

Kim et al. developed LSTM-based PM10/PM2.5 prediction model and showed better performance than CTM. It is surprising that PM2.5/PM10 is more accurately predicted by a machine learning-based model using a few variables than by a complicated 3-dimensional chemical transport model incorporating emission, chemical production/loss, transport and diffusion, wet/dry deposition processes. Especially, this paper is meaningful that machine learning is applied to PM10/PM2.5 forecasting in Korea where the terrain is very complicated, and PM10/PM2.5 comes from diverse sources and is frequently influenced from continental pollution. I have only a few points for better clarification.

1. Authors selected variables for machine learning using their knowledges and experiences. However, the square of the pearson correlation coefficient ( $R^2$ ) in Fig 3. and 4 looks not greater than 0.5. meaning that the input variables have only 50% of explanatory power. Can this not limit the performance of machine learning based model?
2. In major cities in Korea, NO<sub>2</sub> and CO are likely to be correlated due to share the common emission source. Does the dependency between input variables worsen LSTM performance or have little effect on it?
3. The high pollution events of PM10/PM2.5 in Korea are usually caused by long-range transport(LRT) and atmospheric congestion(AC). In most cases both LRT and AC play a role sequentially in polluted days. However, LSTM showed poor prediction at LRT case of May 25 to 28, 2016. Did authors consider any other model or any combination of LSTM and CNN(or DNN) in order to capture both LRT and AC?
4. Air quality forecasting is usually intended for high pollution events. Did authors consider to estimate the LSTM by categorical statistics such as critical success index(CSI), probability of detection(POD), false alarm ratio(FAR), and etc? If then, as high pollution events are not frequent, did authors consider the issue of data imbalance between normal and polluted days?
5. Several things such as data representation, activation function, weight initialization, pre-processing, hyper parameter are important for

determining machine learning model. I believe that authors performed a number of test to find the optimal method. Did authors not present for any reason all the information about them?

6. Correction of missing data is very important, especially, in machine learning algorithm. Authors developed the pre-trained deep LSTM model in order to generate missing data. As a result, the performance of the pre-trained deep LSTM model varies considerably with pollutant species. Does this affect the low dependance of SO<sub>2</sub> and NO<sub>2</sub> on PM<sub>10</sub>/PM<sub>2.5</sub> prediction or not?