Estimating Greenhouse Gas Emissions from Satellite Imagery

Joëlle Hanna, Michael Mommert, Damian Borth
AIML Lab, School of Computer Science, University of St. Gallen

Universität St.Gallen

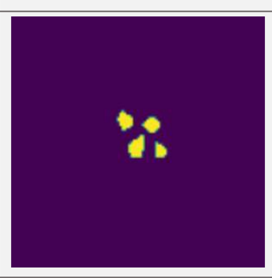
Introduction

Burning fossil fuels leads to air and water pollution, and constitutes the main driver of climate change. Emission reporting, which is necessary to properly assess the amount of released GHG emissions, is only required in some countries, resulting in insufficient global coverage. Such data may be used to enforce environmental protection regulations by the mean of pollutant certificate trading schemes.

Goal: In this work, we aim to predict the emission rates of GHGs from power plants at a given time through observations of the emitted smoke plumes from Earth-observing satellites.

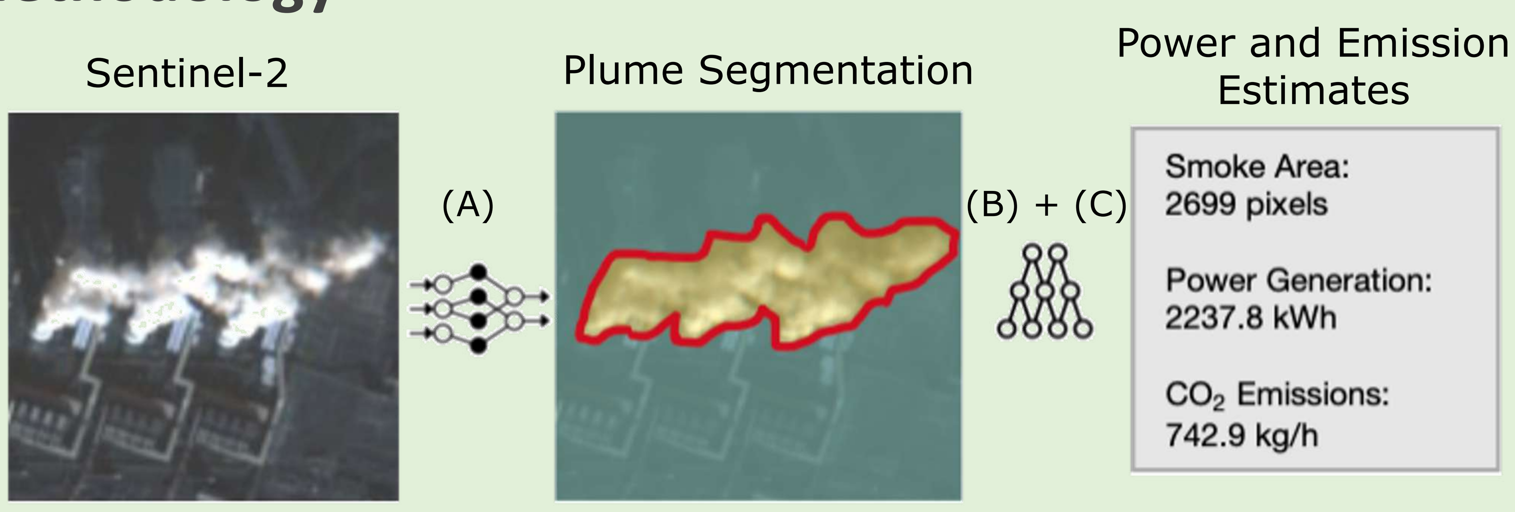
Dataset

Our dataset is constituted of 1600 (1) Sentinel-2 multispectral images centered over European Power Plants[1], with (2) their corresponding plume segmentation mask[1], (3) the actual generation output at the timestamp[2,3], and (4) concurrent weather information[4].

True Color Image	Plume Mask	Power Generation (MW)	Temperature (K)	Humidity (%)	Wind - u (m/s)	Wind - v (m/s)
		638	298.5	66.7	-1.39	0.54
		2050	290.4	53.1	-1.53	2.58
		500	274.5	84.1	5.41	1.47

We carefully divide our dataset into train (80%) and test (20%) sets making sure not to include the same site in more than one set.

Methodology



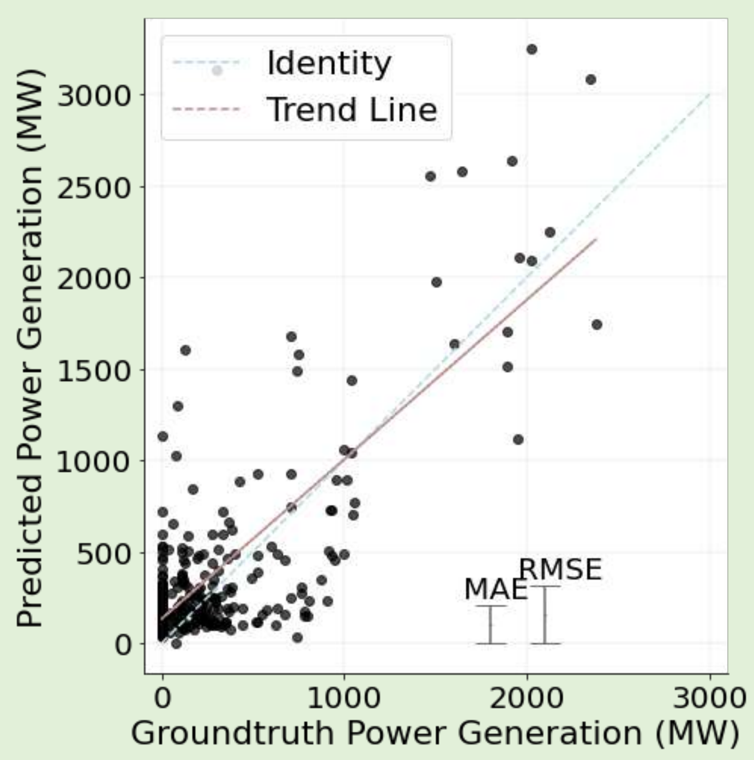
(A): We first segment the power plant's plumes from the satellite observation using a U-Net[5], to extract the corresponding plume areas. The model is trained using a dice loss, reaches a IoU of 0.61 and an accuracy at the image-level of 0.951.

(B): We feed the smoke plume area as well as weather data to an Extreme Gradient Boosting model[6] to predict the generation output of power plants. Weather data was added since environmental conditions may affect the shape and extent of the plume.

(C): We estimate the CO_2 emission rates from the predicted power generation output by applying a fuel-dependant conversion factor[7] distinguishing between coal (0.326 kg CO2/kWh) and natural gas (0.181 kg CO2/kWh)

Results

Predicted versus actual power generation for different satellite observations



Evaluation Metrics:

- RMSE $= \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N}}$
- MAE $= \frac{\sum_{i=1}^{N}|\hat{y}_i - y_i|}{N}$
- $R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$

Performance metrics derived for the test set and target variables.

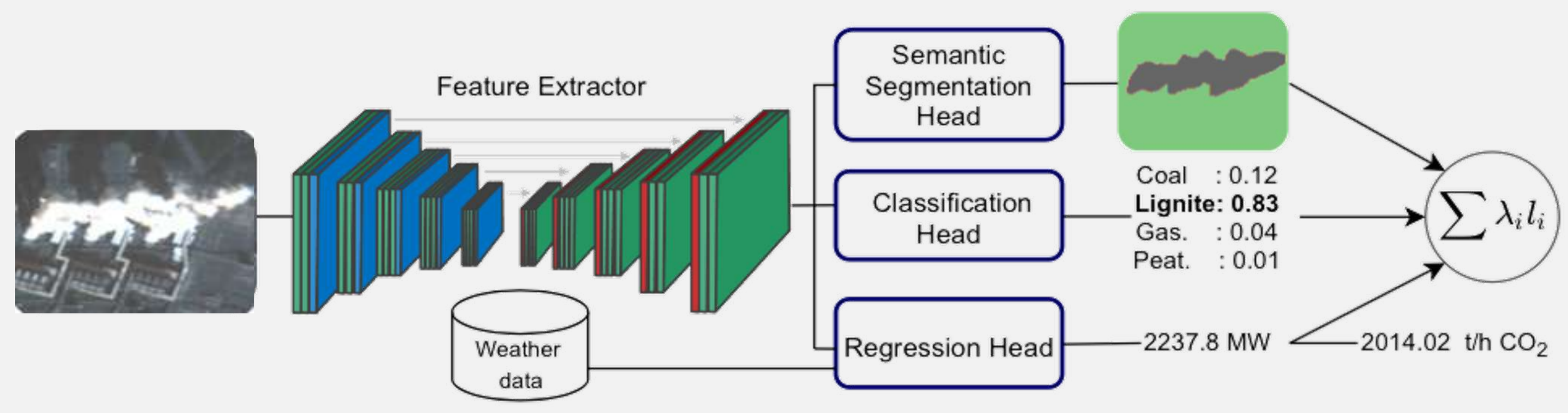
Target	Test set
Power generation - MAE (MW)	180
Power generation - RMSE (MW)	220
Power generation - R^2	0.69
CO_2 - MAE (converted, t/h)	50
CO_2 - RMSE (converted, t/h)	65

We are able to predict power generation rates to within 180 MW (MAE) translating into CO_2 emissions estimates to within 50 t/h (MAE).

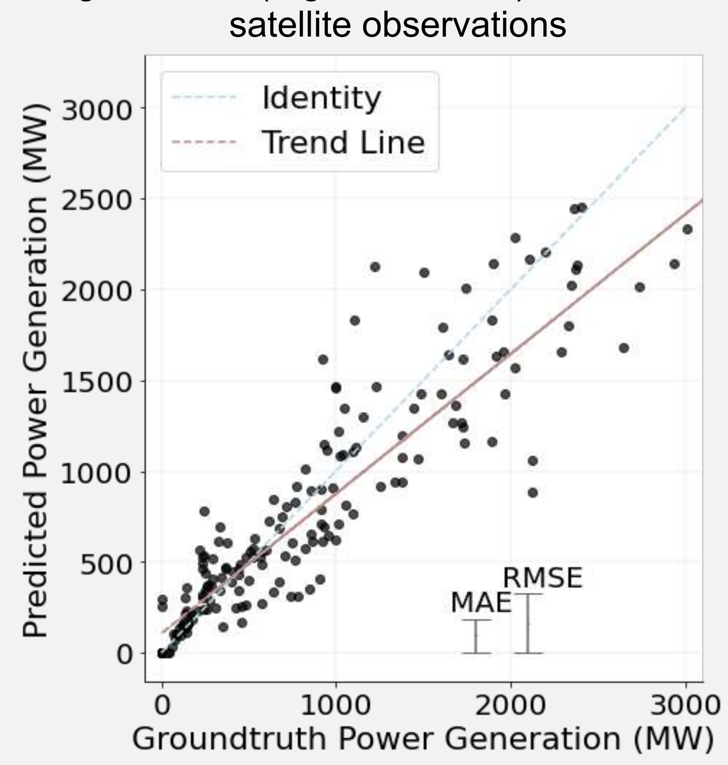
Multitask Learning Approach[8], extension of this Work – to be presented at 'Tackling Climate Change with Machine Learning' NeurIPS Workshop 2021

As an extension to this work, we propose an end-to-end method to predict power generation rates for fossil fuel power plants from satellite images based on which we estimate GHG emission rates. We present a multitask deep learning approach able to simultaneously predict: (i) the pixel-area covered by plumes from a single satellite image of a power plant, (ii) the type of fired fuel, and (iii) the power generation rate. We then convert the predicted power generation rate into estimates for the rate at which CO_2 is being emitted.

Diagram of the proposed multitask learning method



Predicted versus actual power generation (regression task) for different satellite observations



Performance metrics derived for the test set and target variables, for the regression task.

Target	Test set
Power generation - MAE (MW)	139
Power generation - RMSE (MW)	261
Power generation - R^2	0.83

We find that our multitask learning approach significantly boosts the perfromance of our approach. It also allows us to learn power plant types at the same time, enabling us to use better conversion factor for the estimation of CO_2 rates

References

[1] M. Mommert et al. Characterization of industrial smoke plumes from remote sensing data, 2020. Tackling Climate Change with Machine Learning Workshop at NeurIPS 2020
[2] Entsoe. Actual generation per generation unit table, 2021. https://transparency.entsoe.eu/generation/r2/actualGenerationPerGenerationUnit/show.
[3] K. Kanellopoulos et al. Jrc open power plants database (jrc-ppdb-open), December 2019. https://doi.org/10.5281/zenodo.3574566.
[4] H. Hersbach et al. The era5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146:1999–2049, 2020
[5] Ronneberger, O., Fischer, P., and Brox, T. U-Net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, May 2015
[6] Chen, T. and Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIG KDD International Conference on Knowledge Discovery and Data Mining, http://doi.acm.org/10.1145/2939672.2939785.
[7] Energy Information Administration. How much carbon dioxide is produced when different fuels are burned, 2020. https://www.eia.gov/tools/faqs/faq.php?id=73&t=11.
[8] J. Hanna et al. Multitask Learning for estimating power plant greenhouse gas emissions from satellite imager, 2021. Tackling Climate Change with Machine Learning Workshop at NeurIPS 2021

Take-home messages

- It is possible to predict power generation output (and by extension CO_2 emissions) directly from Satellite Images.
- Using a multi-task approach, we are able to outperform our baseline in terms of MAE and R^2, by making use of the entire image (as opposed to only the plumes area with our baseline).