illustrated in Figure 7 for the case of four-fold cross-validation, which we used throughout the pipeline. This

goes some way to avoiding “local overfitting”, where plants that are close to one another behave similarly, leading to inflated cross-validation scores relative to our true target plants.

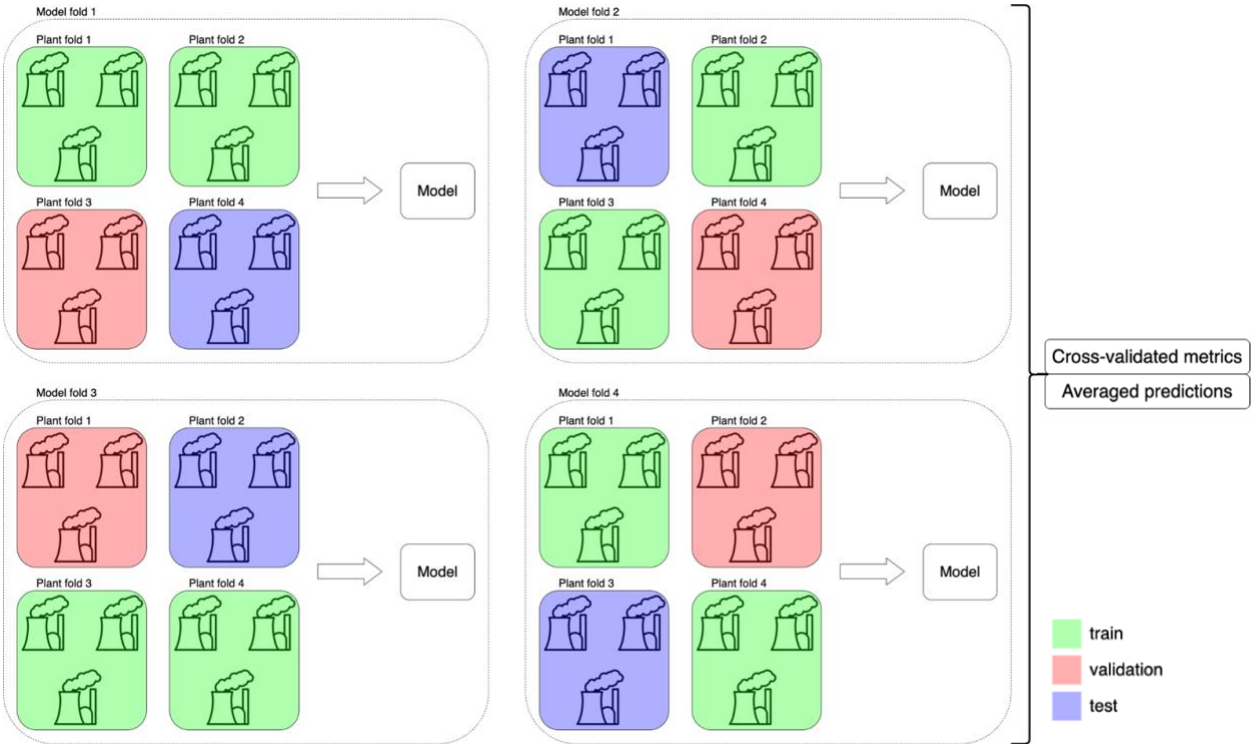


Figure 7 Diagram illustrating the construction and rotation of plant folds in our cross-validation loop.

2.3.7 Model Inference

To produce our final estimates for plants outside of the training set, we followed the same data processing steps laid out in the previous sections, but with a relaxed cooling type filter. Meaning, not all power plants employ one type of cooling technology. By relaxing the cooling type filter, we allow plants with less than 80% cooling type purity through the models, e.g., a power plant for which only 79% of its capacity uses NDT for cooling, and combine outputs where multiple models are applicable.

Predictions for all images of inference plants were generated over the period 2015-2021, using our trained models described in section 2.3.4. We then used these predictions to generate time series features and 30-day capacity factor predictions using trained models as described in section 2.3.5. The result is a set of predictions, one per day per plant, of the capacity factor of that plant over the preceding 30 days. We sampled these appropriately to produce estimates of the annual plant-level capacity factor.

2.3.8 Country-Level Capacity Factor Model

In addition to our machine learning models, we produced a second set of simpler capacity factor estimates applicable to all of the plants in our dataset, regardless of type. We used EIA and EMBER country-level annual estimates of capacity and generation by fuel type to calculate the annual average fuel-specific capacity factor in each year reported for each country in the world. We then assumed an even distribution of capacity factor within each country.

These estimates were used for any plants where we do not have a capacity factor estimate from the machine learning models described above. In inference countries, with average capacity factors significantly different from the average of the training set, our ML models make systematic errors. We averaged our satellite-derived estimates with the simple estimates described above to improve their calibration.

2.3.9 Emissions Estimation

To create the final plant-level emissions estimates, the estimated annual power generation of each plant was multiplied by an appropriate emissions factor. A power plant consists of one or more generating units, each of which may have a different fuel source and prime mover type, and therefore a different carbon intensity, from other units at the plant.

For each unit, an emissions factor was calculated through a combination of region-, fuel-, and prime-mover-specific average carbon intensities. Carbon intensity values for combinations of energy source and prime mover technology were obtained from WattTime's internal database, which combines USA EPA and JRC data, and country corrections from IEA. The emissions factor of each unit was determined as follows:

1. A base value was gathered from the combination of the unit's energy source (e.g., coal, gas, oil, etc.) and prime mover technology (e.g., combined cycle, simple cycle, etc.). This factor accounts for the typical efficiency differences between fuel and prime mover types.
2. If the combination of the energy source and prime mover did not have a value in the database, it was assigned the average carbon intensity of the energy source.
3. The final emissions factor was calculated by applying a country correction, a scalar that was multiplied to the base value to account for regional differences in power plant efficiency (due to age, technology level, size), fuel quality, and the impact of ambient conditions on carbon intensity.

Next, we converted our plant-level capacity factor estimates to the unit level, by assuming that each unit at the plant ran with the same capacity factor predicted from inference or the appropriate country fuel-specific average capacity factor where inference was not available. We then calculated the total annual emissions at the unit level before aggregating these up to the plant level to generate our final plant-level emissions estimates.

Finally, the unit-level estimates were also aggregated to the country level to provide the annual country level total electric power sector emissions estimates.

2.3.10 Country Level Emissions and Selecting Assets for Publication

We published annual country level emissions predictions for 227 countries for the years 2015-2021. Country level estimates are aggregated unit-level emissions.

We selected the fossil fuel plants to publish by gathering and validating the top 500 emitters for each of the years between 2019 and 2021, resulting in 565 plants where we report asset-level emissions. Of these plants, 48 plants have ground truth generation derived estimates, 218 plants have machine learning-derived emissions estimates, and 299 use capacity factors based on disaggregation of public country and region-level data. In addition to emissions estimates, we published a subset of inventory data relating to each plant, namely country, fuel type, latitude and longitude, capacity, and ownership. In the Climate TRACE asset data reporting schema, assets with remote sensing based inference capacity factors are denoted as "i" in the "other1" field, assets in the training set with ground truth based capacity factors are denotes as "t" in the "other1" field, and assets using only country-level capacity factors are denoted as "a" in the "other1" field.

3. Results

3.1 Model Performance

Table 2 summarizes the cross-validation performance of our best sounding-level models. For classification models, we report the classification accuracy; for regression models, we report the root mean square error (RMSE) on the capacity factor; both metrics are calculated on the validation set. To contextualize these numbers, we include "baseline" results using a simple benchmark model. In the case of classification, this benchmark predicts that all plants are always "on". For regression, the benchmark predicts a constant capacity factor equal to the mean capacity factor observed in the relevant training set, which varies from around 0.4 to 0.5.

Table 2 Summary of sounding-level model cross validation performance for each satellite and plant type. Baseline results are italicized. NDT = natural draft towers and FGD = wet flue gas desulfurization.

Sounding models			
Satellite sensor	**Plant filter**	**Classification accuracy** (*and baseline)*	**Regression RMSE** (*and baseline*)
PlanetScope	Coal NDT	**99.74%** (*78.15%*)	**0.1707** (*0.3303*)
PlanetScope	Coal FGD	96.94% (*81.35%*)	0.2234 (*0.3361*)
Sentinel 2	Coal NDT	99.39% (*74.93%*)	0.1993 (*0.3375*)
Sentinel 2	Coal FGD	93.44% (*79.20%*)	0.2358 (*0.3451*)
Landsat 8	Coal NDT	97.34% (*79.80%*)	0.2286 (*0.3376*)
Landsat 8	Coal FGD	94.21% (*82.59%*)	0.2533 (*0.3397*)

We outperform the simple baseline model significantly in all cases. We observe better performance in models trained against higher-resolution, larger datasets, as expected. We also see better performance in NDT models, compared to FGD models. NDT plumes are almost always large and clearly visible by eye in satellite images, while FGD plumes can be harder to spot, since they originate from a smaller aperture and appear to disperse more readily, especially in unfavorable weather conditions.

Table 3 summarizes the results of our two generation ensemble models. Here we measure performance using the RMSE between the true and predicted 30-day capacity factors in the validation set. The baseline used is again the average value of the target in the relevant training set – 0.42 for NDT plants and 0.48 for FGD plants. The 30-day averaging of the target has a smoothing effect, leading to a lower baseline RMSE compared to the sounding-level regression tasks. Our generation models combine multiple models, satellites, and timescales to outperform this baseline significantly.

Table 3 Summary of generation ensemble model cross validation performance for each plant type. Baseline results are italicized. NDT = natural draft towers and FGD = wet flue gas desulfurization.

Generation models	
Plant filter	**Regression RMSE** (*and baseline*)
Coal NDT	0.1356 (*0.2690*)
Coal FGD	0.1940 (*0.2657*)

Here, we again observe better performance on NDT plants than FGD plants, though both comfortably improve on the baseline. This is explained both by the better performance of the NDT sounding-level models, as well as the reduced number of sounding-level predictions going into the FGD models due to the weather cuts described in section 2.3.2, which lead to periods with very little sounding-level information for some plants.

3.2 Global and Country-Level Emissions Estimates

Table 4 gives our total estimates of global emissions electricity generation for the years 2015-2021. Global emissions have steadily increased from 2015 to 2021. Climate TRACE emissions estimates do decrease in 2020, during the COVID-19 pandemic, which rebound in 2021. Similar results have been reported in Davis et al. (2022).

Table 4 Total global emissions estimates for electricity generation for the years 2015-2021.

Year	Estimated Emissions (Gt CO_2)
2015	12.37
2016	12.59
2017	12.73
2018	13.13
2019	13.02
2020	12.49
2021	13.31

Table 5 provides a sample of our country level emissions estimates by providing the top five most-emitting countries. This table includes the percent of emissions derived from satellite data. China

is the largest emitter for 2021, about three times higher than the USA and ~11.5 times greater than Japan. Emissions data for all 227 countries for 2015-2021 can be downloaded in full from the Climate TRACE website.

Table 5 The five largest emissions estimates by country for the year 2021.

Country	2021 Estimated Emissions (Mt CO_2)	% of Capacity Satellite-Derived
China	4816	58%
USA	1584	12% (mostly ground truth)
India	1196	29%
Russia	553	26%
Japan	421	14%

3.3 Asset level Results for Top Emitters

Table 6 displays a sample of our asset level emission estimates, showing the top five emitters for 2021. The top five emitters are located in Asia, specifically Taiwan, Korea, China, and Russia. The 565 power plants with the highest emissions in 2019-2021 can be downloaded in full from the Climate TRACE website.

Table 6 The five plants with the largest emissions in 2021.

Asset Name	Country	Emissions Estimate Method	Fuel type	Emissions ($MtCO_2$)
Taichung power station	Taiwan	Country + Fuel Type Imputation	coal, oil	34.190
Taean power station	Korea	Country + Fuel Type Imputation	coal, other fossil fuel, oil	32.306
Datang Tuoketuo power station	China	Satellite + Machine learning	coal	31.609
Surgut GRES-2 power station	Russia	Country + Fuel Type Imputation	gas	31.590
Dangjin power station	Korea	Country + Fuel Type Imputation	coal	31.020

4. Discussion

4.1 Global- and Country-Level Estimates and Comparisons to Existing Data

4.1.1 Existing Emissions Inventories

There are a variety of emissions inventories that include the energy sector that make both global and country level emissions estimates. Inventories available for energy include the Potsdam Institute for Climate Impact Research (PIK), United Nations Framework Convention on Climate Change (UNFCCC) National Submission, Emissions Database for Global Atmospheric Research (EDGAR) (Crippa et al., 2022), CAIT, IEA global energy related emissions data and Carbon Monitor (Liu et al., 2020). The Climate TRACE electricity generation sector currently covers emissions from thermal power plants with fuel sources including coal, natural gas, fuel oil, diesel, biomass and waste for the period 2015-2021 for 227 countries.

There are some challenges in comparing Climate TRACE estimates to other inventories. The UNFCCC includes a breakdown of the different sub-sectors of energy, with the sectors and time period available varying by country. For Annex 1 countries, the category 1.A.1.a 'Public Electricity and Heat Production' is included in the UNFCCC estimates, which is more comparable to our data. The broader category of 1.A.1 Energy Industries used for non-Annex 1 countries, which includes electricity generation, as well as other energy industries, such as the production of fossil fuels makes comparison difficult. Because data for Non-Annex 1 countries is sparsely available for a similar time period of comparison, and also includes non-electricity related emissions sources, it was not used to generate world-level comparisons.

Some estimates such as the IEA and the PIK include additional sectors, such as transport and manufacturing, with electricity generation in their estimate. To use these estimates as direct comparisons to our data the additional sectors need to be removed. To generate a suitable comparison with PIK for Annex 1 countries, the UNFCCC estimate for all IPCC1 subcategories other than 1.A.1.a 'Public Electricity and Heat Production' were subtracted from PIK's IPCC1 estimate. As a result, some of the variance between PIK and other inventories may be due to differences in subcategories other than electricity generation.

The inventories have different temporal coverages in terms of range and years sampled, which sometimes vary by country. The UNFCCC's best coverage is for Annex I countries, like Germany which has yearly data going back decades with the most recent data point covering 2020. For Indonesia, for example, the most recent two emissions estimates were in 2000 and 1994. PIK has more granular and recent coverage for non-Annex 1 countries, with yearly estimates for countries like India and Indonesia, most recently in 2019 going back decades. However, as we require UNFCCC subcategory data to reliably compare our electricity emissions, the PIK data's

comparability is limited by UNFCCC temporal granularity and recency. CAIT has yearly estimates from 1990-2019 and EDGAR has yearly estimates from 1970-2021.

The most recent directly comparable data is provided by Carbon Monitor, which provides near real-time data in the range 2019-2022. Carbon Monitor covers a more limited range of countries than our inventory, however, including China, India, U.S., Europe (EU27 & UK), Russia, Japan, Brazil, and rest-of-world (ROW) as a single category. For these countries Carbon Monitor represents the most recent data point available of the existing inventories compared.

Additionally, some data sources rely on some of the same data sources for their analysis processed and combined in different ways. For example, our data (through EMBER), CAIT, and EDGAR rely on IEA data for parts of their analysis. On a country level, therefore, there is a large variation in which data sets are available and in which time range for comparison.

4.1.2 Global Emission Estimates Compared to Other Inventories

Figure 8 highlights global and Annex I emissions. The top chart in Figure 8, for years 2015 to 2018, EDGAR reports higher emissions, an average of 13.53 billion metric tons, whereas Climate TRACE reports a slightly lower average value, 12.68 billion metric tons. EDGAR total global emissions for "Public Electricity and Heat Production" steadily increased from ~13.4 to ~13.9 billion metric tons for 2015 to 2018, respectively, a 3.6% increase. Similarly, Climate TRACE emission estimates report an increase from ~12.34 to ~13.10 billion metric tons for 2015 to 2018, respectively, a 6% increase. Both EDGAR "Public Electricity and Heat Production" data and Climate TRACE emission estimates show the pandemic-related decrease in emissions in 2020, and the rebound in 2021 as demand for electricity increased as a result of re-openings. This is similarly reported in Liu et al. (2020) and Davis et al. (2022).

In contrast to the global emissions, Annex I countries display a generally decreasing trend (Figure 8, bottom chart) for all reported emissions prior to 2021, then both EDGAR and Climate TRACE both show a rebound as seen in the comparisons above. In 2015, the emissions are 5.05, 4.65, and 4.63 billion metric tons for EDGAR, UNFCCC, and Climate TRACE, respectively. For 2018, all report lower emissions, 4.77, 4.34, 4.41 billion metric tons for EDGAR, UNFCCC, and Climate TRACE, respectively.

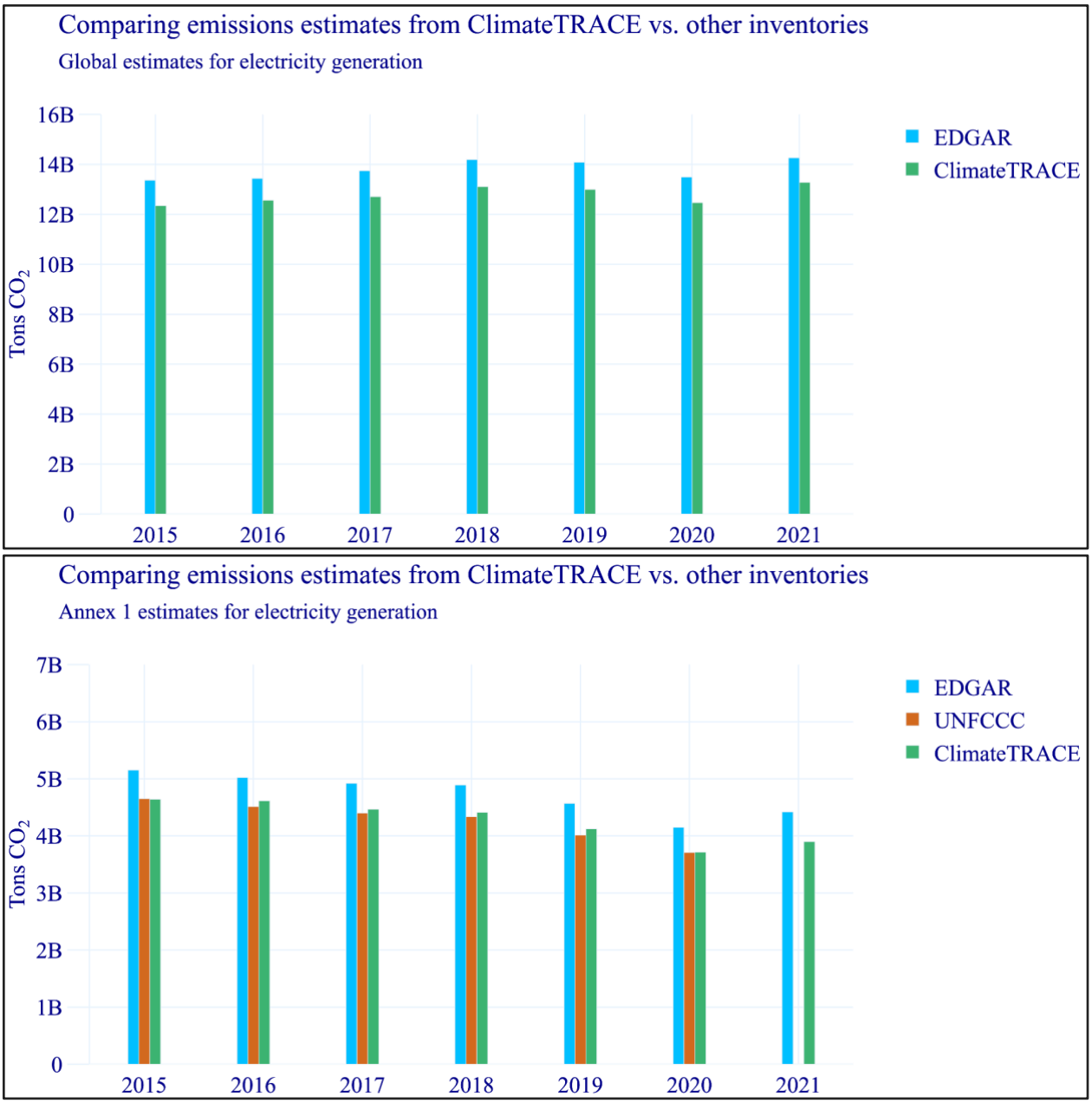


Figure 8 Top chart, total global emission estimates for electricity generation. Bottom chart, total Annex I emission estimates. EDGAR = blue bars, UNFCCC = red bars, and Climate TRACE = green bars. The Y-axis displays billions of metric tons of CO_2. UNFCCC Annex I emission estimates were only available up to 2020 at the time of this work.

The comparisons to both EDGAR 'Public Electricity and Heat Production' and UNFCCC 'Public Electricity and Heat Production' for Annex 1 and global totals show a similarity in trends and a lower estimate than EDGAR. One possible explanation for these differences is that the Climate TRACE estimates do not include the emissions from heat production that is not associated with

the generation of electricity. This is the case in China, Russia, and Europe where district heating infrastructure is prevalent (IEA Emissions Intensity Index 2021).

4.1.3 Selected Country-Level Comparisons

Here we present country-level emissions estimates for the countries listed in Table 5. These are compared to the existing emissions inventories described in section 4.1.1 where available, namely CAIT, UNFCCC, PIK and EDGAR. In addition, we highlight South Africa and Laos, countries in which we predominantly rely on satellite-derived predictions for our estimates.

For China, for years where inventories overlap with Climate TRACE estimates, all show increasing emissions estimates (Figure 9). However, only Climate TRACE, EDGAR, and Carbon Monitor provide the most recent estimates and display the emissions decrease in 2020. Climate TRACE China predictions are lower compared to EDGAR and CAIT for years 2015 to 2021 and 2015 to 2019, respectively. For EDGAR, this dataset includes heat production, which will increase emissions estimates to be higher than ours. This is the case in all regions, such as China, Russia, and Europe where district heating infrastructure is prevalent (IEA Emissions Intensity Index 2021). Carbon Monitor emissions estimates are slightly higher than our estimates. These differences may stem from the challenge in assembling a comprehensive inventory of Chinese power plants. We estimate that we include >95% of the coal capacity in China in our inventory. Missing capacity can bias our estimates low, given how unit capacities are core to both asset- and country-level estimates. EIA and EMBER capacity and generation estimates show that plant-level capacity factors in China are high relative to the countries in our ML training set, which can lead to underestimation of capacity factors in plants covered by our ML models.

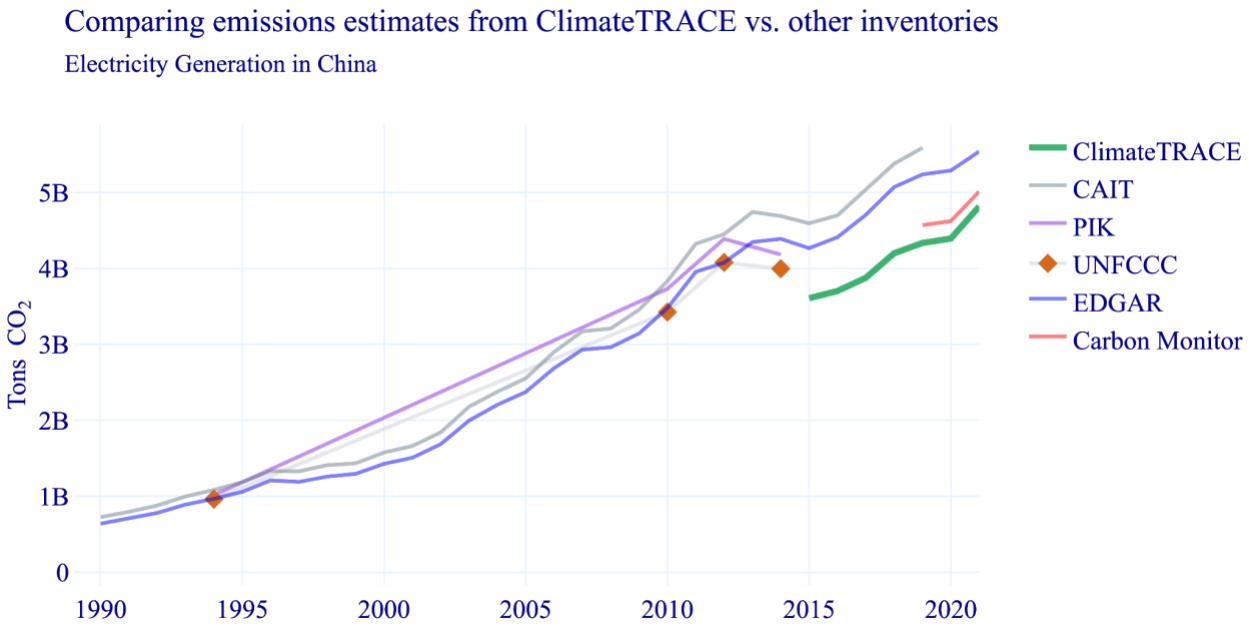


Figure 9 Cross inventory comparison for China. The most recent UNFCCC emissions, and therefore PIK, data point is from 2014 and is not used for comparison. Climate TRACE results are compared to EDGAR in the 2015-2021 overlap, Carbon Monitor in 2019-2021, and CAIT in 2015-2019. The UNFCCC data is for the 1.A.1 Energy Industries sector.

In Figure 10, our predictions in the USA align well with other sources. In particular, Climate TRACE estimates align closely with EDGAR estimates in the years 2016-2021 and display the generally decreasing emissions trend reported in both Carbon Monitor and EDGAR. Overall, all inventories display decreasing emissions for overlapping years. This is largely expected, since most of our plant-level emissions estimates are either directly derived from high-quality asset-level generation data or country-level capacity data.

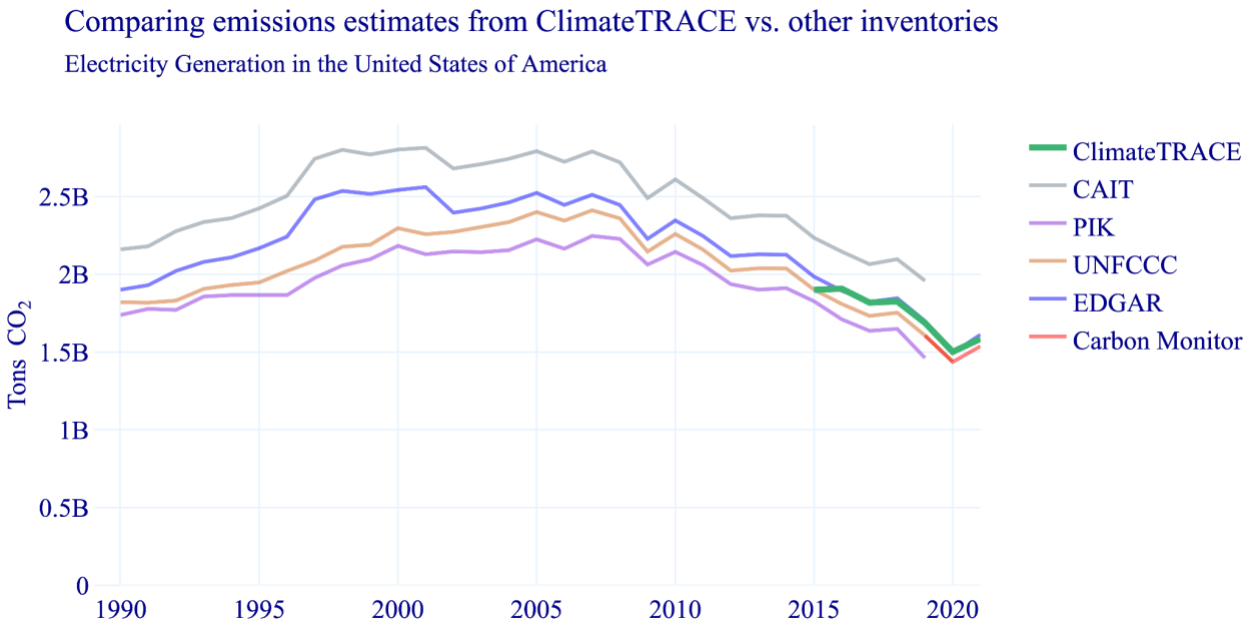


Figure 10 Cross inventory comparison for the USA. The UNFCCC and PIK have data between 1990-2019, while EDGAR has data shown from 1990-2021. We compare these inventories in the range 2015-2019 and to Carbon Monitor in the range 2019-2021. UNFCCC data is for the sector 1.A.1.a 'Public Electricity and Heat Production'.

In India, our estimates broadly agree with other sources (Figure 11). EDGAR data matches our emissions data well. For the time ranges covered by Carbon Monitor and CAIT, our estimates match within 4-9 % for CAIT and 5-7% for Carbon Monitor, with the same dip in emissions in 2020 as seen in the Carbon Monitor data.

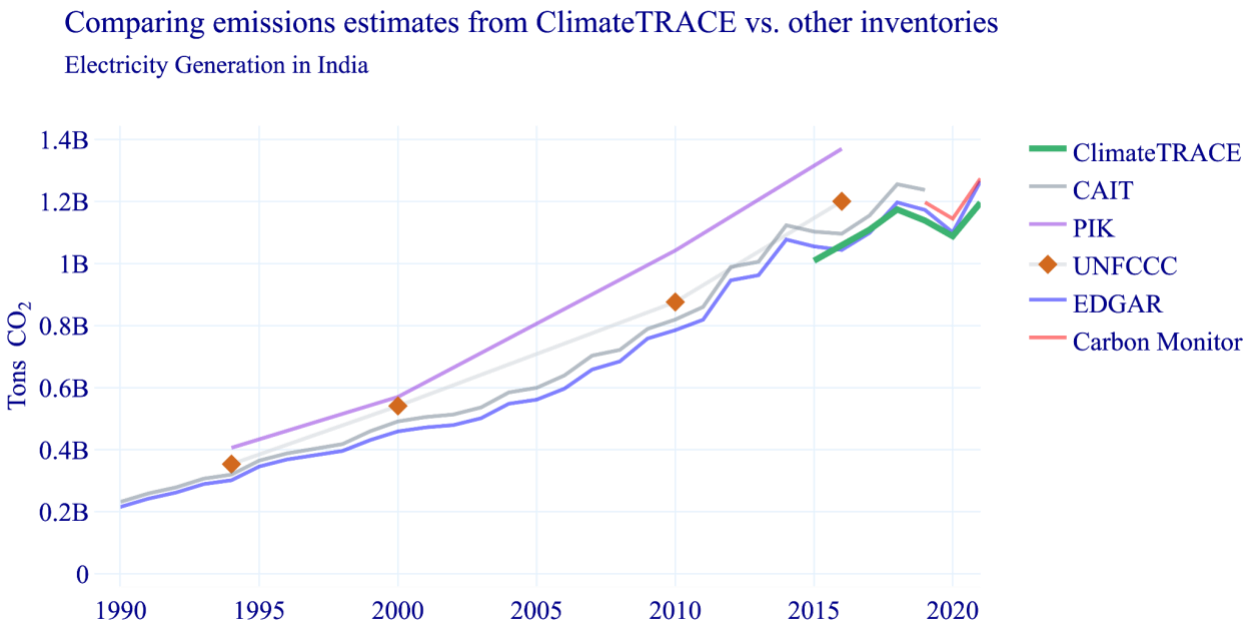


Figure 11 Cross inventory comparison for India. We compare with UNFCCC and PIK data in the 2015-2016 overlap, EDGAR in 2015-2018, CAIT in 2015-2019, and Carbon Monitor in 2019-2022. The UNFCCC data is for the 1.A.1 Energy Industries sector.

Russia is a country where there is a larger discrepancy between Climate TRACE estimates and other data sources (Figure 12). Our 2021 estimate only represents 64% of that of Carbon Monitor. Given Climate TRACE's reliance on generation estimates from EIA and EMBER and that we calibrate country-specific carbon intensities with published data, the relatively large difference in the Russian emissions estimates has several potential sources. One possibility is that we are underestimating thermal power generation in Russia, resulting in lower Climate TRACE emission estimates. Another could be that Russia has higher carbon intensities than we estimate. Finally, our inventory of Russian power plants may be incomplete. This hypothesis has a larger likelihood for Russia data than in regions as the USA, Europe and Australia where, as raised in section 2.2, there is more granular data available on power generation.

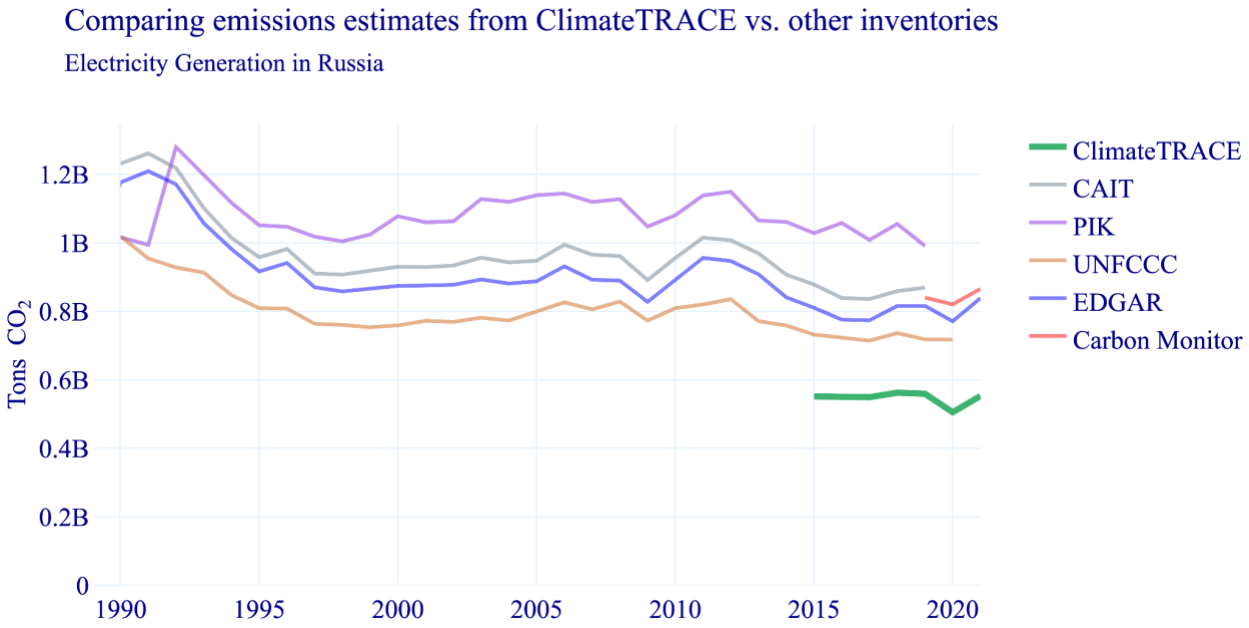


Figure 12 Cross inventory comparison for Russia. We compare to UNFCCC in 2015-2020, EDGAR in 2015-2021, CAIT 2015-2019, and Carbon Monitor 2019-2021. UNFCCC data is for the sector 1.A.1.a 'Public Electricity and Heat Production'.

In Japan we observe a relatively wide range in inventory emissions estimates (Figure 13). Our estimates fall somewhere in the middle of this range. We estimate low compared to Carbon Monitor and EDGAR through 2019-2021. Satellite-derived estimates do not contribute significantly to our Japan estimates, so these differences are likely explained by disagreements in emissions factors or country-level capacity factors employed by either source.

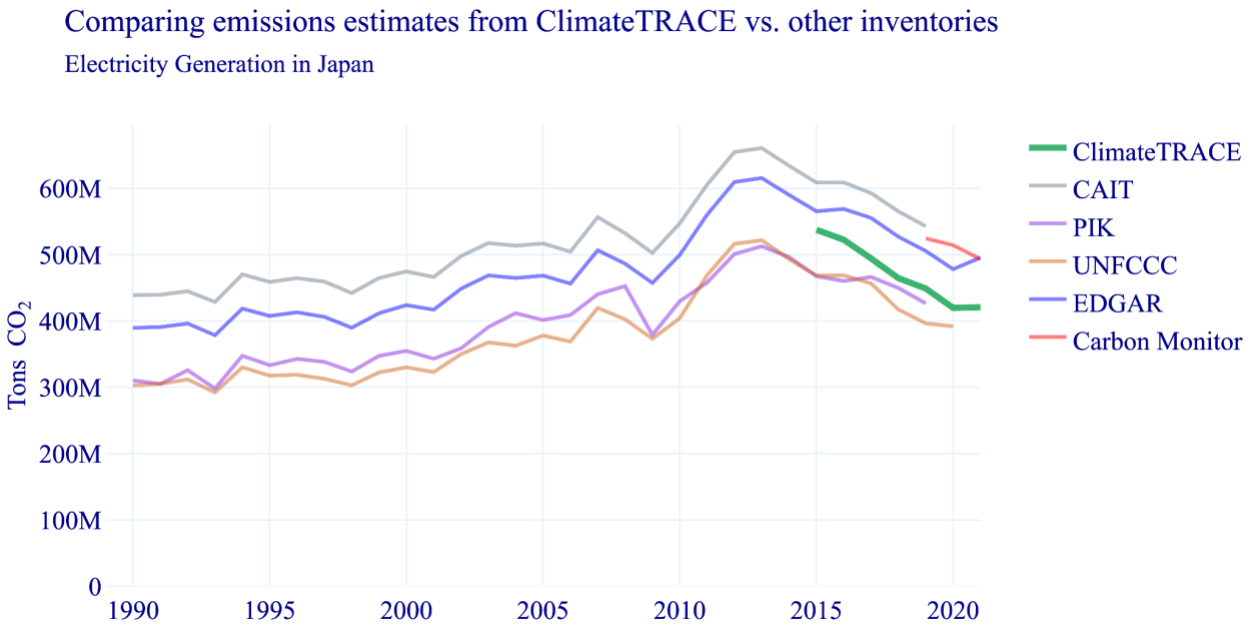


Figure 13 Cross inventory comparison for Japan. We compare to CAIT in 2015-2019, EDGAR in 2015-2021, PIK in 2015-2019, Carbon Monitor in 2019-2021, and UNFCCC 2015-2020. UNFCCC data is for the sector 1.A.1.a 'Public Electricity and Heat Production'.

Our estimates for South Africa contain a significant contribution from ML predictions (Figure 14). We predict significantly lower emissions than CAIT in the periods that the data overlap and follow more closely to EDGAR. We expect our ML estimates to be less accurate in this period due to a relative lack of satellite imagery. In particular, the majority of PlanetScope satellites were launched in 2017 or later and Sentinel-2B was launched in 2017, making satellite derived estimates in years 2015 to 2017 less accurate due to the limited satellite coverage.

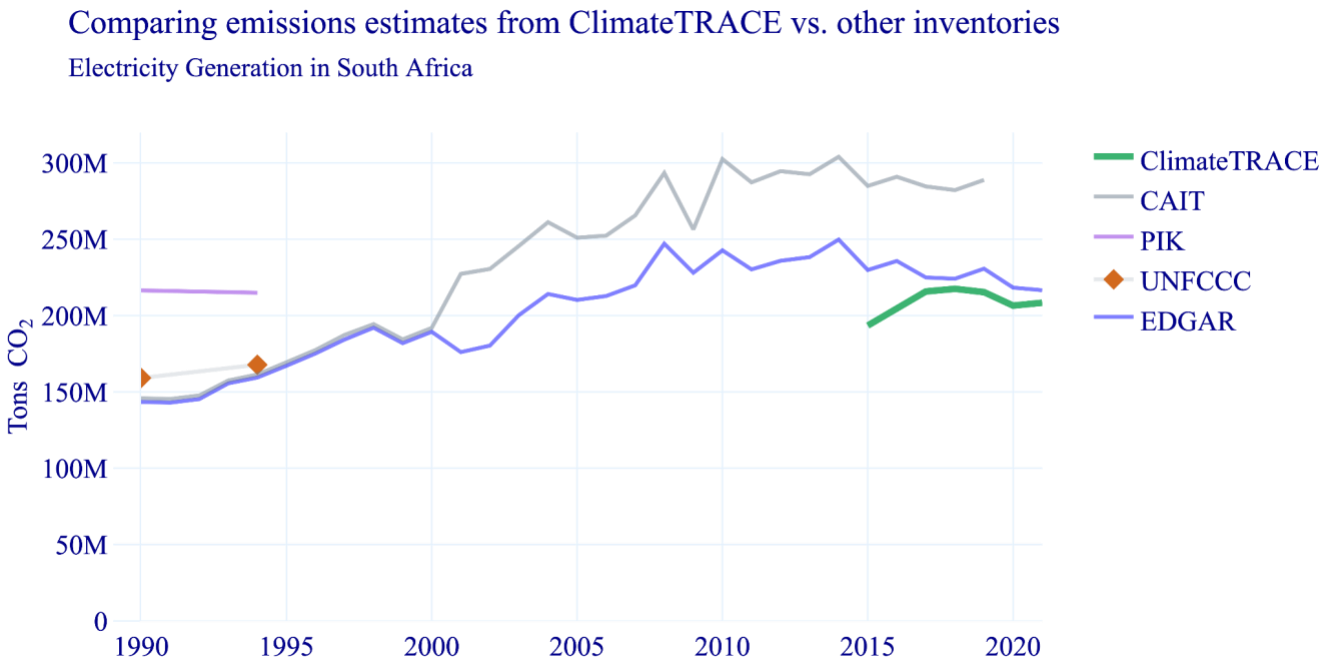


Figure 14 Cross inventory comparison for South Africa. UNFCCC's, and therefore PIK's, most recent data point is from 1994, so not used in the comparison. We compare to CAIT in 2015-2019 and EDGAR in 2015-2021. Carbon Monitor does not cover South Africa. The UNFCCC data is for the 1.A.1 Energy Industries sector.

Figure 15 displays Laos country emissions, where the estimate comprises a large fraction of satellite-derived predictions. Our predictions roughly match the shape of the CAIT and EDGAR datasets, a rapid increase from 2015 to 2017, but with a 10-20% relative underestimation in the years 2017-2021. We note that average capacity factors for fossil fuel plants in Laos are high, around 62-68%, which can cause underestimation by our ML models. The UNFCCC and PIK do not publish data for Laos.

Laos, along with many other smaller countries, have less recent inventory data to compare to than countries with larger capacities and, in some cases, are not covered by large data providers like UNFCCC. Our global, satellite-based approach has the ability to cover countries which are more challenging to estimate using other techniques.

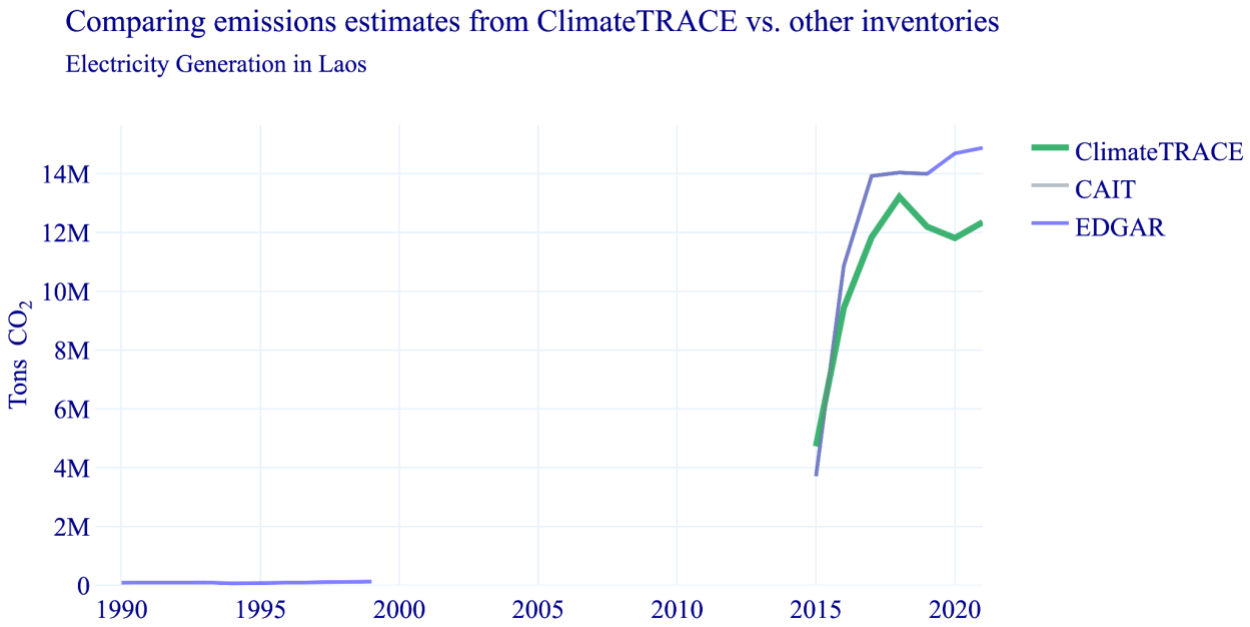


Figure 15 Cross inventory comparison for Laos. We compare to CAIT in 2015-2019 and EDGAR in 2015-2021. These two inventories have very similar values in this time range. EDGAR additionally covers 1990-1999. PIK, Carbon Monitor, and UNFCCC do not cover Laos. Note, the CAIT values overlap with the EDGAR values.

4.2 Asset-Level Estimates

Of the total asset level inventory of 565 plants, 218 have their estimates derived from our machine learning approach, 48 have their estimates derived from the ground truth, while 299 have their estimates based on country and fuel type imputation (Figure 16). These plants globally represent the largest emitters in the top carbon dioxide emitting sector, making this a critical increase in transparency on the asset level in the sector.

Of the 218 plants with satellite-derived estimates, 27 are in India, Taiwan, Brazil, or Turkey. These countries that have some plant level generation publicly available that we have not yet matched their data on a per plant basis to our power plant inventory sufficiently to use as ground truth comparisons. For the remaining 191 plants with satellite-derived data, we are not aware of any publicly available asset level data at the time of this work.

Our 565 assets are located in 41 different countries, shown in Figure 16. For 2021, this includes 272 power plants in China, 58 in India, and 50 in the US. Of these assets, 459 are coal-only plants as of 2021, while the remaining are gas, oil, biomass or other less-common fossil fuels. Many plants use a combination of two or more fuel types.

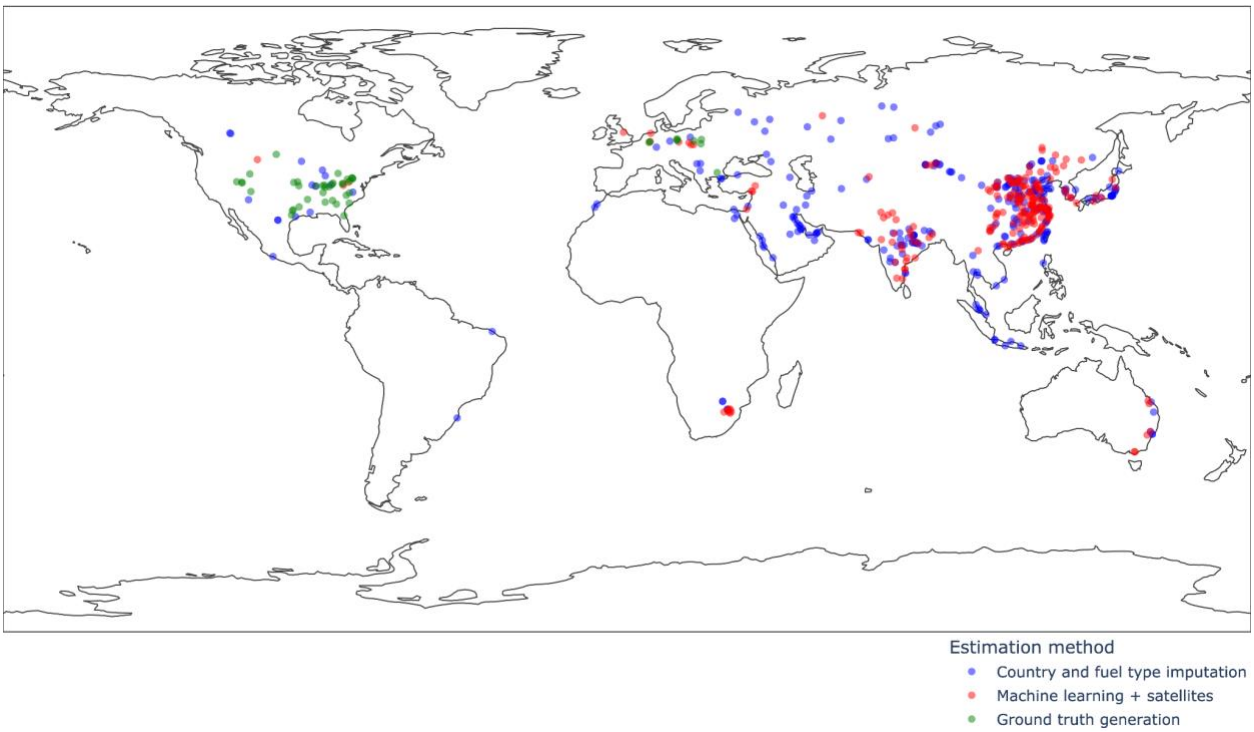


Figure 16 Map with all of the 565 power plants that are published in the Climate TRACE asset-level data, color-coded by estimation method- country and fuel type imputation (blue), machine learning and satellites (red), and ground truth generation (green).

The asset-level data allows for more localized emissions analysis, such as that seen in Figure 17. The five biggest emitters from Table 6 have their CO_2 emissions plotted between the years of 2019 and 2021. As expected, all have a drop in emissions during 2020, given the COVID-19 pandemic, but only the Taean and Dangjin power stations continued the emissions reduction path while the other power plants increased their emissions in 2021.

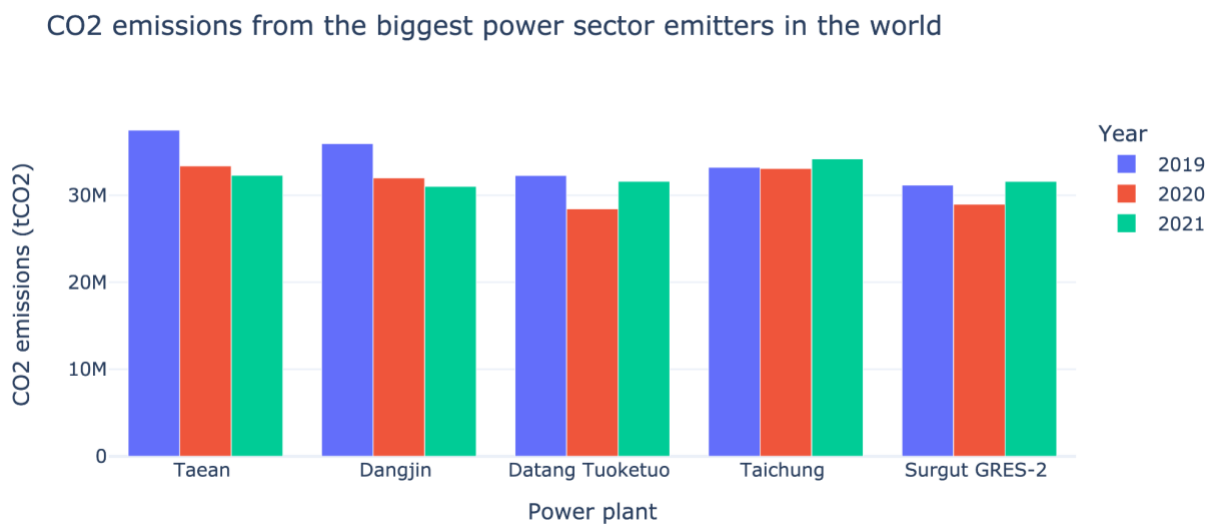


Figure 17 Estimated CO_2 emissions between the years of 2019 and 2021 for the five largest emitting power based on 2021. See Table 6 for additional information on these power plants.

5. Conclusion

By monitoring generation activity at individual power plants using automated plume detection in satellite imagery, it was possible to build a robust picture of generation activity which is more uniformly and recently available than reported data and more easily standardized and compared. Our approach is particularly impactful for assets for which no generation or emissions data publicly exists. The use of satellite imagery that is available in near real-time allows estimates to be prepared more quickly than other inventory methods and can track the emissions down to an asset level.

The Climate TRACE emissions estimates have been compared to other inventories at the global level, Annex I country aggregate level, and for selected countries. While there are some challenges in comparing Climate TRACE estimates to other inventories, the trends in our estimates are generally consistent with trends reported in other estimates. For both global and Annex I total emissions, Climate TRACE estimates were generally less relative to other emission inventories (Figure 8). One possible explanation for the lower estimates was that the Climate TRACE estimates do not include heat production emissions which are not associated with the generation of electricity.

In the five largest emitting countries, Climate TRACE emissions estimates for USA, India, and Japan match similar trends and fall within the range of estimates compared to other inventories (Figures 10, 11, and 13). While Climate TRACE reports lower emissions for China and Russia than other estimates (Figures 9 and 12). These lower emissions may be due to incompleteness in our asset inventory, an underestimation of total thermal generation, and/or an underestimate of the carbon intensity of the plants in these countries.

To continue to improve and refine our emissions estimates, we plan to expand our ability to model signals from more types of power plants, including those using mechanical draft or non-evaporative cooling processes. Currently, alternative sensors, including those with higher optical and thermal resolution than previously employed and those that monitor pollutants co-emitted with carbon dioxide, are being researched. We also plan to increase the precision of the carbon intensity modeling of individual power plants. Furthermore, the integration of continuously reported grid and plant-level data from more regions of the world can further improve the capability to comprehensively predict asset level emissions globally.

Finally, different approaches to processing already used satellite inventories are being planned, as well as the development of new machine learning approaches. In addition, our existing machine learning techniques and our methods in ensuring high quality of input and output data are in continuous development.

6. References

1. Bruckner, T., Bashmakov, I., Mulugetta, Y., Chum, H., de la Vega Navarro, A., Edmonds, J. , Faaij, A., Fungtammasan, B., Garg, A., Hertwich, E., Honnery, D., Infield, D., Kainuma, M., Khennas, S., Kim, S., Nimir, H.B., Riahi, K., Strachan, N., Wiser, R. and Zhang, X., 2014. Energy Systems. In: *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Edenhofer, O., R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T. Zwickel and J.C. Minx (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
2. Climate Watch Historical GHG Emissions. 2022. Washington, DC: World Resources Institute. Available online at: **https://www.climatewatchdata.org/ghg-emissions**
3. Couture, H., O'Connor, J., Mitchell, G., Söldner-Rembold, I., D'souza, D., Karra, K., Zhang, K., Kargar, A.R., Kassel, T., Goldman, B. and Tyrrell, D., 2020a. Towards tracking the emissions of every power plant on the planet. In *NeurIPS Workshop* (Vol. 3). Available: https://www.pixelscientia.com/pub/Couture-CCAI-NeurIPS2020.pdf
4. Couture, H., 2020b. How to track the emissions of every power plant on the planet from space. IEEE Spectrum. Available at: https://spectrum.ieee.org/energywise/energy/fossil-fuels/how-to-track-the-emissions-of-every-power-plant-on-theplanet-from-space)
5. Crippa, M., Guizzardi, D., Banja, M., Solazzo, E., Muntean, M., Schaaf, E., Pagani, F., Monforti-Ferrario, F., Olivier, J., Quadrelli, R., Risquez Martin, A., Taghavi-Moharamli, P., Grassi, G., Rossi, S., Jacome Felix Oom, D., Branco, A., San-Miguel-Ayanz, J. and Vignati, E., CO2 emissions of all world countries - 2022 Report, EUR 31182 EN, Publications Office of the European Union, Luxembourg, 2022, doi:10.2760/730164, JRC130363

6. Cusworth, D.H., Duren, R.M., Thorpe, A.K., Eastwood, M.L., Green, R.O., Dennison, P.E., Frankenberg, C., Heckler, J.W., Asner, G.P. and Miller, C.E., 2021. Quantifying global power plant carbon dioxide emissions with imaging spectroscopy. *AGU Advances*, *2*(2), p.e2020AV000350.
7. Davis, S.J., Liu, Z., Deng, Z., Zhu, B., Ke, P., Sun, T., Guo, R., Hong, C., Zheng, B., Wang, Y. and Boucher, O., 2022. Emissions rebound from the COVID-19 pandemic. *Nature Climate Change*, *12*(5), pp.412-414.
8. Dos Reis, A.A., Werner, J.P., Silva, B.C., Figueiredo, G.K., Antunes, J.F., Esquerdo, J.C., Coutinho, A.C., Lamparelli, R.A., Rocha, J.V. and Magalhães, P.S., 2020. Monitoring pasture aboveground biomass and canopy height in an integrated crop–livestock system using textural information from PlanetScope imagery. *Remote Sensing*, *12*(16), p.2534.
9. Elavarasan, R.M., Shafiullah, G.M., Padmanaban, S., Kumar, N.M., Annam, A., Vetrichelvan, A.M., Mihet-Popa, L. and Holm-Nielsen, J.B., 2020. A comprehensive review on renewable energy development, challenges, and policies of leading Indian states with an international perspective. *IEEE Access*, *8*, pp.74432-74457.
10. EMBER 2022, *Ember Data Explorer, 2022*, Available on: https://ember-climate.org/data/data-explorer/ (Accessed 3 October 2022)
11. IEA 2022. *Global energy-related CO2 emissions, 1990-2020***.** Available at: https://www.iea.org/data-and-statistics/charts/global-energy-related-co2-emissions-1990-2020 (Accessed 25 September 2022)
12. IEA Electricity Information 2022. *Electricity Information, July 2020***.** Available at: https://www.iea.org/data-and-statistics/data-product/electricity-information#documentation (Accessed 25 September 2022)
13. IEA Emissions Intensity Index 2021. *CO2 emissions intensity index for district heat production and heat production by country and region, 2021***.** Available at: https://www.iea.org/data-and-statistics/charts/co2-emissions-intensity-index-for-district-heat-production-and-heat-production-by-country-and-region-2021 (Accessed 3 October 2022)
14. Kuhlmann, G., Broquet, G., Marshall, J., Clément, V., Löscher, A., Meijer, Y. and Brunner, D., 2019. Detectability of CO2 emission plumes of cities and power plants with the Copernicus Anthropogenic CO2 Monitoring (CO2M) mission. *Atmospheric Measurement Techniques*, *12*(12), pp.6695-6719.
15. Liu, F., Duncan, B.N., Krotkov, N.A., Lamsal, L.N., Beirle, S., Griffin, D., McLinden, C.A., Goldberg, D.L. and Lu, Z., 2020. A methodology to constrain carbon dioxide emissions from coal-fired power plants using satellite observations of co-emitted nitrogen dioxide. *Atmospheric Chemistry and Physics*, *20*(1), pp.99-116.
16. Liu, Z., Ciais, P., Deng, Z., Lei, R., Davis, S. J., Feng, S., Zheng, B., Cui, D., Dou, X., He, P., Zhu, B., Lu, C., Ke, P., Sun, T., Wang, Y., Yue, X., Wang, Y., Lei, Y., Zhou, H., . . . Schellnhuber, H. J. (2020). COVID-19 causes record decline in global CO2 emissions. *arXiv*. https://doi.org/10.1038/s41467-020-18922-7

17. Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U. and Gascon, F., 2017, October. Sen2Cor for sentinel-2. In *Image and Signal Processing for Remote Sensing XXIII* (Vol. 10427, pp. 37-48). SPIE.
18. Marchese, F., Genzano, N., Neri, M., Falconieri, A., Mazzeo, G. and Pergola, N., 2019. A multi-channel algorithm for mapping volcanic thermal anomalies by means of Sentinel-2 MSI and Landsat-8 OLI data. *Remote Sensing*, *11*(23), p.2876
19. Mia, M.B., Fujimitsu, Y. and Nishijima, J., 2017. Thermal activity monitoring of an active volcano using Landsat 8/OLI-TIRS sensor images: A case study at the Aso volcanic area in southwest Japan. *Geosciences*, *7*(4), p.118.
20. Moon, M., Richardson, A.D. and Friedl, M.A., 2021. Multiscale assessment of land surface phenology from harmonized Landsat 8 and Sentinel-2, PlanetScope, and PhenoCam imagery. *Remote Sensing of Environment*, *266*, p.112716.
21. Planet, 2022. Planet imagery product specifications. https://assets.planet.com/docs/Planet_Combined_Imagery_Product_Specs_letter_screen.pdf
22. Shikwambana, L., Ncipha, X., Malahlela, O.E., Mbatha, N. and Sivakumar, V., 2019. Characterisation of aerosol constituents from wildfires using satellites and model data: A case study in Knysna, South Africa. *International Journal of Remote Sensing*, *40*(12), pp.4743-4761.
23. UNFCCC 2022. G*HG data from UNFCCC*. 1990-2020, Available at: https://unfccc.int/topics/mitigation/resources/registry-and-data/ghg-data-from-unfccc
24. Vaughn, T.L., Bell, C.S., Pickering, C.K., Schwietzke, S., Heath, G.A., Pétron, G., Zimmerle, D.J., Schnell, R.C. and Nummedal, D., 2018. Temporal variability largely explains top-down/bottom-up difference in methane emission estimates from a natural gas production region. *Proceedings of the National Academy of Sciences*, *115*(46), pp.11712-11717.