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 #1

1. "*The authors should better clarify what the intended use-case is for the machine learning model. In particular, there is some ambiguity in the word 'prediction' as it means different things in machine learning (where it means the 'guess' of the model) and atmospheric chemistry (where it refers to concentration estimates in the future): is the goal of the LSTM model to make a prediction of $PM_{2.5}/PM_{10}$ concentrations 24 hours from now - based on current conditions? If so, I assume the inputs/outputs have been prepared in such a way that they incorporate this 24-hour time lag? Or is the LSTM designed to make an optimal prediction of the concentration at a given time based on current conditions? In this case, it is not fully clear what the use-case for such a model would be. In general, the mapping between input features and the predictor variable needs more explanation. For example, based on Figures 9 and 10 one would conclude that the variables needed to make a model prediction are the current meteorological conditions as well as the previous day pollutant concentrations? If this is the case, was there a rationale for this choice? Also, $PM_{2.5}$ concentration is not used as an input for the PM_{10} prediction model (Figure 9), but seems to be used for the prediction of $PM_{2.5}$ (Figure 10)?*"

Reply) The aim of the LSTM system is daily prediction of PM_{10} and $PM_{2.5}$ with one hour resolution. In this study, we used the meteorological conditions and atmospheric pollutant concentrations of the previous day as input parameters to predict the next 24 hour PM concentrations (i.e., there is a 24-hr time lag between the independent and dependent variables). Regarding this, we added a brief explanation about the mapping between the input and target values in the revised manuscript (please, refer to pp. 4:30-5:2).

Based on the reviewer's comment, we conducted a sensitivity test for the new input variable of $PM_{2.5}$ to the PM_{10} predictions. The accuracy of PM_{10} predictions with and without the added independent variable of $PM_{2.5}$ is summarized in Table R1. Even with the $PM_{2.5}$, no significant changes were found in the LSTM-based PM_{10} predictions. In terms of IOA, most of the LSTM-based PM_{10} predictions without $PM_{2.5}$ are correlated better with the observed PM_{10} than did those with $PM_{2.5}$ except for the Seoul-2 and Daegu sites. Several AIR KOREA sites did not have monitored $PM_{2.5}$. In addition, this variable is not always available for the PM_{10} predictions because of instrument malfunction such as at the Daejeon site. Therefore, the current combination of independent variables is more suitable for daily PM_{10} prediction, we believe (please, also check out Table R1 attached in this reply).

2. "*The motivation to choose LSTM over another architecture should be discussed in more detail. Was LSTM selected because urban PM concentrations are expected to be dominated by local processes (e.g., emissions) and thus have a local, time-persistent signal? This would be a reasonable argument, but possibly also limits the usefulness of this approach to (urban) areas where PM concentrations are primarily determined by local processes? Based on the current version of the manuscript, it is not obvious why a simpler architecture (e.g., XGBoost) wouldn't yield a comparable (or even better) result.*"

Reply) We added more explanations for why we selected the deep LSTM neural network to develop the daily PM prediction system in South Korea (please, see pp. 5:6-12)

3. "*The authors should clarify whether they trained just one LSTM model (for all 7 locations combined) or an individual LSTM for each station. If the former, can the LSTM model then also be used for PM predictions for a different city? This would be a powerful argument for this methodology and worthwhile testing.*"

Reply) We developed individual PM prediction models for seven selected sites (please, see pp. 3:15-17).

4. "*Did the authors consider to use the logarithmic of NO2 and SO2 before normalizing the inputs? These species are often log-normally distributed and applying the regular normalization function to them (Eq. 1) might not be optimal. The generated missing values for both NO2 and SO2 are much worse than the predictions for the other four species (Figure S2), which might be further indication that these two species are not treated optimally. At the very least, a justification for using non-logarithmic concentration values for NO2 and SO2 should be provided.*"

Reply) To optimize data preprocessing, we conducted several sensitivity tests including converting the input variables into logarithmic values. The generation accuracy of missing SO_2 and NO_2 with and without logarithmic conversion (LC) of SO_2 and NO_2 is summarized in Table R2. As shown in Table R2, in general the generated missing values without LC showed better correlations with the observations than did those with LC.

In addition, we evaluated the impact of LC on the PM predictions. The performance of LSTM-based PM predictions with and without the LC of SO_2 and NO_2 is summarized in Table R3. The PM predictions without LC were more accurate ($0.62 \leq IOA_{PM10} \leq 0.79$; $0.63 \leq IOA_{PM2.5} \leq 0.79$) than were those with LC ($0.60 \leq IOA_{PM10} \leq 0.74$; $0.63 \leq IOA_{PM2.5} \leq 0.77$).

5. "*The paragraph on model overfitting is confusing (page 6, line 14ff.): an overfitted model will produce better skill scores against the training data vs. the validation data since it has learned to fit well to the training data, but the model doesn't generalize. The results shown in Table 1 are not particularly encouraging in that regard and need more explanation. It would also be helpful to provide more information on the network architecture, in particular the number of hidden nodes.*

Given that the number of input features is relatively small (11 variables per station per hour) and the training period only covers 2.3 years, it seems plausible that a complex model with too many modes will (a) overfit or (b) not converge to a (local) minimum in time because the training sample is too small. With regards to the latter, it would be instructive to show the MSE as a function of training cycles."

Reply) We added more detailed explanations for model training in the revised manuscript (please, refer to pp. 6:29 – pp. 7:10).

6. *"Another issue that should be addressed in the context of overfitting is the correlation of input variables: I assume some of the input features are highly correlated (e.g. NO_2 and SO_2, $PM_{2.5}$ and PM_{10}, temperature and O_3, etc.). While this is not a problem for the LSTM, per se, it lowers the amount of (independent) information contained in the training data and will likely slow convergence of the LSTM model as the model 'wastes time' learning these correlations first. In that regard it is surprising to see that, for a number of stations, the $PM_{2.5}$ prediction strongly depends on the previous day $PM_{2.5}$ concentration but shows little dependency (or even a negative dependency) on PM_{10} concentration (Figure 10). Is this an expected result?"*

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

We also agree with the reviewer's opinion that the number of available independent variables is very important for improving the performances of the LSTM-based PM prediction model. We organized the current data set with 11 to 12 independent variables that all could be collected from the ground-based observations (i.e., AIR KOREA and AWS networks). As mentioned in Sect. 4, if more useful independent variables are obtained from ground observations, we expect that the performances of the LSTM-based model could be improved further. This is also our research topic in the future.

The model training is a process to find the optimized weight and bias vectors to minimize the outcome of cost function (i.e., general scientific knowledge is not taken into account in the model training). In the deep learning, it is impossible to identify the causal relationships between independent and dependent variables (black box). In general, the prediction of the deep neural network model is mainly determined by input features used in the model training. Therefore, it is expected that the unusual correlations between PM_{10} and $PM_{2.5}$ would be originated from the principle of model training and the characteristics of training data.

7. *"The CTM used in this study was run at 15x15 km2 horizontal resolution, which can make it challenging to compare its output against ground-based observations due to representation error. This is particularly true for urban sites that might be heavily influenced by local, smallscale emission sources that are difficult to capture at this model resolution. As such, the comparison between CTM vs. LSTM predictions is somewhat unfair as it seems likely that a CTM with a local bias correction applied to it would perform significantly better. While this might be difficult to quantify, it should at*

least be addressed in the revised version of the manuscript."

Reply) This is well-known sub-grid variability problem!! We added discussion into pp. 9:9-17. Please, take a look at this part!

Minor comments:

"Page 4, line 12: it would be helpful to provide the number of missing values (in %) for the pollutant concentrations "

Reply) The number of missing observations is summarized in Table S1.

"Page 6, line 17: I assume the authors mean 'overtuned', not 'overturned'"

Reply) We corrected it (please, see pp. 7:9).

"Page 21/22: the authors should explain why the LSTM predictions are missing for Daejeon from approximately 5/27 to 6/7.'"

Reply) We added an explanation of why there were no PM_{10} predictions on those days (please, see pp. 23).

"Appendix, equation S4: Isn't the sigmoid function defined as: s(x) = 1 / (1 + e-x) ?'"

Reply) In this study, we used the hard sigmoid function to activate the LSTM gate. We corrected the description (please, check out pp. S2:13-14).

Table R1. Accuracy of the LSTM-based PM_{10} predictions with and without the input variable of $PM_{2.5}$ [1)]

Site	Statistical parameter	Without $PM_{2.5}$	With $PM_{2.5}$
Seoul-1	IOA	0.62	0.61
	RMSE	24.22	23.81
	MB	-3.2	-2.81
	MNGE	49.72	49.50
	MNB	-5.27	-4.64
Seoul-2	IOA	0.76	0.79
	RMSE	21.19	21.29
	MB	-1.29	0.02
	MNGE	46.72	47.53
	MNB	-2.4	0.04
Daejeon	IOA	0.67	-
	RMSE	19.17	-
	MB	6.28	-
	MNGE	72.01	-
	MNB	15.51	-
Gwangju	IOA	0.67	0.66
	RMSE	18.92	18.10
	MB	1.69	-1.88
	MNGE	74.68	67.85
	MNB	3.96	-4.39
Daegu	IOA	0.71	0.72
	RMSE	16.46	15.71
	MB	6.02	4.82
	MNGE	44.12	41.81
	MNB	15.26	11.45
Ulsan	IOA	0.79	0.76
	RMSE	18.57	17.25
	MB	-1	0.08
	MNGE	37.33	34.05
	MNB	-2.21	0.18
Busan	IOA	0.74	0.71
	RMSE	16.58	17.39
	MB	0.41	1.48
	MNGE	44.37	46.33
	MNB	1.03	3.69

[1)] The units for RMSE and MB are $\mu g/m^3$, and those for MNGE and MNB are in %.

Table R2. Accuracy comparison of missing SO_2 and NO_2 generations using log and non-log scale SO_2 and NO_2[1)]

Site	non-log scale					log scale				
	Statistics	Tr. SO_2	Val. SO_2	Tr. NO_2	Val. NO_2	Statistics	Tr. SO_2	Val. SO_2	Tr. NO_2	Val. NO_2
Seoul-1	IOA	0.74	0.46	0.92	0.88	IOA	0.74	0.47	0.90	0.88
	RMSE	1.83	1.93	9.48	9.87	RMSE	1.83	2.25	10.33	9.50
	MB	0.13	-0.75	0.54	-0.39	MB	-0.46	-1.37	-1.46	-1.89
	MNGE	43.75	24.62	26.55	26.77	MNGE	35.62	27.73	26.13	25.35
	MNB	2.72	-17.00	1.49	-1.26	MNB	-9.64	-26.56	-4.07	-5.96
Seoul-2	IOA	0.84	0.56	0.91	0.88	IOA	0.74	0.51	0.90	0.88
	RMSE	1.13	1.32	8.15	9.69	RMSE	1.30	1.30	8.24	9.96
	MB	-0.08	0.21	1.29	0.86	MB	-0.44	-0.26	-1.33	-2.51
	MNGE	14.52	18.31	21.05	27.30	MNGE	14.59	16.32	17.56	24.17
	MNB	-1.36	3.50	3.19	2.41	MNB	-7.72	-4.38	-7.24	-7.06
Daejeon	IOA	0.93	0.84	0.84	0.77	IOA	0.92	0.81	0.85	0.76
	RMSE	0.79	0.86	6.03	6.09	RMSE	0.84	0.91	6.11	6.24
	MB	-0.01	-0.27	0.06	-0.79	MB	-0.14	-0.32	-0.66	-1.20
	MNGE	27.08	21.93	54.69	45.76	MNGE	25.02	23.36	45.08	41.44
	MNB	-0.19	-9.65	0.45	-7.03	MNB	-5.46	-11.52	-5.43	-10.79
Gwangju	IOA	0.87	0.67	0.79	0.79	IOA	0.75	0.59	0.77	0.79
	RMSE	1.06	1.11	9.07	9.38	RMSE	1.32	1.21	9.66	9.28
	MB	-0.06	-0.42	-0.28	1.36	MB	-0.40	-0.65	-0.56	0.49
	MNGE	24.21	23.74	37.55	49.96	MNGE	25.36	23.47	34.72	46.32
	MNB	-1.64	-13.83	-7.46	6.41	MNB	-11.81	-21.24	-13.75	2.33
Daegu	IOA	0.81	0.68	0.88	0.89	IOA	0.68	0.59	0.87	0.85
	RMSE	2.38	2.46	8.83	11.30	RMSE	2.72	2.55	9.08	12.25
	MB	0.08	0.01	0.66	-5.08	MB	-0.94	-1.08	-0.30	-6.60
	MNGE	65.92	54.14	38.95	28.68	MNGE	48.15	43.04	37.11	29.65
	MNB	2.15	0.23	2.82	-15.71	MNB	-27.10	-28.78	-1.30	-20.42
Ulsan	IOA	0.85	0.70	0.89	0.88	IOA	0.77	0.76	0.90	0.87
	RMSE	5.72	5.92	7.64	7.99	RMSE	6.46	4.61	7.37	8.13
	MB	-0.22	1.65	0.77	2.51	MB	-1.06	0.64	-1.21	0.63
	MNGE	44.70	52.01	41.18	34.88	MNGE	36.62	41.42	29.06	31.05
	MNB	-2.82	26.50	3.98	12.53	MNB	-13.51	10.21	-6.26	3.16
Busan	IOA	0.65	0.63	0.75	0.67	IOA	0.65	0.63	0.68	0.60
	RMSE	2.48	2.56	9.34	11.85	RMSE	2.51	2.38	10.04	13.08
	MB	-0.23	0.58	0.13	-2.78	MB	-0.69	0.11	-2.44	-5.84
	MNGE	26.36	36.95	46.96	40.16	MNGE	23.07	30.56	41.69	37.52
	MNB	-3.36	9.26	0.61	-10.56	MNB	-10.19	1.84	-11.40	-22.21

[1)] Tr. and Val. represent training and validation; the units for RMSE and MB are $\mu g/m^3$, and those for MNGE and MNB are in %.

Table R3. Accuracy comparison of PM predictions using log and non-log scale NO_2 and SO_2[1)]

Station	Statistical parameter	non-log scale		log scale	
		PM_{10}	$PM_{2.5}$	PM_{10}	$PM_{2.5}$
Seoul-1	IOA	0.62	0.71	0.60	0.69
	RMSE	24.22	12.51	24.66	13.03
	MB	-3.2	-1.33	1.06	0.08
	MNGE	49.72	56.03	54.99	59.06
	MNB	-5.27	-4.58	1.74	0.27
Seoul-2	IOA	0.76	0.77	0.72	0.76
	RMSE	21.19	15.14	22.25	15.92
	MB	-1.29	-1.09	-3.90	-3.56
	MNGE	46.72	57.6	46.00	53.65
	MNB	-2.4	-3.48	-7.28	-11.39
Daejeon	IOA	0.67	-	0.66	
	RMSE	19.17	-	17.57	
	MB	6.28	-	6.16	
	MNGE	72.01	-	52.56	
	MNB	15.51	-	20.84	
Gwangju	IOA	0.67	0.63	0.67	0.63
	RMSE	18.92	11.53	19.02	11.65
	MB	1.69	-0.23	-3.55	-1.49
	MNGE	74.68	82.74	63.13	78.93
	MNB	3.96	-0.98	-8.31	-6.52
Daegu	IOA	0.71	0.78	0.72	0.73
	RMSE	16.46	9.91	17.50	10.27
	MB	6.02	0.00	8.17	-0.17
	MNGE	44.12	39.07	45.98	39.56
	MNB	15.26	0.01	19.40	-0.66
Ulsan	IOA	0.79	0.79	0.74	0.77
	RMSE	18.57	12.75	16.93	12.35
	MB	-1.00	2.52	-0.60	-1.41
	MNGE	37.33	64.04	32.63	50.31
	MNB	-2.21	9.39	-1.47	-5.46
Busan	IOA	0.74	0.79	0.72	0.74
	RMSE	16.58	11.13	17.22	11.71
	MB	0.41	0.82	1.81	-2.46
	MNGE	44.37	38.63	46.13	36.08
	MNB	1.03	3.05	4.50	-9.12

[1)] The units for RMSE and MB are $\mu g/m^3$, and those for MNGE and MNB are in %