A Machine Learning Approach to Spatiotemporal Emission Modelling

Kelly Zheng University of Waterloo, ON, N2L 3G1, Canada
Roydon Fraser University of Waterloo, ON, N2L 3G1, Canada
Jesse Thé Lakes Research Inc., ON, N2L 3L3, Canada
Email: {k7zheng, rafraser}@uwaterloo.ca, jesse.the@weblakes.com ,

Abstract

Despite the growing impact of emissions on health and environment, there remains an unmet need for emission concentration prediction and forecasting. The accumulating monitoring station and satellite data available makes the problem well suited for machine learning. This work formulates the spatiotemporal prediction of emission concentration as a machine learning task. To this end, an evaluation framework including baseline models and metrics of per-pixel loss and intersection over union accuracy, as well as a simple spatiotemporal ConvLSTM machine learning model were developed. The ConvLSTM model successfully generates one-hour ahead emission concentration forecasts with increasingly lower loss (6.5% and 30.5% less) and higher accuracy (18.4% and 18.6% higher) compared to the input-independent and random baseline models at the end of training. Crucially, compared to commonly used, physics-based models for emission monitoring, the model generalizes to unseen emission sources with no significant decrease in accuracy.

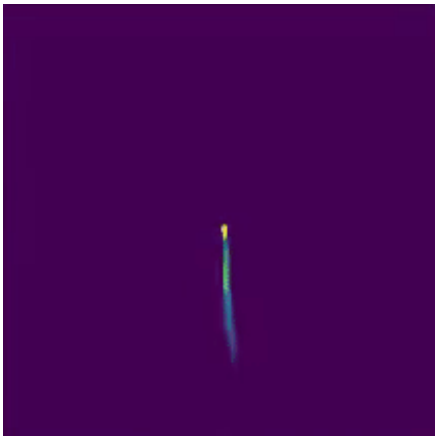

Fig. 1: Animated emissions from a point source.

1 Introduction

Despite the widespread and growing impact of emissions on our health and environment, there remains an unmet need for emission concentration models. Every year, more than seven million people die prematurely due to air pollution [1], and one in five people die of air pollution from greenhouse gas emissions [2]. Yet every year, our emissions continue to grow [3].

Conventional emission models represent the spatial and temporal variation of substances such as greenhouse gases (GHG) emitted either at the emission source or after release [4]. At the source, an accounting of all emissions is done to conduct emission inventories. The emission inventory consists of data reported by emitters in accordance with regulations using prescribed, sector-specific calculation methods inducing the emission factor, a number that relates the quantity of emissions from a source to units of activity associated with emission release. However, these factors are typically averages of available data and result in emission inventories that are mostly useful as a knowingly gross estimation of emission trends [5, 6]. After release, air dispersion models such as Gaussian plume models AERMOD or CALPUFF are used to simulate the mixture of an emission with the atmosphere. These models characterize key processes that control the dispersion and may be used to predict the concentration further in space and time. However, these models are limited in their lack of generalizability as well as required computational time and resources [7–9].

2 Methodology

This work focuses on formulating the emission modelling problem as a machine learning problem and setting up a foundation where increasingly complex models can be built upon. To evaluate feasibility and properly scope the proposed research, this paper first formulates the problem. Foundational preliminary work then consisted of two tasks: preparing the data used to develop models with and developing the baselines and metrics used to evaluate future models on. Finally, a simple classical ConvLSTM model is developed and deployed. The results are used to demonstrate the developed evaluation framework in action and to do a crude comparison of the model to the baselines.

2.1 Problem Formulation

To scope, the modelling problem focus was selected as forecasting emission concentration from point sources, as shown above in Figure 1, chosen due to the ease of data collection and the fact that they account for 70% of all emissions [3]. The developed model was required to generate one hour ahead predictions of emission concentration, given an input sequence of past six-hour emission concentrations from the same point source. The forecast would be made with the assumption of no significant change in land use or buildings.

2.2 Data

Data for preliminary work was provided by Lakes Environmental Research and included concentration data from various point sources. Concentration was measured at 10m above the ground within a square geofenced region around various point sources. Concentration values were normalized and centered within a 164x164 matrix. The dimension of each datapoint was (162, 162, 1): 162 pixels in latitude, 162 pixels in longitude, and each point with one emission concentration measurement. The number of datapoints in each dataset was 35,040 datapoints with one datapoint every 15 minutes generated over one year.

2.3 Baselines

Baselines developed were simple benchmarks that serve as meaningful reference points to compare the results of solutions to the toy problem developed against. By comparing solutions to baseline outputs, the impact of changes to the models developed and changes in data fed to the models on the predictions generated can be evaluated. Two baselines were developed: a random chance benchmark which corresponds to filling a matrix with randomly generated values and a zero-rule benchmark which corresponds to generating a zero-matrix, the majority value of concentration as there are generally no emissions except where the plume is located.

2.4 Metrics

In order to evaluate the models against the baselines developed, metrics were chosen to quantify their losses and prediction accuracy. During training, a loss function that measures the difference between the pixel values of predicted and true images was first created. Two standard metrics used to do so are per-pixel loss functions using mean absolute error and mean square error. A per-pixel loss function quantifies the total of all errors between the exact pixel in each image. This is generally a calculation of mean absolute error, which is also known as L1 loss.

Instead of mean absolute error, this work uses mean square error or L2 loss. This simply involves squaring the error instead of taking the absolute value. Because mean square error has mathematical properties which makes it easier to calculate gradients and backpropagate error [10]. The equation used for mean square error

for the per-pixel loss function, is shown below in Eq. 1.

$$loss(y,\hat{y}) = \frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2 \quad (1)$$

There are several drawbacks to using per-pixel loss functions. In summing the error of each pixel, content similarities are ignored. In future work, a more refined loss function will be created that primarily captures high-level differences between image content. For example, where a per-pixel loss function would return a large error for two images that are identical except for being shifted by one pixel, a perceptual loss function could return a small to zero amount of error [11]. However, for this preliminary work, a differentiable and quickly calculated indication of accuracy was selected. The per-pixel L2 loss function is a good foundational metric that can later be modified to add additional terms like perceptual loss or log-loss that consider predicted probabilities to.

Model accuracy is a measure of how similar images of the predicted plumes generated by the trained models are to images of the true plume. The first indicator that could be used is the final per-pixel loss, as defined above using either L1 or L2 loss. However, because the accuracy will be evaluated on the trained models, accuracy metrics can be developed that are not differentiable and better capture high-level differences between image content. As accuracy in this application is human-interpretable, one measure of accuracy that could use is simply the evaluated human accuracy. Manually looking at each predicted plume image overlayed on a true plume image can be a check of how well the two images line up. However, interpretation needs to be scaled up in order to evaluate the accuracy of model predictions deployed on a large amount of test data. One metric that captures some of this interpretation is intersection over union (IoU) shown below in Eq. 2.

$$IoU(y,\hat{y}) = \frac{y \cap \hat{y}}{y \cup \hat{y}} \quad (2)$$

$IoU \in [0,1]$ where if y and $\hat{y}$ are identical images, their calculated IoU would be 1. To define the plume areas of intersection and union in the predicted and true images, pixel values in both of images are converted to True or False based on a threshold of 0, then take the logical and as well as the logical or for the intersection and union, respectively. The total IoU calculation for a test dataset then records the average IoU values across all predicted and true images.

2.5 Models

In this preliminary work, a simple classical ConvLSTM model [12] was developed consisting of two ConvLSTM2D layers with batch normalization, followed by a Conv3D layer for the spatiotemporal outputs. The ConvLSTM network was chosen as a simple first model to check the correctness of the training and evaluation skeleton. The ConvLSTM developed was trained on the emission concentration dataset for the tasks of concentration prediction 1 and 2 hours ahead. The predictions were compared to the baselines using the metrics. The deployment of the model on the concentration prediction task is described below.

3 Results

The ConvLSTM model took inputs of 162x162x1 resolution in a six-hour window. The model then outputs the next hour prediction. Figure 2 shows the results of the trained model used to predict 1-hour ahead forecast for concentration.

3.1 Loss

Figure 3 shows the average mean squared error loss of the model predictions during training plotted against epoch. With each epoch, the model updates its internal parameters, so it is expected for loss to start high and trend downwards as parameters that better describe the data are selected. It is also expected for validation loss (orange) to be higher than training loss (blue) as the model is trained on the training dataset, then applied to the unseen validation dataset. In the loss vs. epoch plot, both training and validation loss trend downwards, with validation loss about 10-20% higher than training loss. Overall, Figure 3 suggests that despite a large amount of loss, the model is behaving as desired in terms of overfitting.

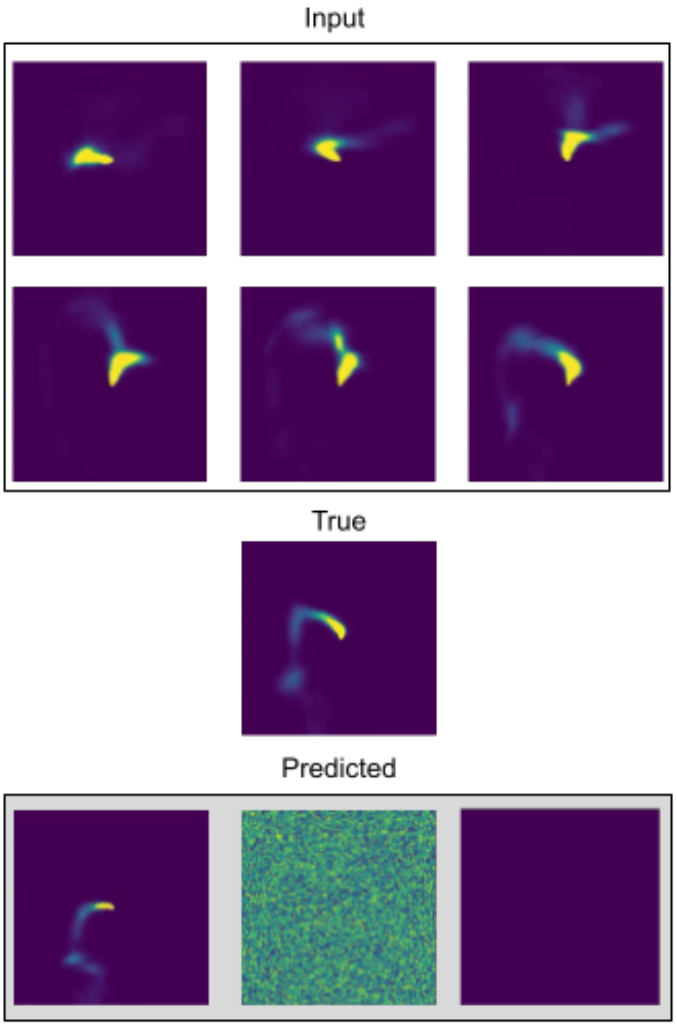

Fig. 2: Input and predicted concentration plumes using the simple ConvLSTM model. The first six frames are the input sequence. The middle frame is the true 1-hour ahead prediction. The last three frames are the ConvLSTM model as well as the baseline random model and input-independent model predictions respectively. Each frame is an hour time-step.

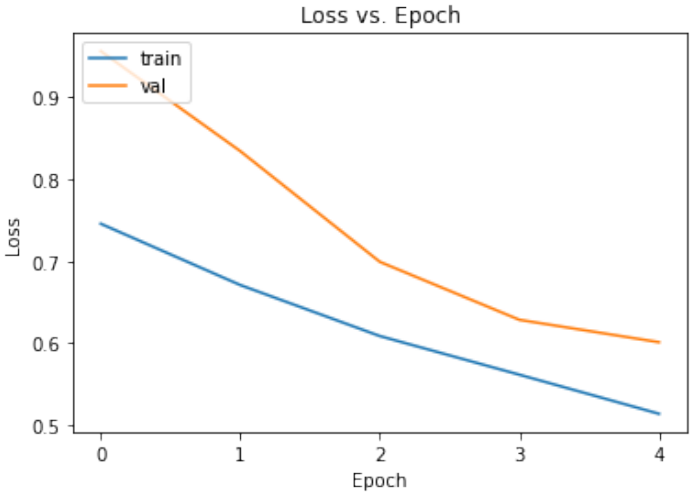

Fig. 3: Mean squared error loss plotted against epoch. Training loss is in blue while validation loss is in orange.

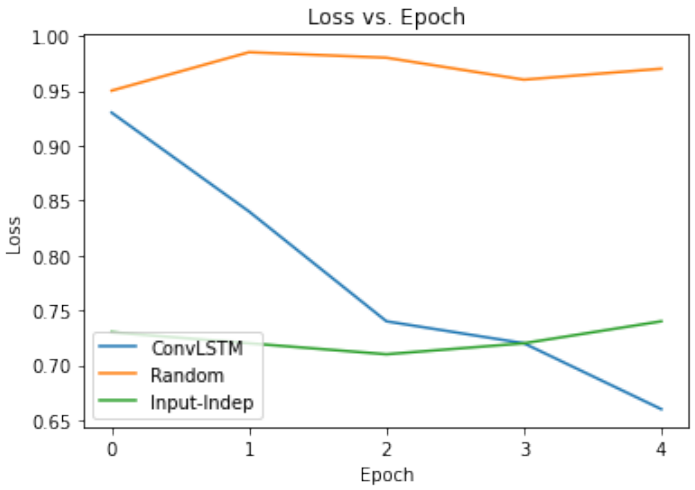

Fig. 4: Comparison of mean squared error loss plotted against epoch for all models run on the test dataset.

3.2 Accuracy

Accuracy was calculated using the intersection over union method for the ConvLSTM model 1-hour forecasts using the validation dataset. The average IOU was calcuated as 13.7% with most predictions under 20%. As an IOU of 100% would indicate two identical images, the current IOU average of 13.7% is low and indicates that the plumes images predicted using the ConvLSTM model are dissimilar from the true plume images. This is expected, as a very simple ConvLSTM model architecture was used with only two layers and there was little training of the model with only five epochs. The lack of accuracy is also supported by the high loss of the model, which was around 60% at the end of training for the validation test set, as seen before in Figure 3.

3.3 Comparison to Baselines

Though the classical ConvLSTM model was intended to serve mainly as a very simple test of the problem formulation and evaluation framework, having kept the training and validation hyperparameters the same, a crude comparison of the model to the baselines can be done. To do the comparison, the ConvLSTM model as well as the two baseline models (random and input-independent) were run on the reserved, unseen test dataset. The ConvLSTM model was then compared to the two baselines using the evaluation framework to calculate loss (Figure 4), accuracy (Table 1), and performance on the 1-hour ahead forecasts.

Table 1: Comparison of intersection of union values of the 1-hour ahead concentration forecasts using all model on the test dataset.

Model	**IoU** (%)
ConvLSTM	18.9
Random	0.2
Input-Independent	0.0

Overall, the ConvLSTM model outperformed both baselines in terms of loss and accuracy, which was expected. This was especially pronounced in the loss vs epoch plot seen in Figure 4 after a few epochs, as the ConvLSTM model loss decreased significantly (from 0.93 to 0.66 MSE) with training while the baselines remained the same (around 0.71-0.74 MSE for the input-independent baseline model and around 0.95 to 0.98 MSE for the random baseline model). Accuracy as measured by IoU showed similar outperformance of the baselines by the ConvLSTM model, with 18.6% IoU compared to 0.2% and effectively 0% IoU for the random and input-independent baselines.

4 Conclusion

This work formulates the task of emission concentration prediction as a machine learning task, describes the development of an evaluation framework including metrics of per-pixel loss and an intersection over union accuracy, as well as baseline models; and finally presents the training of a simple ConvLSTM model to demonstrate usage of the evaluation framework. The developed model successfully generates 1-hour ahead emission concentration forecasts with lower loss and higher accuracy compared to baselines. The ConvLSTM model successfully generated 1-hour ahead emission concentration forecasts with increasingly lower loss (6.5% and 30.5% less) and higher accuracy (18.4% and 18.6% higher) compared to the input-independent and random baseline models at the end of training. Comparison to the baseline models shows that the evaluation framework functions as expected and hints to the promise of models with greater complexity trained on richer data which might further improve predictions and generalize even better to unseen data. Future work increasing model and feature complexity is anticipated to further improve accuracy and generalization to unseen emission sources.

Acknowledgements

We would like to thank Natural Sciences and Engineering Research Council of Canada and Lakes Research for supporting this work.

References

[1] WHO, “Air pollution,” 2021. [Online]. Available: https://www.who.int/westernpacific/health-topics/air-pollution

[2] K. Vohra, A. Vodonos, J. Schwartz, E. A. Marais, M. P. Sulprizio, and L. J. Mickley, “Global mortality from outdoor fine particle pollution generated by fossil fuel combustion: Results from GEOS-Chem,” *Environmental Research*, vol. 195, p. 110754, Apr. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0013935121000487

[3] IPCC, “AR6 WGI Full Report,” Tech. Rep., 2021. [Online]. Available: https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Full_Report.pdf

[4] EPA, “Air Emissions Reporting Requirements (AERR),” 2018. [Online]. Available: https://www.ecfr.gov/current/title-40/chapter-I/subchapter-C/part-51

[5] J. Southerland, “Abridged History of Emission Inventory and Emission Factor Activities.” EPA, Apr. 2005, p. 18.

[6] S. Bromberg, “The Underappreciated Emission Inventory; Foundation for Pollution Control Decisions,” *EM*, pp. 17 – 20, Aug. 1997. [Online]. Available: https://pubs.awma.org/gsearch/em/1997/8/bromberg.pdf

[7] Exponent, “Official CALPUFF Modeling System,” 2020. [Online]. Available: http://www.src.com/

[8] EPA, “Air Quality Dispersion Modeling - Preferred and Recommended Models,” Nov. 2016. [Online]. Available: https://www.epa.gov/scram/air-quality-dispersion-modeling-preferred-and-recommended-models

[9] N. S. Holmes and L. Morawska, “A review of dispersion modelling and its application to the dispersion of particles: An overview of different dispersion models available,” *Atmospheric Environment*, vol. 40, no. 30, pp. 5902–5928, Sep. 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1352231006006339

[10] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986. [Online]. Available: http://www.nature.com/articles/323533a0

[11] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual Losses for Real-Time Style Transfer and Super-Resolution,” *arXiv:1603.08155 [cs]*, Mar. 2016, arXiv: 1603.08155. [Online]. Available: http://arxiv.org/abs/1603.08155

[12] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. WOO, “Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting,” in *Advances in Neural Information Processing Systems*, vol. 28. Curran Associates, Inc., 2015. [Online]. Available: https://proceedings.neurips.cc/paper/2015/hash/07563a3fe3bbe7e3ba84431ad9d055af-Abstract.html