Machine Learning Approaches for Air Quality Modeling Applications

Satish Vutukuru*, Alexander Cohan#, Fernando Garcia Menendez*

*North Carolina State University #Lake Michigan Air Directors Consortium

1. Introduction

Traditional air quality modeling is based on mechanistic approaches that simulate the underlying chemical and physical phenomena occurring in the atmosphere. However, in recent years, machine learning and statistical modeling approaches have undergone rapid advancement. So far there have been only limited efforts to apply the latest generation of machine learning techniques for air quality modeling applications.

We present preliminary results from our application of neural networks based deep learning architectures to forecast air quality in the Atlanta area. We use modeling output from a regional air quality model, domain-wide emissions, historical air quality, and meteorological data to train the model. We then predict air quality for a future period and evaluate the model performance.

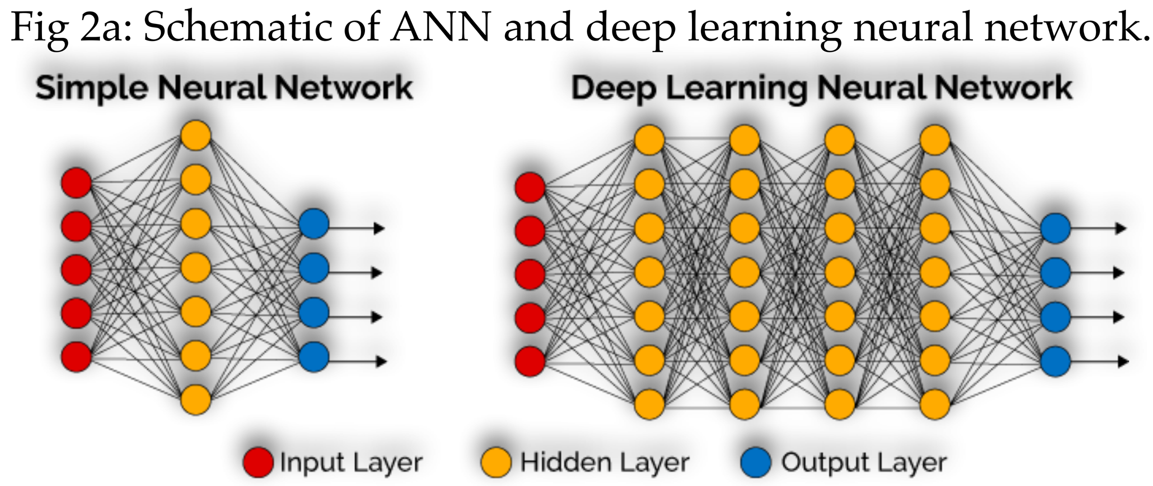
2. Artificial Neural Networks and Deep Learning

Artificial neural networks (ANN) is a machine learning paradigm for learning data representation through a series of non-linear data transformations. Recent advances in learning algorithms, such as backpropogation, led to the rise of *deep learning* which involves training a neural network containing many layers.

Deep learning overcomes *bias-variance* trade-off, as models can be improved by additional training data and/or adding more layers.

Two types of ANN: convolutional neural networks (CNN) that are useful for learning tasks on images and recurrent neural networks (RNN) for predicting sequences such as a time series.

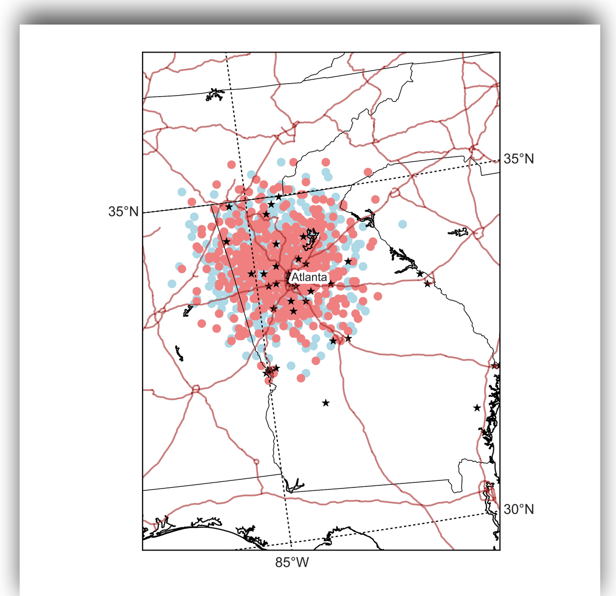
Fig 2a: Schematic of ANN and deep learning neural network.



3. Modeling Set-up

We used input and output data for CMAQ model to train an artificial neural network.

Fig 3a: Modeling Domain.



<u>Input:</u>
Meteorological parameters,
Local Emission parameters, and
Historical concentrations at monitors from past 24 hours

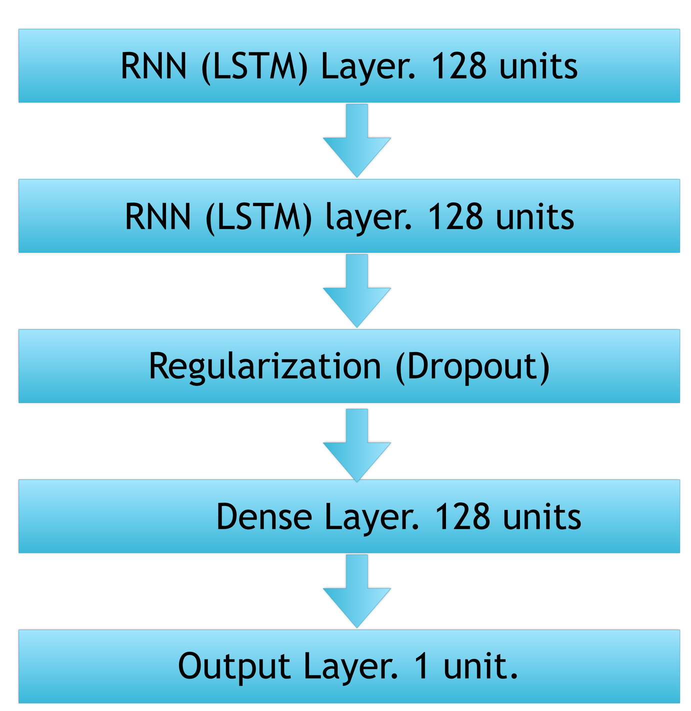
<u>Output:</u> Predicted air quality at the next hour.

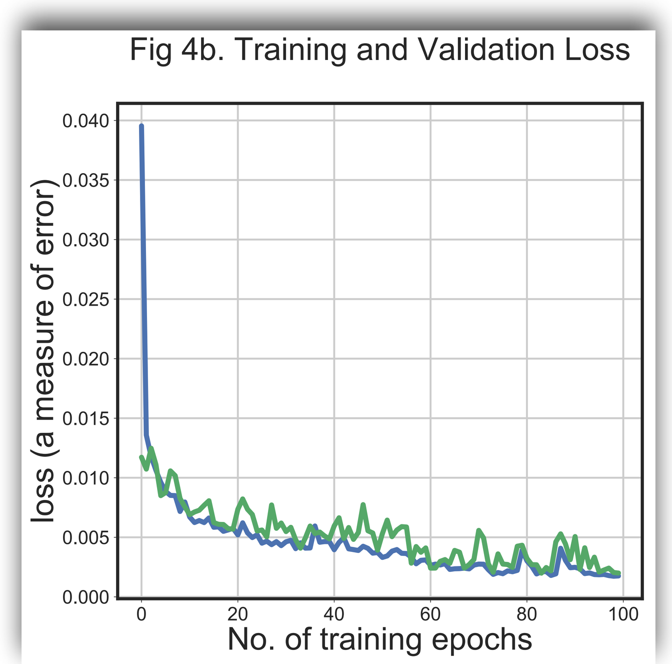
We developed three datasets to build and test the model:
a) *training dataset*: data from 180 grid cells to train the model
b) *validation dataset*: data from 20 grid cells to refine the model
c) *test dataset*: data from 100 grid cells to test the model. These data are from simulation output for days not used to either train or validate the model.

4. Deep Learning Network Training

We used a recurring neural network layers comprising of Long short-term memory (LSTM) units to build a deep learning model. The units used *tanh* activation functions. In order to prevent overfitting, a Dropout layer was used as a means of regularizing the network. Model was trained over 100 epochs using *adam* optimization scheme to learn the parameters.

Fig 4a: Schematic of ANN used to develop the air quality model





5. Sample Results

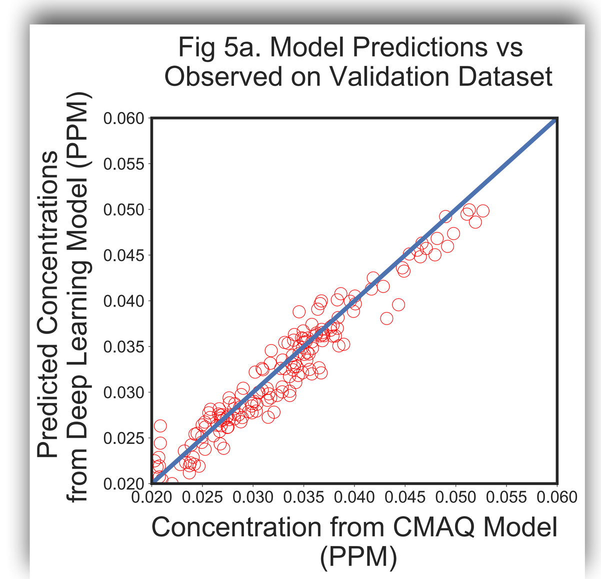


Fig 5a: The deep learning model was able to closely predict ozone concentrations for training and validation datasets. The root mean squared error (RMSE) was 0.002 PPM.

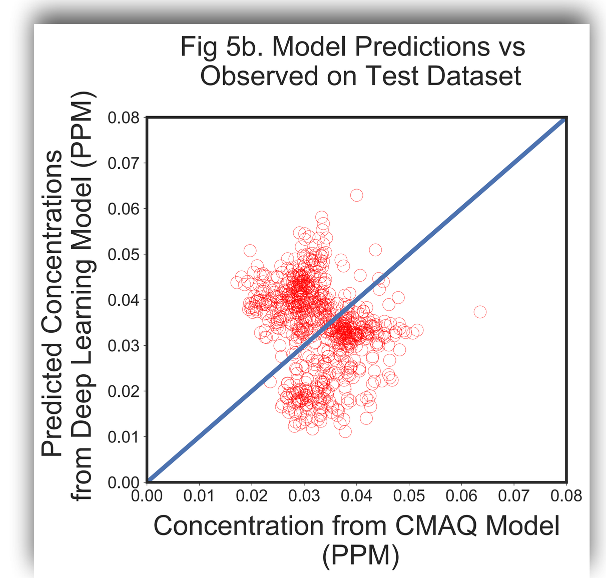


Fig 5b: When used to forecast air quality for future days, the model performed well although exhibited relatively high variance. This is attributed low number of days used for training data (3 days).

6. Discussion & Key takeaways

★ We demonstrated application of latest deep learning networks for air quality modeling.

★ A combination of Eulerian Models and Machine Learning models yield a broader toolset for modelers and regulatory agencies.

★ Some areas where machine learning is useful:

- Design of air quality monitoring network
- Data-assimilation and real-time forecasting.

References

- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.
- Zheng, Y., Liu, F., & Hsieh, H. P. (2013, August). U-Air: When urban air quality inference meets big data. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1436-1444). ACM

Acknowledgements: We thank Dr. Yongtao Hu and Dr. Talat Odman from GeorgiaTech for providing us with CMAQ simulation data for the Atlanta region.