Proceedings of the 8th World Congress on Civil, Structural, and Environmental Engineering (CSEE'23)
Lisbon, Portugal – March 29 – 31, 2023
Paper No. ICEPTP 198
DOI: 10.11159/iceptp23.198

Air Quality Prediction with a Combined Machine Learning and Chemical Transport Model

Xiaoxiao Feng[1,2], Xiaole Zhang[1,2], Jing Wang[1,2*]
[1]Institute of Environmental Engineering (IfU), ETH Zürich, Zürich, 8093, Switzerland
[2] Swiss Federal Laboratories for Materials Science and Technology, Dübendorf 8600, Switzerland
fengx@ethz.ch; zhanxiao@ethz.ch; jing.wang@ifu.baug.ethz.ch

Extended Abstract

Air quality forecast provides significant information and helps people plan ahead to reduce the health and economic risks associated with air pollution. However, the accurate air quality prediction is a difficult task to achieve. The air quality models reproduce the physical and chemical processes in atmosphere and result in the pollutant concentrations on the ground surface. However, the deficiency of the mechanisms and the uncertainties of the model input (emission inventories, meteorological information et al.) lead to considerable deviations from the observations. On the contrary, machine learning models do not aim at explaining the physical and chemical mechanism. They calculate the pollutant concentrations by training and learning the past phenomenon. The training samples are important, and it is uncertain that the past stories will be repeated in the future and cause the prediction biases.

In this study, we combined a traditional air quality model, the Community Multiscale Air Quality Modeling System (CMAQ), together with a type of neural network, long short-term memory network (LSTM), to predict the future concentrations of the fine particle matter ($PM_{2.5}$) in Beijing, China. The new algorithm can improve the prediction accuracy of CMAQ and that of using LSTM independently. When only running the CMAQ model, CMAQ can capture the variations of the temporal $PM_{2.5}$, but there were still overestimations and underestimations over time. The root-mean square error (RMSE) of CMAQ results was around 40 compared with the ground observations. When predicting $PM_{2.5}$ concentrations 2 hours, 12 hours, 24 hours and 48 hours ahead of time with the combined system, the RMSEs were 4.4, 10.2, 14.4 and 19.6 respectively, lower than that of CMAQ and within the goal and criteria of modelling $PM_{2.5}$ recommended by US Environmental Protection Agency (US EPA) [1].

The combined system took advantages from both CMAQ model and LSTM model. For different prediction hours, the prediction with the combined system was more accurate than the single LSTM without CMAQ inputs, and the errors of prediction were smaller than the single LSTM. The difference of RMSE, mean fraction bias (MFB) and mean fraction error (MFE) between the combined system and the single LSTM grew larger when making a longer prediction of $PM_{2.5}$ concentrations. The combined system benefited from the accuracy of LSTM for the short time prediction and the stable performance of CMAQ for the long time prediction.

References

[1] US EPA. Modelling Guidance for Demonstrating Air Quality Goals for Ozone, PM2. 5 and Regional Haze. 2018.