Employing Deep Learning to Quantify Power Plant Greenhouse Gas Emissions via Remote Sensing Data

Aryan Jain
Amador Valley High School
1155 Santa Rita Rd, Pleasanton, CA 94566
Pleasanton, California 94588
aryanjn09@gmail.com

Abstract

Greenhouse gasses (GHG) emitted from fossil-fuel-burning power plants pose a global threat to climate and public health. GHG emissions degrade air quality and increase the frequency of natural disasters five-fold, causing 8.7 million deaths per year. Quantifying GHG emissions is crucial for the success of the planet. However, current methods to track emissions cost upwards of $520,000/plant. These methods are cost prohibitive for developing countries, and are not globally standardized, leading to inaccurate emissions reports from nations and companies. I developed a low-cost solution via an end-to-end deep learning pipeline that utilizes observations of emitted smoke plumes in satellite imagery to provide an accurate, precise system for quantifying power plant GHG emissions by segmentation of power plant smoke plumes, classification of the plant fossil fuels, and algorithmic prediction of power generation and CO_2 emissions. The pipeline was able to achieve a segmentation Intersection Over Union (IoU) score of 0.841, fuel classification accuracy of 92%, and quantify power generation and CO_2 emission rates with R^2 values of 0.852 and 0.824 respectively. The results of this work serve as a step toward the low-cost monitoring and detection of major sources of GHG emissions, helping limit their catastrophic impacts on climate and our planet.

Introduction

Fossil-fuel power plants are one of the largest emitters of Greenhouse gasses, accounting for 73% of the U.S.' GHG emissions and 65% of global GHG emissions (on Climate Change and Edenhofer 2014). The pollutants produced by these emissions serve as major contributors to the climate crisis and have had devastating impacts on air quality and the environment. GHG emissions cause 8.7 million deaths per year and have increased the frequency of natural disasters such as wildfires and powerful storms five fold (Smol 2012). These public health and environmental impacts cost billions in annual damages.

Preventing the permanent effects of climate change and air pollution requires identifying the sources and distributions of GHG emissions on a precise scale. However, keeping track of GHG emissions from all global power plants is difficult, as the quality of emissions data varies depending on each country's reporting protocols, maturity of their infrastructure, and availability of proper monitoring systems. For example, in a developed country such as the United States, every major power plant has on-site Continuous Emissions Monitoring Systems (CEMS) that reports data to the Environmental Protection Agency. But these systems are very expensive, costing over $500,000 for installation and $20,000 annually for maintenance (US EPA 2016), making them impractical for use in many lesser-developed countries. Additionally, the lack of reliable infrastructure causes many countries to provide vague, inaccurate, and outdated estimations of their GHG emissions. An examination of GHG emission reports from 196 countries found gaps in nation's declared emissions versus estimates by the United Nations totalling to 10.6 billion tons of globally under-reported emissions per year (Mooney et al.).

These issues require new, low-cost alternatives to estimate and report GHG emissions on a more precise scale. In recent years, the use of satellite data has emerged as a potential candidate to monitor the progression of climate change and global warming (Boesch et al. 2021). Equipped with an array of sensors and instruments to measure various atmospheric conditions, spectrometer satellites have helped inform our understanding of the dynamics of changes in Earth's temperature. Launched in 2009 and 2014, spectrometer satellite missions Greenhouse Gasses Observing Satellite (GOSAT) and Orbiting Carbon Observatory (OCO-2) have provided carbon dioxide (CO_2) emission data on a global and national level (Eldering et al. 2017). However, spectrometer satellite instruments are imprecise and low-resolution ($\geq$ 10 km resolution), and cannot identify the granular emissions ($\leq$ 2km) of individual power plants (Apte et al. 2017).

When active, fossil-fuel burning power plants emit a smoke plume as a byproduct of the electricity generation process. These plumes can be captured by optical satellite imagery and fed into a deep learning model to produce accurate estimates of the plant's GHG emissions (Cusworth et al. 2021). Pairing deep learning with high-resolution optical satellite imagery serves as a promising method to estimate power plant GHG emissions with accuracy rates near spectrometer measurements, while simultaneously maintaining the ability to monitor emissions on a global scale. Moreover, this method does not require huge investments or elaborate infrastructure, serving as a low-cost alternative to fill-

ing long existing gaps in emissions data around the world. In this work, I present an end-to-end deep learning pipeline to estimate CO_2 emissions, the most dominant greenhouse gas in the atmosphere, at an individual power-plant scale. My pipeline processes a single multi-spectral satellite image and associated weather data to extract smoke plumes from power plants and estimate power generation and CO_2 emission rates. The results of this work serve as a step towards the detection and monitoring of major sources of power plant GHG emissions on a global scale at a low-cost.

Previous Works

Previous works have explored the relations between plume imagery and GHG emission rates, and the applications of machine learning in predicting power plant behavior. Cusworth et al. (Cusworth et al. 2021) employed airborne visible/infrared imaging spectrometers (AVIRIS-NG) to quantify the carbon dioxide (CO_2) and methane (CH_4) emissions of 17 power plants from their smoke and vapor plumes. Aided by plant-specific annotations, Climate TRACE, a coalition working towards tracking all greenhouse gas emissions from anthropogenic activities, has been able to estimate plant generation and emission rates from satellite imagery. Couture et al. (Couture et al. 2020) details Climate TRACE's methods in annotating cooling towers, flue stacks, and water outlets to aid in their model's predictions. Both Cusworth and Climate TRACE's respective approaches are reliant on extensive data preparation and annotation, thus making it difficult to produce a generalizable solution that can scale across large regions. More recently, Hannna et al. (Hanna et al. 2021) demonstrated the promise of using plume segmentation to inform more generalizable model predictions, feeding a satellite image as an input to a pipeline capable of plume segmentation, power plant classification, and power generation prediction. This work builds off Hannna's research by adding CO_2 flux rates to the dataset, and comparing various state-of-the-art machine learning architectures to produce a pipeline that performs well across the plume segmentation, fossil fuel classification, power generation regression, and CO_2 regression tasks.

Methods

Dataset

The dataset from Hanna et al. is comprised of 2131 samples of multi-spectral satellite images taken by ESA's Sentinel-2 satellites (Drusch et al. 2012). The resolution of the imagery is 120px $\times$ 120px at 10m/px to cover an area of 1.2km $\times$ 1.2km on the ground. Each image has a corresponding smoke plume mask that is used to train the segmentation models. These samples are paired with the plant's longitude and latitude coordinates, country, weather data (temperature, humidity, wind), type of fossil fuel, and power generation rates. Using reported annual CO_2 emissions and power plant generation capacities sourced from the European Union Emissions Trading System (Verena Graichen 2019), I convert the provided power generation rates into CO_2 emission, or flux, rates. The CO_2 flux rate of the plants ranges from 307 tons/hour to 2834 tons/hour, with the average flux rate being 1548 tons/hour.

Data Preprocessing

The data was split with 70% (1507 samples) going into the training set and 30% (624 samples) going in the testing sets such that each set did not contain images of the same power plant. All images in the dataset were normalized to reduce the effect of background objects or noise in the image. Then, all the images from the training set were duplicated five fold to increase the size of the training data to 7535 samples, and they all underwent a data augmentation process where they were randomly mirrored, flipped, cropped, and rotated a random amount between 0° and 360° both clockwise and counter-clockwise. This augmentation serves to generate a diverse set of possible plume orientations and center locations that should help the model better generalize and prevent over fitting to the training set.

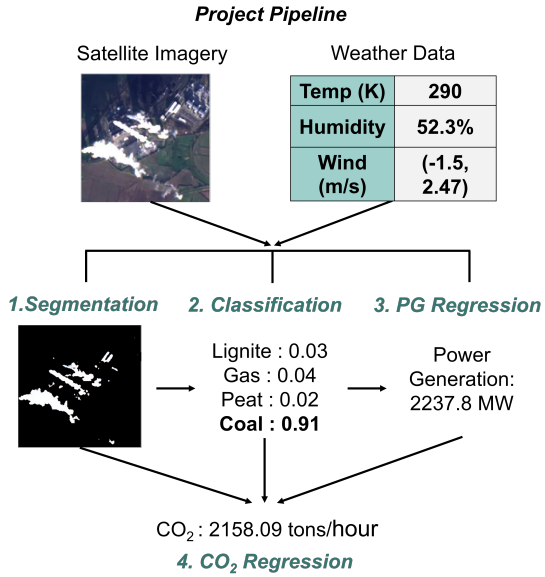

Figure 1: Diagram of the model pipeline. It takes a multi-spectral satellite image as input and learns to do four tasks: (1) semantic segmentation of smoke plumes, (2) classification of type of fossil fuel, and (3) regression with respect to power generation and (4) CO_2 emission rates.

Model Pipeline

The pipeline needs to accomplish four tasks: (1) semantic segmentation of smoke plumes in the satellite imagery, (2) classification of the type of fossil fuels being used by the power plant, (3) prediction of the plant's power generation rate, and (4) prediction of the CO_2 flux rate. Figure 1 shows the structure and flow of the model pipeline, and how the models for tasks 2-4 use outputs of other models to help inform their predictions. This is most significantly utilized for task 4, the prediction of the CO_2 flux rate, which uses the output of all three previous tasks as input to the model. For

each task, I evaluated 3 state-of-the-art model architectures that have shown to generally perform well in their respective tasks.

Segmentation

For the segmentation task, I chose FCN (Fully Convolutional Network), U-Net, and DeepLabV3 for experimentation (Long, Shelhamer, and Darrell 2014), (Ronneberger, Fischer, and Brox 2015), (Chen et al. 2017). The FCN model consists of a set of max-pooling and convolution layers to identify and segment features in an image. The U-Net is based on FCN, but it employs an Encoder-Decoder architecture consisting of contracting and expanding convolutional layers. DeepLabv3 is a pre-trained model that also employs an encoder-decoder architecture with spatial pyramind pooling layers and atrous convolution techniques to learn about the larger context of the image it is segmenting. I measure performance on this task using Intersection Over Union (IoU) and the loss function is binary cross entropy.

Classification

For classification, I employed transfer learning, and tested pre-trained models Res-Net 50, VGG-16, and InceptionV3, which were all created with different metrics to optimize (He et al. 2016), (Simonyan and Zisserman 2015), (Szegedy et al. 2016). ResNet prioritizes finding the simplest solution through shortcut connections. VGG-16 is an optimized convolutional neural network model (CNN) with a focus on faster learning without over-fitting. InceptionV3 uses multiple kernal sizes to adapt to finding both larger, global features and smaller, area-specific features in an image, which is necessary for this task, as plumes can span across the entire satellite image or be a single spot in its corner. The chosen loss function is cross entropy loss, and the evaluation metric for this task is accuracy and Area Under the Curve (AUC).

Regression

Tasks 3 and 4 are regression problems, in which I evaluated Linear Regression, Artificial Neural Networks (ANN), and XGBoost (eXtreme Gradient Boost) (Chen and Guestrin 2016). Linear Regression models the relationship between a set of variables through a linear equation. ANNs employ the neural network architecture and have done well in regression tasks. XGBoost is an implementation of the gradient boosted trees algorithm that learns to fit data by minimizing a regularized (L1 and L2) objective function. L1 loss was selected as the loss function and performance was measured through the R^2 coefficient, Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE).

Results

To train each model, I performed a hyperparameter search using the library Optuna, a framework that automates the training process by automatically adjusting the hyperparameters to maximize each of the listed performance metrics above (Akiba et al. 2019). The results from the training and test sets of all the models discussed is shown in Table 1.

Table 1: Model Training Results

Model	Task	Metric	Train	Test
FCN	Seg.	IoU	.752	.684
DeepLabv3+	Seg.	IoU	.836	.769
U-Net	Seg.	IoU	**.903**	**.841**
VGG-16	Cls.	Acc.	76%	69%
Inceptionv3	Cls.	Acc.	86%	81%
ResNet50	Cls.	Acc.	**94%**	**92%**
Lin Reg	Pwr Reg.	R^2	.803	.651
ANN	Pwr Reg.	R^2	.837	.809
XGBoost	Pwr Reg.	R^2	**.893**	**.852**
Lin Reg	Flux Reg.	R^2	.723	.542
ANN	Flux Reg.	R^2	.815	.748
XGBoost	Flux Reg.	R^2	**.861**	**.824**

Plume Segmentation

The best performing segmentation model was the U-Net, achieving an IoU score of 0.903 on the training set and 0.841 on the test set. The model performed very well on samples where the plume masked the majority of the image, and performance declined on images with smaller plumes with more complicated shapes. I found that the model heavily utilized both associated weather data and certain multi-spectral imagery bands as key features that influenced its predictions. Particularly, the model used outside factors such as humidity and wind speeds to help it gain a larger context of the plume, and how it could have possibly been influenced by conditions that could not be captured by the satellite imagery. Moreover, the Short-wave Infrared (SWIR) and Water Vapor imagery bands were able to capture thermal and visual details about the smoke plume that helped the model differentiate the plume from other background noise in the image, such as clouds, light buildings, or other terrain.

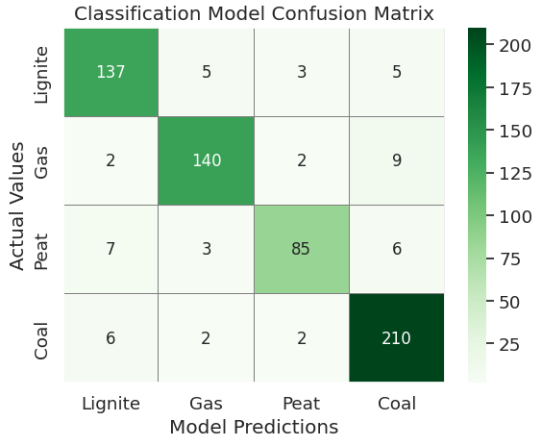

Figure 2: Confusion Matrix of ResNet-50 Model for Fossil-Fuel Classification Task.

Fossil Fuel Classification

For fossil-fuel classification, the ResNet50 model reached an accuracy rate of 94% on the training and 92% on the test set, much higher than InceptionV3 and VGG-16. This model was able to generalize very well across the four classes, coal, peat, gas, and lignite, and the test set results are displayed in the Confusion Matrix (Figure 2). One possible source of bias in this model comes from the unequal distributions of classes in the dataset, where coal is present more than twice as much as peat.

Power Plant Regression

The XGBoost model outperformed Linear Regression and ANN, gaining a R^2, MAPE of .861, 8.7% and .824, 10.2% on the training and test sets respectively. The output of the second model, the fossil-fuel classification prediction, had the most influence over these power generation predictions, as the per-hour CO_2 emissions from coal power plants are much larger than the emissions from peat or natural gas power plants (Raghuvanshi, Chandra, and Raghav 2006). Initially, the model was largely over-fitting to the training set, and this was reduced through increased data augmentation and the addition of several dropout layers, which both served as regularization techniques increasing the model's variances to the training data. This enabled a better general fit and increased performance on the test set, where it was giving predictions on plants it had never seen before.

CO_2 Flux Rate Regression

XGBoost was also the best performing model for CO_2 flux rate regression, achieving an R^2 value of .824 and a MAPE of 10.8% on the test set. Figure 3 exhibits this high performance, where the .87 line slope indicates a high correlation between the model's predictions and the ground truth data. Model performance on the CO_2 emission rate predictions was heavily dependent on the accuracy of the power generation predictions, as seen from the direct relationship between power generation and CO_2 flux rate mentioned above. The XGBoost model was able to generalize very well to the data, and it is a promising algorithm to further evaluate to see if it can continue to perform well across other regions.

Conclusions and Future Work

In this work, I developed an end-to-end deep learning pipeline that successfully predicted power generation and CO_2 emission rates across Europe via high resolution remote sensing data, an important step toward a future of accurate emissions monitoring across the globe. My pipeline performed well across all of its tasks (plume segmentation, fossil-fuel classification, power generation regression, and CO_2 flux rate regression) and demonstrates the promise of the plume segmentation approach acting as a possible generalizable solution to measure emissions across many power plants.

This project identified a number of features, techniques, and models that hold promise for evaluation in future works. The use of Shortwave Infrared (SWIR) imagery for differentiating plumes and other pollutants from background noise

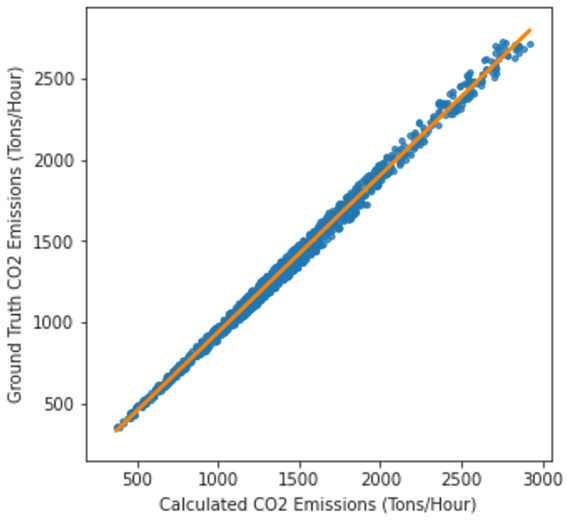

Figure 3: XGBoost Model Predicted CO_2 Emissions v.s. Ground Truth CO_2 Emissions (Flux Rate).

can serve as a key component for creating adaptive models to generalize to regional patterns and operate at night. The application of XGBoost in regression tasks can be further evaluated to see if the model can maintain its high accuracy rates across a larger sample size of data. Data accessibility remains a key component to the expansion of this work. Currently, the model has only trained on European power plants, and additional testing is required to measure model bias and see if this performance can translate to other regions and countries, such as the United States and China, in order for it to be truly globally scalable. In the near future, I aim to make this work more accurate and precise, with a focus on expanding to lesser-developed regions such as India and Brazil. Moreover, as more data becomes available, the pipeline can extended to predict other gases, such as such as methane (CH_4) and nitrous oxide (N_2O).

Global emissions monitoring systems will radicalize climate action efforts, providing a new level of reliable and transparent data that can aid governments and companies in designing effective climate policy. For example, by helping to identify "super-emitter" power plants, this pipeline pinpoints locations where government regulation is necessary and renewable alternatives will have the most impact. The results of this work serve as a step toward the low-cost monitoring and detection of major sources of GHG emissions, helping limit their catastrophic impacts on climate and our planet.

Acknowledgments

Part of this research was done in affiliation with WattTime, a member of the Climate TRACE coalition. I would like to thank Hannes Koenig, Jeremy Freeman, Heather Couture, and everyone else at WattTime for their help and mentorship that aided in the development of this work.

References

Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; and Koyama, M. 2019. Optuna: A Next-generation Hyperparameter Optimization Framework. *arXiv:1907.10902 [cs, stat]*. ArXiv: 1907.10902.

Apte, J. S.; Messier, K. P.; Gani, S.; Brauer, M.; Kirchstetter, T. W.; Lunden, M. M.; Marshall, J. D.; Portier, C. J.; Vermeulen, R. C.; and Hamburg, S. P. 2017. High-Resolution Air Pollution Mapping with Google Street View Cars: Exploiting Big Data. *Environmental Science & Technology*, 51(12): 6999–7008. Publisher: American Chemical Society.

Boesch, H.; Liu, Y.; Tamminen, J.; Yang, D.; Palmer, P. I.; Lindqvist, H.; Cai, Z.; Che, K.; Di Noia, A.; Feng, L.; Hakkarainen, J.; Ialongo, I.; Kalaitzi, N.; Karppinen, T.; Kivi, R.; Kivimäki, E.; Parker, R. J.; Preval, S.; Wang, J.; Webb, A. J.; Yao, L.; and Chen, H. 2021. Monitoring Greenhouse Gases from Space. *Remote Sensing*, 13(14): 2700. Number: 14 Publisher: Multidisciplinary Digital Publishing Institute.

Chen, L.-C.; Papandreou, G.; Schroff, F.; and Adam, H. 2017. Rethinking Atrous Convolution for Semantic Image Segmentation.

Chen, T.; and Guestrin, C. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, 785–794. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-4232-2.

Couture, H. D.; O'Connor, J.; Mitchell, G.; Söldner-Rembold, I.; D'souza, D.; Karra, K.; Zhang, K.; Rouzbeh Kargar, A.; Kassel, T.; Goldman, B.; Tyrrell, D.; Czerwinski, W.; Talekar, A.; and McCormick, C. 2020. Towards Tracking the Emissions of Every Power Plant on the Planet. In *NeurIPS 2020 Workshop on Tackling Climate Change with Machine Learning*.

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): e2020AV000350. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020AV000350.

Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; Meygret, A.; Spoto, F.; Sy, O.; Marchese, F.; and Bargellini, P. 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, 120: 25–36.

Eldering, A.; Wennberg, P.; Crisp, D.; Schimel, D.; Gunson, M.; Chatterjee, A.; Liu, J.; Schwandner, F. M.; Sun, Y.; O'Dell, C.; Frankenberg, C.; Taylor, T.; Fisher, B.; Osterman, G.; Wunch, D.; Hakkarainen, J.; Tamminen, J.; and Weir, B. 2017. The Orbiting Carbon Observatory-2 early science investigations of regional carbon dioxide fluxes. *Science (New York, N.Y.)*, 358(6360): eaam5745.

Hanna, J.; Mommert, M.; Scheibenreif, L. M.; and Borth, D. 2021. Multitask Learning for Estimating Power Plant Greenhouse Gas Emissions from Satellite Imagery. In *NeurIPS 2021 Workshop on Tackling Climate Change with Machine Learning*.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. Las Vegas, NV, USA: IEEE. ISBN 978-1-4673-8851-1.

Long, J.; Shelhamer, E.; and Darrell, T. 2014. Fully Convolutional Networks for Semantic Segmentation.

Mooney, C.; Eilperin, J.; Butler, D.; Muyskens, J.; Narayanswamy, A.; and Ahmed, N. ???? Countries' climate pledges built on flawed data, Post investigation finds.

on Climate Change, I. P.; and Edenhofer, O., eds. 2014. *Climate change 2014: mitigation of climate change: Working Group III contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. New York, NY: Cambridge University Press. ISBN 978-1-107-05821-7 978-1-107-65481-5. OCLC: ocn892580682.

Raghuvanshi, S. P.; Chandra, A.; and Raghav, A. K. 2006. Carbon dioxide emissions from coal based power generation in India. *Energy Conversion and Management*, 47(4): 427–441.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv:1505.04597 [cs]*. ArXiv: 1505.04597.

Simonyan, K.; and Zisserman, A. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. ArXiv:1409.1556 [cs].

Smol, J. P. 2012. Climate Change: A planet in flux. *Nature*, 483(7387): S12–S15. Number: 7387 Publisher: Nature Publishing Group.

Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; and Wojna, Z. 2016. Rethinking the Inception Architecture for Computer Vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2818–2826. Las Vegas, NV, USA: IEEE. ISBN 978-1-4673-8851-1.

US EPA, O. 2016. EMC: Continuous Emission Monitoring Systems.

Verena Graichen, S. G., Johanna Cludius. 2019. European Union Emissions Trading System (EU ETS) data from EUTL — European Environment Agency.