Response to reviewer's comments:

First of all, thank you for your valuable comments and suggestions. In the revised manuscript, we attempt to improve the manuscript based on your comments and suggestions. The added/modified parts are painted in a red color in the revised manuscript. Here, we would like to reply to some specific comments raised by you below:

Author's comments to anonymous referee #2

1. "*Authors selected variables for machine learning using their knowledges and experiences. However, the square of the pearson correlation coefficient (R2) 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?*"

Reply) The predication accuracy of deep neural network models is, in general, known to be very high. The performances of these models are mainly determined by input data used in the model training. We organized the current data set with 11 to 12 independent variables that were all information that could be collected from the ground-based observations (i.e., AIR KOREA and KMA AWS networks). This indicates that we used almost all chemical and meteorological variables available from the observations. In the model training, we used the observations for a period of 2.3 years because there was limited data availability. We expect that the performance of the LSTM-based PM prediction model would improve if more independent variables were obtained from ground observations and longer time-series observations were utilized in the model optimization in the future.

2. "*In major cities in Korea, NO_2 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?.*"

Reply) We added more detailed discussions about the multicollinearity issue in the revised manuscript (please, see pp. 5:27-6:4).

3. "*The high pollution events of PM_{10}/$PM_{2.5}$ 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.*"

Reply) We may be able to improve the performances of the LSTM-based PM prediction model by combining different types or methods of neural network model that can predict high pollution events more accurately. To develop these models (or methods), it is essential to identify high PM episode events and collect more amounts of variables, but these preliminary studies require considerable time. One example is the balancing the data for better predictions of high PM events. This issue is discussed in reply 4. However, this was not working very well. We think this issue will be able to be covered by future work!!

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?.*"
Reply) We added a more detailed discussion on data imbalance in the revised manuscript (please, refer to pp. 11:3-18).

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?*"
Reply) We carried out several pre-tests to find out the optimized structure of the deep LSTM model. Recent deep learning studies have not provided detailed information about determining model structure because such descriptions must be extensive. In addition, the structure of deep neural network should change according to the configuration of independent and dependent variables. Therefore, we did not describe these contents in the manuscript.

The results of important sensitivity tests to determine the structure of $PM_{2.5}$ prediction model for Seoul-1 site are presented in Fig. R1. As shown in Fig. R1, the validation cost of the LSTM model training was the lowest when there were 100 hidden nodes (i.e., hidden neurons) and 5 hidden layers. In addition, the deep LSTM model showed optimal performances, when ReLU was embedded as an activation function, which is similar to previous studies (Nair and Hinton, 2010). Recent studies rarely used the sigmoid function, because of the gradient vanishing problem. For weight initialization, we applied the Xavier algorithm. This initialization method finds the optimal initial weight vectors according to the structure of the deep neural network (Glorot and Bengio, 2010). Because we adopted ADAM as an optimizer, the learning rate, which determines adjustment rate of weight and bias, continuously changed to find the global minima (Kingma and Ba, 2015).

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 SO2 and NO2 on PM10/PM2.5 prediction or not??"

Reply) As we described in Sec. 2.3, one of the main criteria in selecting the PM prediction sites was the number of missing observations. The percentage of missing observations at seven sites is summarized in Table S1. As shown in Table S1, the fractions of missing observations are relatively small. Therefore, the values generated by the pre-trained model are unlikely to affect the dependencies of atmospheric pollutants. In order to confirm this, we performed the LSTM model training without missing observations. The dependence of the independent variables on the PM predictions of the previous model (including the missing values generated by the pre-trained model) and the newly trained model (excluding the missing observations) is summarized in Table R1. As shown in Table R1, the dependencies of SO_2 and NO_2 were also low, although the missing observations were not considered in the model training. In addition, we compared the performances of both models to evaluate the effects of missing observations on the PM predictions. The prediction accuracy of the two models is summarized in Table R2. In general, the PM predictions by the previous model were superior to those by the newly trained model, except at Gwangju site. This is because the missing observations generated by the pre-trained model enabled us to train the LSTM model for more various atmospheric conditions.

References

Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the 13th International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 13-15 May, 249-256, 2010.

Kingma, D. and Ba, J.: Adam: A method for stochastic optimization, in: Proceedings of the 3rd International Conference on Learning Representations, San Diego, USA, 3-8 May, arXiv:1412.6980v9, 2015.

Nair, V. and Hinton, G. E.: Rectified linear units improve restricted Boltzmann machines, in: Proceedings of the 27th International Conference on Machine Learning, Haifa, Israel, 21-24 June, 432, 2010.

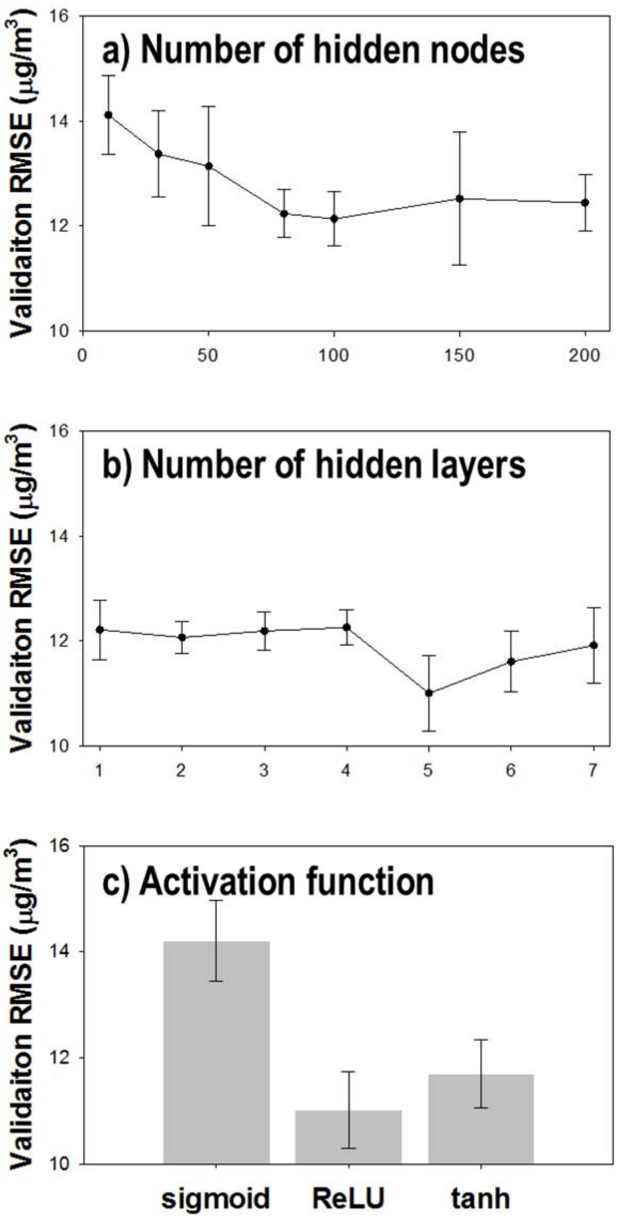

Figure R1. Results of sensitivity test to determine the structure of $PM_{2.5}$ prediction model for Seoul-1 site.

Table R1. Dependency comparison between the LSTM model with and without considerations of missing observations[1)]

Species	Station	Model	Input variable											
			TA	WD	WS	RN	RNH	RH	SO_2	O_3	NO_2	CO	PM_{10}	$PM_{2.5}$
PM_{10}	Seoul-1	LSTM w/ MO	-26.93	18.78	-1.44	0.20	-0.27	2.65	12.92	12.47	15.64	-0.35	42.98	-
		LSTM w/o MO	-19.25	16.98	4.30	0.48	-0.13	2.81	12.58	11.84	16.35	2.22	36.75	-
	Seoul-2	LSTM w/ MO	-22.07	32.29	10.05	0.22	-0.14	2.91	-0.87	4.74	12.42	-5.70	47.55	-
		LSTM w/o MO	-32.59	31.08	3.43	0.32	-0.25	3.79	1.34	10.07	18.54	-2.05	49.41	-
	Daejeon	LSTM w/ MO	-24.17	45.67	-9.97	0.23	-0.25	11.72	-0.90	24.37	9.23	0.98	39.97	-
		LSTM w/o MO	-16.74	22.07	-15.13	0.79	-0.33	18.57	1.28	26.08	5.65	2.83	43.93	-
	Gwangju	LSTM w/ MO	-18.34	23.68	-13.94	1.02	-0.50	-	-9.62	31.84	20.09	1.58	43.61	-
		LSTM w/o MO	-1.43	25.78	-9.18	0.87	-0.58	-	-9.62	31.15	17.66	6.53	48.56	-
	Daegu	LSTM w/ MO	8.85	16.59	-4.39	-0.04	-0.40	-	2.40	10.18	10.49	8.65	37.87	-
		LSTM w/o MO	2.11	5.87	-7.05	-0.01	-0.63	-	3.30	16.42	14.81	10.97	40.94	-
	Ulsan	LSTM w/ MO	17.19	11.93	-8.32	0.20	-0.51	11.13	-1.39	19.78	14.16	-3.99	60.12	-
		LSTM w/o MO	16.19	14.95	-6.33	0.03	-0.50	5.10	0.31	23.72	11.47	-6.26	47.84	-
	Busan	LSTM w/ MO	-2.83	22.95	-2.95	-0.03	-0.08	-10.40	-0.35	18.30	5.67	12.24	38.48	-
		LSTM w/o MO	11.44	19.25	-10.77	-0.04	-0.04	-18.37	1.80	14.66	11.20	7.74	30.97	-
$PM_{2.5}$	Seoul-1	LSTM w/ MO	-25.85	24.23	-5.67	0.40	-0.33	8.32	5.74	11.06	8.04	1.64	10.54	37.05
		LSTM w/o MO	-33.87	25.21	-6.30	0.10	-0.38	11.41	-5.99	14.74	13.66	2.20	11.74	35.94
	Seoul-2	LSTM w/ MO	6.46	17.38	-8.70	0.14	-0.16	10.28	-0.36	5.24	12.21	-6.17	18.29	33.89
		LSTM w/o MO	-21.21	20.68	-8.84	-0.10	-0.18	10.33	-3.38	3.75	10.72	-6.73	16.32	28.33
	Daejeon	LSTM w/ MO	-24.17	45.67	-9.97	0.23	-0.25	11.72	-0.90	24.37	9.23	0.98	39.97	-
		LSTM w/o MO	-16.74	22.07	-15.13	0.79	-0.33	18.57	1.28	26.08	5.65	2.83	43.93	-
	Gwangju	LSTM w/ MO	-5.86	9.68	-8.93	-0.49	-0.49	-	-3.92	18.55	16.29	-2.82	7.27	28.80
		LSTM w/o MO	-1.31	13.15	-3.54	1.25	-0.58	-	-13.49	26.10	18.18	-5.90	9.69	31.37
	Daegu	LSTM w/ MO	9.05	-10.16	-6.13	-0.04	-0.38	-	5.20	8.28	10.90	6.74	2.14	44.11
		LSTM w/o MO	11.93	-0.96	-10.68	-0.06	-0.52	-	1.52	7.96	2.22	11.78	9.80	32.66
	Ulsan	LSTM w/ MO	-8.11	5.63	-10.52	0.07	-0.15	11.56	1.75	7.14	8.54	-3.82	-0.51	83.38
		LSTM w/o MO	4.31	7.13	-10.70	0.32	-0.23	14.02	-1.26	21.07	11.87	-10.40	-1.50	81.17
	Busan	LSTM w/ MO	-11.75	16.47	-24.01	0.19	-0.06	-3.59	0.96	7.83	8.29	16.23	-6.36	48.77
		LSTM w/o MO	17.44	8.70	-25.10	0.06	0.12	1.80	-2.13	7.65	6.35	8.99	1.49	58.50

[1)] LSTM w/ MO and LSTM w/o MO represent the LSTM model with and without consideration of missing observations in the model training; TA, WD, WS, RN, RNH, and RH denote temperature, wind direction, wind speed, daily accumulative precipitation, hourly precipitation, and relative humidity of previous day; SO_2, O_3, NO_2, CO, PM_{10}, and PM_2 are the concentrations of the respective air pollutants on the previous day.

Table R2. Performance comparison between the LSTM model with and without considerations of missing observations[1)]

Station	Species	Model	Statistical parameter				
			IOA	RMSE	MB	MNGE	MNB
Seoul - 1	PM_{10}	LSTM w/ MO	0.62	24.22	-3.20	49.72	-5.27
		LSTM w/o MO	0.57	25.35	0.91	57.16	1.49
	$PM_{2.5}$	LSTM w/ MO	0.71	12.51	-1.33	56.03	-4.58
		LSTM w/o MO	0.71	12.95	-2.56	53.53	-8.81
Seoul - 2	PM_{10}	LSTM w/ MO	0.76	21.19	-1.29	46.72	-2.40
		LSTM w/o MO	0.69	23.53	-4.90	48.35	-9.14
	$PM_{2.5}$	LSTM w/ MO	0.77	15.14	-1.09	57.60	-3.48
		LSTM w/o MO	0.75	16.06	-4.62	51.52	-14.78
Daejeon	PM_{10}	LSTM w/ MO	0.67	19.17	6.28	72.01	15.51
		LSTM w/o MO	0.59	19.13	-0.44	62.07	-1.16
	$PM_{2.5}$	LSTM w/ MO	0.67	12.17	3.99	72.01	16.49
		LSTM w/o MO	0.59	12.15	-0.28	62.07	-1.16
Gwangju	PM_{10}	LSTM w/ MO	0.67	18.92	1.69	74.68	3.96
		LSTM w/o MO	0.72	18.41	0.74	66.60	1.72
	$PM_{2.5}$	LSTM w/ MO	0.63	11.53	-0.23	82.74	-0.98
		LSTM w/o MO	0.68	11.86	2.40	95.92	10.46
Daegu	PM_{10}	LSTM w/ MO	0.71	16.46	6.02	44.12	15.26
		LSTM w/o MO	0.71	16.51	5.34	43.30	12.67
	$PM_{2.5}$	LSTM w/ MO	0.78	9.91	0.00	39.07	0.01
		LSTM w/o MO	0.67	11.55	0.79	43.78	3.06
Ulsan	PM_{10}	LSTM w/ MO	0.79	18.57	-1.00	37.33	-2.21
		LSTM w/o MO	0.69	18.60	-1.55	37.37	-3.41
	$PM_{2.5}$	LSTM w/ MO	0.79	12.75	2.52	64.04	9.39
		LSTM w/o MO	0.72	12.95	-0.14	57.43	-0.21
Busan	PM_{10}	LSTM w/ MO	0.74	16.58	0.41	44.37	1.03
		LSTM w/o MO	0.68	17.62	1.89	48.21	4.69
	$PM_{2.5}$	LSTM w/ MO	0.79	11.13	0.82	38.63	3.05
		LSTM w/o MO	0.77	12.07	0.91	40.98	3.39

[1)] LSTM w/ MO and LSTM w/o MO represent the trained LSTM model with and without missing observations; the units for RMSE and MB are $\mu g/m^3$, and those for MNGE and MNB are in %.