SAND2022-16609C

Machine Learning Applications for Estimation of Greenhouse Gas Emissions Using Multiple Satellite Images

John Ringer (1,2), Hongkyu Yoon (1), Teeratorn Kadeethum (1), and Trevor Harris (3)

(1) Sandia National Laboratories, Albuquerque, NM (2) University of New Mexico, Albuquerque, NM, (3) Texas A&M University, College Station, TX

PRESENTED AT:

INTRODUCTION

Motivations

Estimating air pollution levels is a critical step towards mitigating climate change.

Current conventional methods rely on point-based or surface-based measurements, which are typically limited to cover large areas and are temporally sporadic.

Machine learning (ML) methods can be developed for accurately predicting air pollution levels from remote sensing data.

Dataset

The dataset consists of satellite data taken from the ESA Copernicus missions Sentinel-2 (S2) and Sentinel-S5P satellites, as well as ground measurements of NO_2 levels at 3,000 locations across Europe over the 2018-2020 (Scheibenreif et al., 2021).

- S2 satellite collects high-resolution, multi-spectral images.
- S5P satellite observes gas concentrations in the atmosphere.
- NO_2 ground measurements (microgram/m^3) from the European Environmental Agency's collection of air quality monitoring stations.

Each S2/S5P image covers 1.2x1.2 km region

NO_2 measurements are considered average concentrations for this area during the timeframe under consideration.

Both the S2 and S5P images are processed so that the relevant ground station is centered (Scheibenreif et al, 2021b).

ML MODEL ARCHITECTURES

Various deep learning models to predict NO_2 levels from satellite images.

Static Model

Model consists of 2 CNN architectures (+ dense layer) followed by a "head"

Head consists of a dense layer followed by an activation function, followed by a final dense layer for generating a single NO_2 concentration prediction for the location.

Output is an NO_2 prediction for average concentration over the entire timespan.

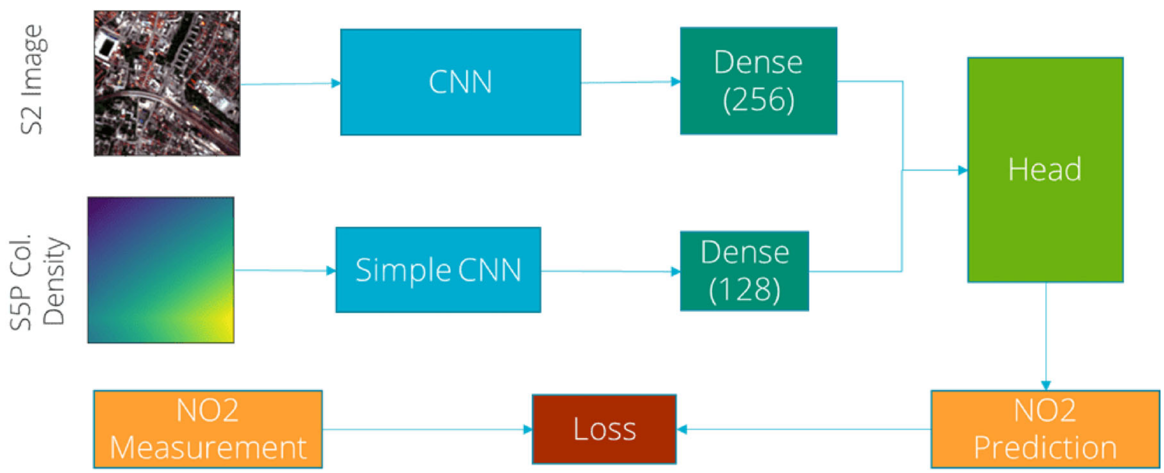

Figure 1

Overview of Static Model

TGLSTM Model

Given that greenhouse gas emissions vary temporarily, it is important to evaluate models against time-series data.

We implemented a model using a Time-Gated LSTM (TGLSTM) to investigate the use of timeseries information.

Monthly S5P data is organized into a timeseries

Output is a vector of NO_2 values, each value a prediction for average concentration over a single month.

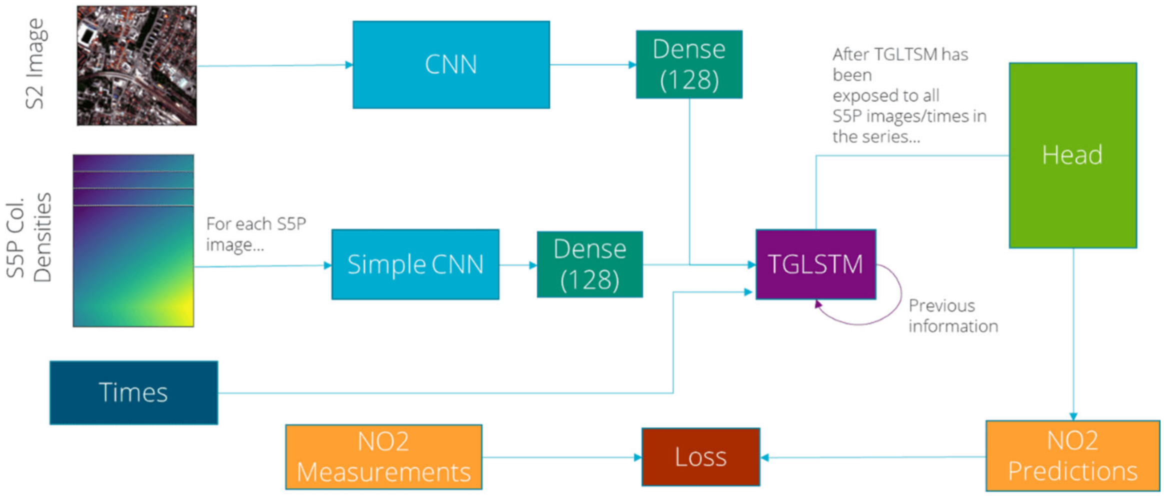

Figure 2

Overview of TGLSTM Model.

*Note: The TGLSTM is used instead of an LSTM layer because the TGLSTM can account for nonuniformly sampled timeseries data (i.e., data where the time between observations is not constant). This is done by incorporating time information as a nonlinear scaling factor using additional time gates (Sahin & Kozat, 2018).

TGLSTM + FFN Model

Spatial information was hypothesized to improve performance due to the differences in seasonal fluctuations between distant regions.

Coordinate information is fed to a feed-forward neural network (FFN) whose output is concatenated to the output tensor of the TGLSTM.

Output is a vector of NO_2 values, each value a prediction for average concentration over a single month.

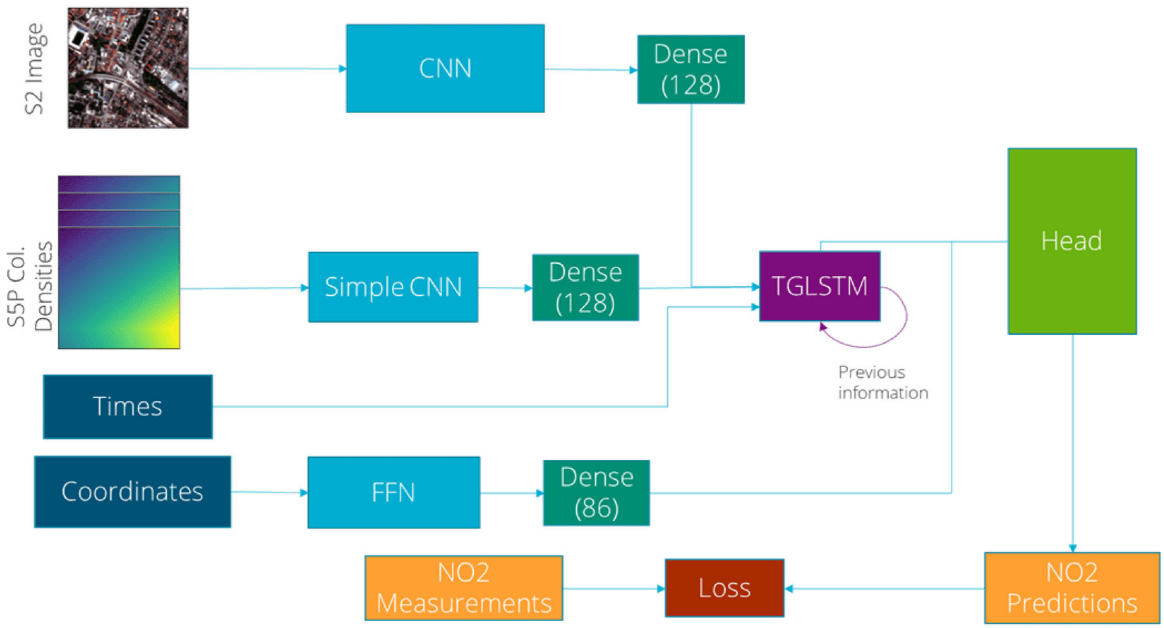

Figure 3

Overvew of TGLSTM+FFN Model

Model Complexity

Within this work we investigated using three different S2 backbone architectures.

Model Type	S2 Backbone	# Trainable Params
Static Model	Resnet-50	24,476,378
Static Model	U-Net Encoder	24,009,690
Static Model	Efficientnet-b1	7,290,010
TGLSTM	Efficientnet-b1	7,475,930
TGLSTM+FFN	Efficientnet-b1	7,505,200

Table 1

Overview of number of trainable parameters for each model.

*Note 1: For TGLSTM and TGLSTM+FFN models we present only models using an Efficient-b1 backbone as this backbone was found to give the best performance with Static Models (see Table 3).

*Note 2: EfficientNet-b1 was chosen over other EfficientNet models as it was found to give high performance while using a relatively small number of parameters.

ML TRAINING/RESULTS

Model Training

Each model with a random seed for weight initialization is trained for 3 runs using the hyperparameters described by Table 2.

Models are trained for 50 epochs. The model weights that are saved and used for evaluation are those that gave the lowest validation loss during training.

Optimizer	Adam
Learning rate	1e-3 (Adam default)
Loss Function	MSE Loss
Runs	3
Epochs	50
Batch Size	96
Training/Validation/Test Split	0.8/0.1/0.1

Table 2
Overview of hyperparameters used for all models

*Note: Each run uses a different random seed. The random seed determines how model weights are initialized, the way data is shuffled into each split (training, validation, test), and the random augmentations used during training.

Quantitative Results

The performance metrics:

- r^2 –coefficient of determination between NO_2 predictions and actual ground measurements.
- MAE – Mean absolute error (mg/m^3)
- MSE – Mean squared error
- X-Mean – average of scores for X metric over 3 runs.
- X-T3 – Best score for X metric over 3 runs.

Model Type	S2 Backbone	r^2 Mean	r^2 T3	MAE Mean	MAE T3	MSE Mean	MSE T3
Static Model	Resnet-50	0.5853	0.6292	5.5132	5.3971	55.3456	54.1262
Static Model	U-Net Encoder	0.5649	0.6336	5.6825	5.4384	57.6435	53.1247
Static Model	Efficientnet-b1	0.5995	0.6248	**5.4277**	5.1953	**53.8364**	**49.4701**
TGLSTM	Efficientnet-b1	0.5814	0.6021	6.3303	6.0987	76.4028	67.4023
TGLSTM+FFN	Efficientnet-b1	**0.6280**	**0.6945**	5.8380	**5.1807**	67.7610	51.7574

Table 3

Overview of model results. Shown values come from the testing split of each dataset.

For reference, statistics for both the averaged data and monthly data are provided below.

Time	NO_2 Mean (micrograms/m^3)	NO_2 Std (micrograms/m^3)
2018-2020	20.97	11.57
monthly	20.52	12.71

Table 4
Relevant metrics for each dataset.

*Note: There is inherently greater difficulty in predicting NO_2 concentrations on a monthly basis (rather than average emissions over the entire timespan) due to fluctuations in emissions from month to month. TGLSTM Model performance may have also been limited by missing information (certain stations did not report measurements for all months).

Qualitative Results

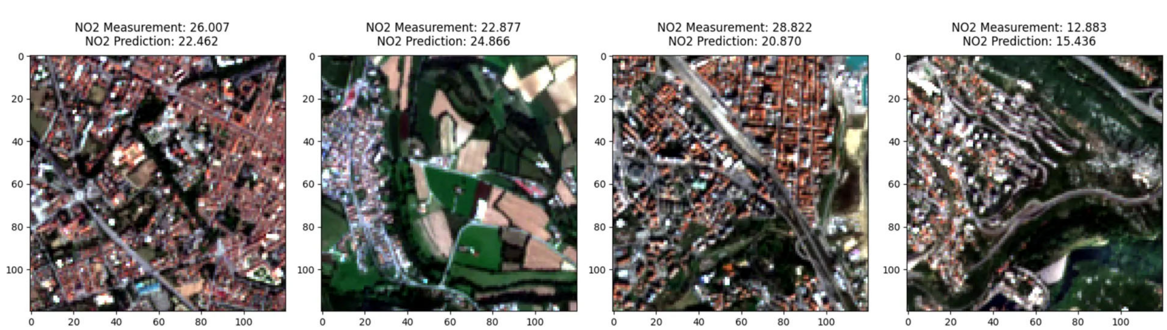

Figure 4

Sample predictions for Static Model w/ EfficientNet-b1 backbone. Images shown are from S2 satellite data

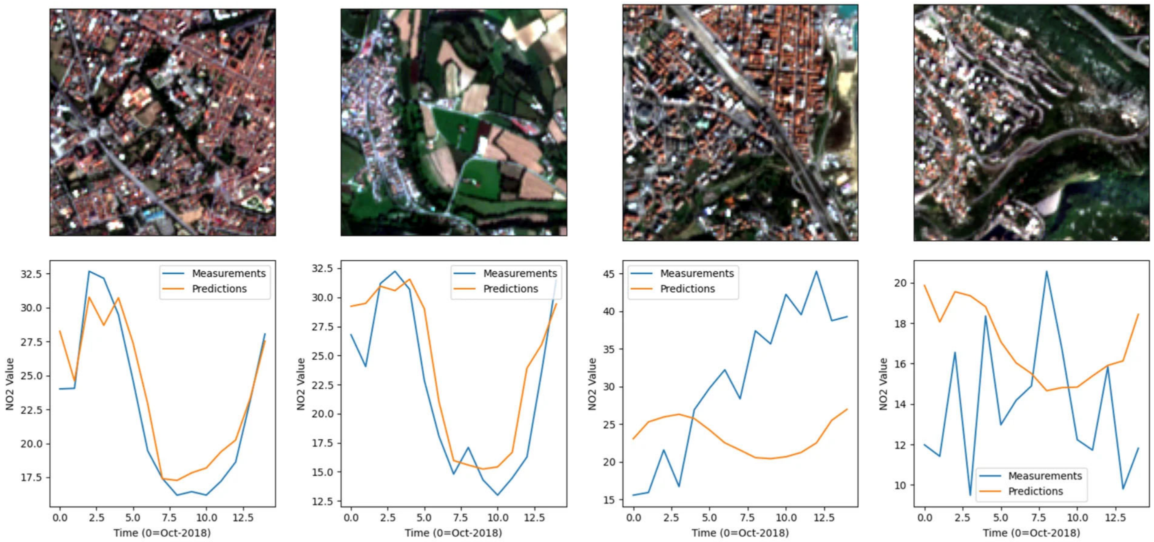

Figure 5

Sample predictions for TGLSTM+FFN Model w/ EfficientNet-b1 backbone. Two quality predictions are shown (left) and two poor predictions (right).

DISCUSSION

Uncertainty Quantification

Quantify uncertainty with conformal inference

• Generates well-calibrated prediction intervals for any prediction algorithm

• Only requires a training and test dataset

• Compute test residuals $r_i=|y_i - \hat{y}_i|$ for $i \in 1,\dots,N_{test}$

• Prediction interval $\hat{y} \pm d_\alpha$ where $d_\alpha=\text{Quantile}(1-\alpha, r_1,\dots,r_{N_{test}})$

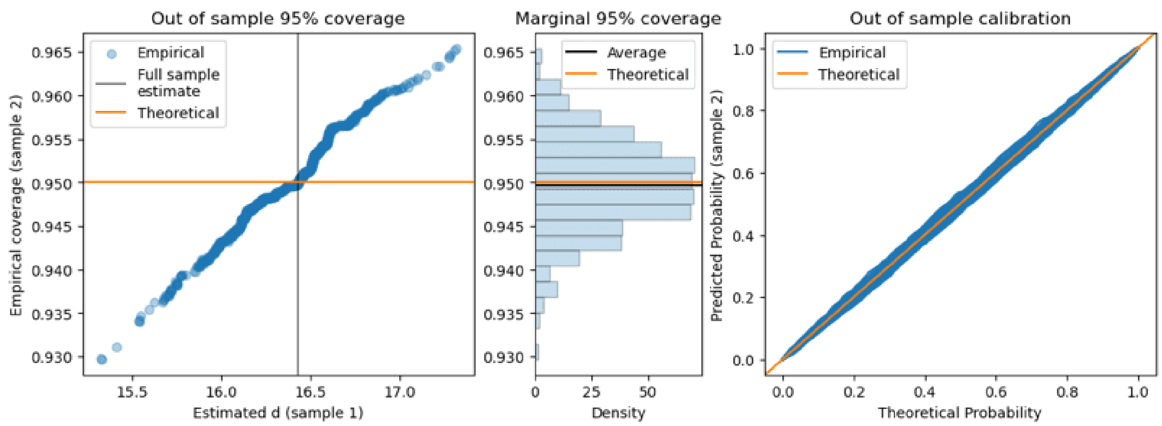

Figure 6

Left – Estimated $d_{0.05}$ against out of sample coverage. Concentrates around the target 95%.

Right – Out of sample calibration. Blue lines are estimates. Orange line is perfect calibration.

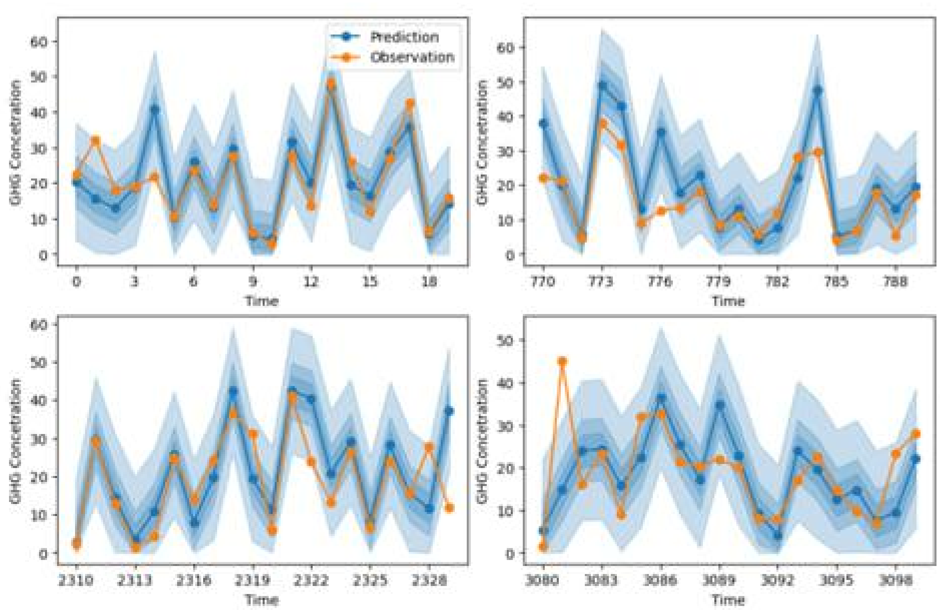

Figure 7

Prediction interval with measured values (orange)

Results Discussion

The Static Model using an Efficient-b1 S2 backbone consistently achieved higher or similar performance when compared to Static Models using S2 backbones with far more parameters.

The TGLSTM+FFN model achieved high-performance, being the only model to have an average r^2 higher than 0.6.

Considering the improvement over the TGLSTM model, spatial coordinates seem to provide relevant information for predicting NO_2 concentrations for a region.

TGLSTM models performed worse when concentration profiles had monthly fluctuation rather than seasonal fluctuations (see Figure 5).

Conclusions

Newly developed deep learning models in this work improved prediction accuracy of spatio-temporal variations of NO_2 concentrations based on two different satellite images and additional information such as time of monthly data and coordinates of ground stations.

We also improve model performance by incorporating architectures that use a far smaller number of parameters.

This research will lead to a new groundwork for estimating greenhouse gas concentrations using remote sensing data including satellite data, which will enhance our capability of tracking the cause of climate change and developing mitigation strategies.

Acknowledgment

This work was supported by Laboratory Directed Research & Development project at Sandia National Laboratories.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

ABSTRACT

Estimation of greenhouse gases emissions is a critical step towards mitigating climate change. Despite the importance of monitoring and limiting these emissions, existing models to estimate greenhouse gases emission and distribution rely heavily on land-use datasets that are both spatially and temporally restricted. To overcome this restriction, combining ground-based measurements with remote sensing data such as satellite images researchers have shown promising outcomes to estimate the distribution of greenhouse gas concentrations as well as emissions. In this work we investigate the use of deep learning (DL) models for the prediction of greenhouse gas distribution over time using satellite images taken from the ESA Copernicus missions Sentinel-2 and Sentinel-S5P satellites where three-channel aerial images were processed from Sentinel-2 and Sentinel-S5P provides a depth-average of NO_2 concentrations. We explore multiple deep learning model architectures to account for multi-modal input data and time-series data. Preliminary results show that DL model predictions based on relatively simple neural network architectures and recurrent neural networks (e.g., LSTM) attain a high degree of accuracy (mean absolute error <6 microgram/m^3) against measurements from ground stations. In addition, the coordinate of ground stations helped improving predictability of NO_2 concentrations over time. Finally, we use last layer Laplace Approximations for efficient post-hoc uncertainty quantification (UQ) and demonstrate the UQ of estimated NO_2 concentrations and other types of greenhouse gas concentrations and/or other locations via transfer learning.

SNL is managed and operated by NTESS under DOE NNSA contract DE-NA0003525.

REFERENCES

Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., et al. (2023). Tackling Climate Change with Machine Learning. ACM Computing Surveys, 55(2), 1–96. https://doi.org/10.1145/3485128

Sahin, S. O., & Kozat, S. S. (2018). Nonuniformly Sampled Data Processing Using LSTM Networks. IEEE Transactions on Neural Networks and Learning Systems, 30(5), 1452–1461. https://doi.org/10.1109/tnnls.2018.2869822

Scheibenreif, L. M., Mommert, M., & Borth, D. (2021a, July 23). Estimation of Air Pollution with Remote Sensing Data: Revealing Greenhouse Gas Emissions from Space. Retrieved July 27, 2022, from https://www.climatechange.ai/papers/icml2021/23

Scheibenreif, L., Mommert, M., & Borth, D. (2021b July 23). Estimation of Air Pollution with Remote Sensing Data: Revealing Greenhouse Gas Emissions from Space (Dataset). Retrieved July 27, 2022, from https://zenodo.org/record/5764262#.YfJiPS1XYUs