Manuscript number: gmd-2019-180

Xueling Wu *, Ying Wang, Siyuan He, Zhongfang Wu: "$PM_{2.5}$ / PM_{10} Ratio Prediction Based on a Long Short-term Memory Neural Network in Wuhan, China".

Dear editor and reviewer,

We would like to thank you for the positive and constructive comments concerning our revised manuscript (ID: gmd-2019-180). These comments are all valuable and very helpful for revising and improving our paper. We have studied the comments carefully. Based on the comments, we have made corrections which we hope will meet with your approval. In the revised manuscript, all the corrections are marked in red. The responses to the reviewers' comments are as follows (in blue font):

Responses to the Referee Comment 1:

Point-by-point responses to the comments:

1. Comment: *Abstract: The MODIS first appeared in keywords, so Moderate Resolution Spectroradiometer should provide abbreviations in the abstract.*

Response: Thank you for pointing this out. We agree with this valuable comment. The correction is as follows (line 12-15):

First, the aerosol optical depth (AOD) in 2017 in Wuhan was obtained based on Moderate Resolution Imaging Spectroradiometer (MODIS) images, with a 1 km spatial resolution, by using the Dense Dark Vegetation (DDV) method. Second, the AOD was corrected by calculating the planetary boundary layer height (PBLH) and relative humidity (RH).

2. Comment: *Introduction 1 The introduction needs significant improvement. Please following* *the state of the paper, which is nor presented well in this paper. Why do authors want to conduct this study? What is the significance of this research? The introduction only introduced the research status, advantages of methods and models, and research purposes but does not specify the significance of this study.*

Response: Thank you for your thoughtful insights. We agree with this valuable comment. We added a description of the purpose of the study in Section 1. The correction is as follows (line 90-93):

At present, air quality monitoring is still mainly based on monitoring stations, and it is difficult to acquire large-scale and accurate prediction results. In order to reduce the dependence on monitoring stations and achieve the goal of broad, rapid and accurate air quality predictions, this paper aims to use a machine learning algorithm, combined with AOD, gaseous pollutant and meteorological data, to obtain a spatially and temporally reliable prediction model.

3. Comment: *In the introduction, you described PM_{10} and $PM_{2.5}$. Why does the paper only point out aerodynamic particle size of $PM_{2.5}$, but not PM_{10}?*

Response: Thank you for your thoughtful insights. We agree with this valuable comment. The correction is as follows (line 28-31):

Particles with an aerodynamic particle size not exceeding 10 μm are called PM_{10}. PM_{10} is primarily produced by industrial production, agricultural production, construction, roadside dust, various industrial processes and natural processes such as the resuspension of local soil and dust storms..

4. Comment: *PM_{10} is primarily produced by natural processes, such as resuspending local soils,* *sandstorms, and roadside dust, and various industrial processes. This sentence need significant improvement. PM_{10} is not only derived from natural processes but also from anthropogenic emissions. In addition, were the roadside dust and various industrial processes generated by natural processes?*

Response: Thank you for your helpful insights. We agree with this valuable comment. The correction is as follows (line 28-31):

Particles with an aerodynamic particle size not exceeding 10 μm are called PM_{10}. PM_{10} is primarily produced by industrial production, agricultural production, construction, roadside dust, various industrial processes and natural processes such as the resuspension of local soil and dust storms..

5. Comment: *In the lines 30-31, "are particularly important for environmental policy and public health research". What does this sentence focus on? Anthropogenic combustion products? Please revise carefully.*

Response: Thank you for your helpful insights. We agree with this valuable comment. The correction is as follows (line 32-34):

$PM_{2.5}$ is mainly produced by anthropogenic combustion for transportation and energy production, and it is particularly important in environmental policy and public health.

6. Comment: *In the lines 36-38, this sentence lacks the reference. Some other sentences lacking references in the introduction. Please check them carefully. Besides, some references are too old in the introduction.*

Response: Thank you for your thoughtful insights. We updated some of the references.

Since some of the research results are derived from classic papers, several older references have been retained. The correction is as follows (line 38-40):

In addition, since the scattering extinction contribution of $PM_{2.5}$ particles accounts for 80% of the extinction of the atmosphere, the concentration of $PM_{2.5}$ is a key factor in determining the visibility of the atmosphere (Sisler and Malm, 1997).

7. Comment: *"Many statistical models have been used for the ground PM estimation of AOD and other predictors, such as linear regression models, random forest models, neural net- work models, and generalized additive models." In this sentence, each model should be added with the corresponding reference.*

Response: Thank you for your thoughtful insights. The correction is as follows (line 77-79):

Many statistical models have been used for the ground PM estimation of AOD and other predictors, such as linear regression models (Kim et al., 2019), random forest models (Stafoggia et al., 2019), neural network models (Sowden et al., 2018), and generalized additive models (Chen et al., 2018).

8. Comment: *Overall, the introduction is too long. Because the introduction does not mean the stack of the references, such as the lines 78-81, and there are similar problems in some sentences in the introduction. In addition, there is a suggestion that the second and third paragraphs should be selectively integrated.*

Response: Thank you for your thoughtful insights. We agree with this valuable comment. We removed the extra references in lines 80-81 and merged the second and third paragraphs. The correction is as follows (line 55-57):

There are many ways to obtain the AOD from satellite sensors such as the Geostationary Operational Environmental Satellites (GOES) (Prados et al., 2007), the Advanced Very High Resolution Radiometer (AVHRR) (Gao et al., 2016), and the Moderate Resolution Imaging Spectroradiometer (MODIS) (Levy et al, 2013).

9. Comment: *Methods: 1. In the line 141, why do gaseous pollutants exist subscripts and particulate matter do not?*

Response: Thank you for your helpful insights. We are very sorry for our incorrect notation. We modified the particulate matter subscripts throughout the text.

10. Comment: *In the lines 143-162 of the section 3.2, are the sources of the gaseous pollutants described in this section the results generated by authors? If not, please add the references to support your statements.*

Response: Thank you for your helpful insights. The monthly average values of gaseous pollutants come from calculations with the daily data released by the China National Environmental Monitoring Center (http://webinterface.cnemc.cn/cskqzlrbxsb2092932.jhtml). We described the data source and re-stated the changes in the data. The correction is as follows (line 157-159):

The calculations in this paper were based on these daily averaged data, which were released by the China National Environmental Monitoring Center (http://webinterface.cnemc.cn/cskqzlrbxsb2092932.jhtml).

11. Comment: *In section 3.3, a total of 5 meteorological stations exist near the Wuhan area. Are* *there too few stations for interpolation? Please provide parameters to prove that the spherical model of the kriging method used in the paper is reasonable.*

> **Response:** Thank you for your thoughtful insights. Since Wuhan is a provincial capital, the number of meteorological stations around Wuhan is higher than that around other cities, and the distribution of the five stations is relatively scattered. The interpolation results from them are reasonable. We also added a description of the kriging interpolation method. The correction is as follows (line 186-195):
>
> We believe that the kriging method is the most appropriate for examining the spatial characteristics of meteorological data. The kriging method is a multi-step process that includes exploratory statistical analysis of the data, variogram modelling, surface creation, and studying the various surfaces. The kriging method interpolates unknown samples according to the distribution characteristics of a few well-known data points in a finite neighbourhood. After taking into account the size, shape, and spatial orientation of the sample points, combining the spatial relationship between the known sample points and the unknown samples, and adding the structural information provided by the variogram, kriging performs a linear unbiased optimal estimation of the unknown samples in the spatial range. After comparing the kriging, natural neighbour, spline, and inverse distance weighted methods, we found that the results acquired by setting 12 interpolation points and using the spherical model of the kriging method were smoother and more suitable for the study area.

12. Comment: *Sections 3 and section 4 are methods, and they should be integrated together. Results and discussion Results and discussion section, it reads like just the results and there is* *little discussion. It is suggested that this paper should be compared with other articles on neural network models to ensure the credibility and stability of the results.*

Response: Thank you for your helpful insights. We are very sorry for the incorrect organization. We changed the title of Section 3 to “Data”. We added discussion and analysis at the end of each paragraph in Section 5. The corrections are as follows (line 326-334, 339-341, 342-346, 358-362, 378-383):

In air quality research, predictions of higher values are particularly important, because only a successful prediction of poor air quality can be used to promptly remind people to take preventive measures, such as wearing masks. This table was produced in site order, i.e., the first and second data entries are from the same site for the last two days of 2017, and the third and fourth data are from another site. The actual data for $PM_{2.5}/PM_{10}$ on the first day were generally lower than those on the next day, and the data from 7 of the sites on the last day were larger than 0.8. Only the LSTM model can produce stable and higher predictions. In the other models, the average value of $PM_{2.5}/PM_{10}$ on the day when the air quality is bad is less than 0.7, while the average value of $PM_{2.5}/PM_{10}$ of the LSTM model is 0.726. This result indicates that LSTM produced better predictions at higher values than the other machine learning model algorithms.

Since the prediction site had no input data for the whole year and is far away from the other 9 stations, the prediction result was less accurate than the time and random prediction results. However, this prediction method can better reflect the applicability of the model to spatial prediction.

The average error rate of the LSTM model was still the lowest, along with the maximum error value and minimum error rate, which was much smaller than that of the other models. In this spatial prediction, the accuracy of the prediction result when the $PM_{2.5}/PM_{10}$ was lower than 0.2 was the lowest, and the accuracy of the prediction result when the $PM_{2.5}/PM_{10}$ was larger than 0.8 was better than that when the $PM_{2.5}/PM_{10}$ was lower than 0.2. The prediction results in other cases were much better.

The random pattern prediction was based on the completely random selection of time and space aspects and can reflect the effect of air quality prediction under various climatic conditions well. The superiority of the LSTM model prediction in the random prediction pattern was more obvious than in the other patterns, which indicates that under irregular conditions, the LSTM model is more suitable for making predictions.

Since LSTM is a time-recurrent neural network that is suitable for processing and predicting events with relatively long intervals and delays in time series, the time pattern prediction results for the three prediction models are the most accurate, and the spatial pattern prediction results without any time data are the least accurate. However, the predictions for the maximum and minimum values were always below average, especially the prediction of the maximum value. The next focuses for improvement will be the optimization of the algorithm and the improvement of the prediction accuracy.

Responses to the Referee Comment 2:

Point-by-point responses to the comments:

1. Comment: *In particular, the authors should clarify the way they split their data. They state on line 271, "80% of the data were used as the training sample for modelling, and 20% of the data were used as the verification sample." It is important to specify how the hyper-parameters of their model were chosen. If they were chosen by optimising the performance against the verification dataset then it is possible that the algorithm has over-fit the hyper-parameters.*

Lines 272-274 seem to imply there was at least some hyper-parameter tuning performed. Ideally, the data should be split three ways, into a training, verification, and test, so that the hyper-parameters are tuned against the verification data, and the algorithm scored against the test data. The authors should reassure the reader that they have taken measures to ensure they have not overfit the hyper-parameters - for instance, perhaps they further split their training set.

In addition, they state on line 279 that for Section 5.1 that there were 570 samples in the training data, and what I infer is 20 samples in the verification data (two days multiplied by ten sites, as per the training data). This appears not to be an 80%/20% split. In any case, the verification data are from one period in the season (end of December)-the algorithm may simply be good at predicting air quality in December but not the rest of the year. A more convincing approach would be to test against multiple cases from throughout the year. This is similar for the spatial prediction, which appears to only be tested at one site.

As such, the authors must reassure us that their approach to validation guards against overfitting in order for this to be suitable for publication.

Response: Thank you for pointing this out. After considering your suggestions, we readjusted our model and code. Except for the data used for prediction, we divided the data set involved in the model construction into three parts: 40% of the data were used as the training samples for modeling, 30% of the data were used as the test samples, and the remaining 30% of the data was used as verification data. We tried to add a regularization term, but the effect did not improve. After adjusting the number of neurons, the number of epochs, and the batch size, the loss function we obtained has converged without overfitting. Moreover, the revised model obtained higher prediction accuracy than the original one. We have generated the following learning curves for three prediction patterns, but in order to avoid the manuscript being too verbose, we do not intend to add the learning curve to the manuscript. The three curves in this figure are the losses of training samples, test samples and verification samples (train, test, and verifi) that increase with the number of epochs. We understand that our spatial and time prediction patterns do not completely cover the whole year and all regions, so we have added a random prediction pattern which randomly selects data from the whole year and the entire region for prediction to reduce fortuity of the other two prediction patterns. The correction is as follows (line 304-313):

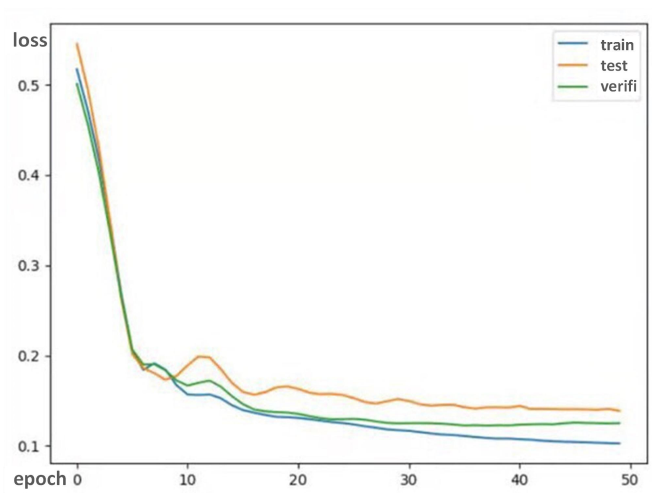

Figure 1 Learning curves

To determine the appropriate number of layers for the LSTM method, except for the data used for prediction, we divided the data set involved in the model construction into three parts: 40% of the data were used as the training samples for modeling, 30% of the data were used as the test samples, and the remaining 30% of the data was used as verification data. We tried to use various LSTM architecture layers for the comparison. After obtaining the results of various LSTM architecture layers, we found that the results obtained using the LSTM architecture with four layers were the best, with the first three layers and the dense layer as the last layer. The role of the dense layer is to complete the final output of unique values. Because the LSTM uses the activation function as the gate, the outputs of the gates must be between 0 and 1, and the output ranges of both types of activation functions must be satisfied. We determined that the activation function for setting the forget gate and the input gate was defined as a sigmoid function. After adjusting the number of neurons, the number of epochs, and the batch size, the loss function we obtained has converged without overfitting.

2. Comment*: Could you discuss why you have focused on the PM2.5/PM10 ratio as opposed to considering them separately?*

Response: Thank you for pointing this out. The explanation of using the $PM_{2.5}$-PM_{10} scale is as follows (line 43-47):

Since fine and coarse particles come from different sources, the $PM_{2.5}$-PM_{10} scale model has different physicochemical properties, which can not only distinguish the type of aerosol in the PM but also provide the mixing ratio of dust and artificial aerosols

(Sugimoto et al., 2015). The $PM_{2.5}$-PM_{10} scale is the main indicator for macro analysis of the source of particulate pollution in a region, which is more practical than considering $PM_{2.5}$ and PM_{10} separately.

3. Comment*: line 75: "random precision", and Section 5.3 "random pattern prediction". Please could you clarify what this is - it was unclear to me. Are you randomly selecting a subset of points in space and then predicting them with the remaining, contemporaneous points? If so, how is this significantly different from the spatial prediction in Section 5.2? Please better explain this task near the beginning of the manuscript.*

Response: Thank you for pointing this out. We have added the explanation of these three prediction patterns to the Section 1-Introduction (line 80-84):

The time precision mentioned in this article refers to the accuracy of inputting time-series data to predict the subsequent period results; the spatial precision refers to the accuracy of inputting all-time data of spatial points to predict the result of another spatial point; the random accuracy refers to the accuracy of inputting data of any time and space to predict the random selection data.

4. Comment*: line 112,113: I found this sentence confusing. Are "monitoring station" and "monitoring site" different things? I'm unclear what the definition is on an "inspection standard"? Does this mean the "truth" data you are using for the verification. I'm unclear what "correlation factors" means.*

Response: Thank you for your thoughtful insights. "Monitoring station" and "monitoring site" have the same meaning, and we replaced "monitoring site" with "monitoring station". "Inspection standard" means truly data and "Correlation factors" refer to $PM_{2.5}$, PM_{10}, and gaseous pollutant data detected by the stations. We have redefined these meanings in the article (line 126-127):

Therefore, we used the truly air quality data from the ground monitoring stations as the inspection standard for verification and extracted the values of $PM_{2.5}$, PM_{10}, and gaseous pollutant with the data from the monitoring stations.

5. Comment*: line 132: You verify your data processing against NASA data. Is NASA has a product, why not just use that?*

Response: Thank you for your thoughtful insights. The commonly used aerosol automatic observation network AERONET jointly established by NASA and CNRS is of good quality and easy to obtain, but the number of stations is limited, and there is no station coverage in the study area. The remote sensing data we collected are better in time and space continuity, and the AOD retrieval algorithm is also applicable to the study area.

6. Comment*: line 175: "higher trend" and "lower trend". The word "trend" changes the meaning. I presume this should read "average temperature is higher in summer and lower in winter", as expected. Otherwise, I don't understand what it means.*

Response: Thank you for your thoughtful insights. We have modified the expression of this sentence (line 197-198):

The average surface temperature and average temperature were higher in summer and lower in winter.

7. Comment*: line 185: Is this standard published. If so, please cite. In my opinion, this can be a* *technical paper as opposed to a peer review paper (but the editor may feel differently).*

Response: Thank you for your thoughtful insights. This standard has been published, so we added a citation as follows (line 206-208):

The national standard method is performed according to the method specified in the Chinese national standard GB/T13201-91 (http://www.mee.gov.cn/gzfw_13107/kjbz/qthjbhbz/qt/201605/t20160522_342349.shtml).

***8. Comment**: line 230: Could you clarify the optical subset approach? Was the R^2 score performed on the output of the LSTM as compared with the observations. If so, it may be better to put this section after the description of the LSTM and make that clear.*

Response: Thank you for your thoughtful insights. We clarified the optical subset approach. In addition, we did not perform R^2 scoring on the output of the LSTM, because the relative error rate can also reflect the accuracy intuitively. Since the maximum relative error and minimum relative error needs to be analyzed at the same time, it is more neatly to display the three relative error rates in a table. The explanation of optimal subset method is as follows (line 249-252):

The process of the optimal subset method is that in a set containing multiple independent variables, freely selecting and combining from each independent variable, combining all independent variables and dependent variables to establish all possible equations, and then the best independent variable combination model is selected from all the fitted regression equations.

***9. Comment**: Table 4: I presume this table shows only the top 10 scoring selections? Presumably* *you scored all combinations of predictors. Please explain.*

Response: Thank you for your thoughtful insights. I added an explanation for table 4 as follows (line 257-258):

This table shows the top 10 scores for R^2 scores and the corresponding factor combinations.

10. Comment*: Line 258: Normally the first gate is expressed at deciding what to forget, rather than what to remember (I appreciate they are equivalent). Figure 7 shows a "Forget gate" so it would be helpful to standardise the terminology.*

Response: Thank you for your thoughtful insights. The correction is as follows (line 286-288):

The input gate determines how much of the input x_t of the network is saved to the cell state c_t at the current moment, the forget gate determines how much the cell state c_{t-1} at the previous moment is retained to the current moment c_t, and the output gate controls how much the cell state c_t is output to the current output value h_t of the LSTM.

11. Comment*: "and the third switch controls whether c is the current output of the LSTM model" I'm not sure about this - correct me if I'm wrong, but isn't c_t combined with h_t-1, and x_t to create h_t, which is the output?*

Response: Thank you for your thoughtful insights. It is true that h_t is created by x_t with the combination of c_t and h_{t-1}. The correction is as follows (line 282-288):

Fig. 6 emphasizes the calculation process of the cell state c, and the overall process of the LSTM model is shown in Fig. 7. The key point of the LSTM model is how to control the state c. The idea of the LSTM model is to use three control switches to control it. The switches implemented in the algorithm are known as "gates", which are fully connected layers whose input is a vector, and the output is a real vector between 0 and 1 (Srivastava and Lessmann, 2018). The input gate determines how much of the input x_t of the network is saved to the cell state c_t at the current moment, the forget gate determines how much the cell state c_{t-1} at the previous moment is retained to the current moment c_t, and the output gate controls how much the cell state c_t is output to the current output value h_t of the LSTM.

12. Comment*: Section 4: I see the link to your code, that you used Keras and their LSTM implementation, which is great. If possible, could you cite Keras directly in the manuscript, and state that you used their implementation of LSTMs.*

Response: Thank you for your thoughtful insights. I cited Keras in the manuscript as follows (line 301-302):

The implementation of the LSTM models is based on Keras which is a high-level neural network Application Programming Interface written in Python.

13. Comment*: line 273: "with the first three layers being the LSTM layer and the last layer being the dense layer". I presume this means that you have three LSTM units follows by a dense layer. However, each LSTM unit can be thought of comprising multiple layers, so this terminology is confusing. In addition, could you explain to the reader the purpose of the final dense layer.*

Response: Thank you for your thoughtful insights. This sentence does means that we have three LSTM units follows by a dense layer. To avoid misunderstandings, we adjusted the description and explained the purpose of the dense layer as follows (line 307-310):

We tried to use various LSTM architecture layers for the comparison. After obtaining the results of various LSTM architecture layers, we found that the results obtained using the LSTM architecture with four layers were the best, with the first three layers and the dense layer as the last layer. The role of the dense layer is to complete the final output of unique values.

14. Comment*: line 282: I don't understand the difference between a "multilayer perceptron" and an "artificial neural network". In addition, I would have thought the LSTM, multilayer perceptron and artificial neural network, all rely on back propagation, so I don't understand the terminology "back propagation artificial neural network". Please clarify.*

Response: Thank you for your thoughtful insights. The reason why we discriminate between MLP and BP is that MLP is a back propagation neural network model with a three-layer architecture after adjusted in Python by us, while the BP neural network acquired by the clementine software with non-adjustable parameters. After careful consideration, we decided to delete the comparison with MLP to avoid ambiguity.

15. Comment*: Section 5.3: as mentioned above, I don't understand what is being done in this section.*

Response: Thank you for your thoughtful insights. The output of the random prediction pattern in Section 5.3 is randomly selected data in any time and space. This pattern is different from the time and space pattern. The applicability of LSTM reflected by the random prediction pattern is more obvious than the other two patterns. The explanation for section 5.3 is in the manuscript as follows (line 358-362):

The random pattern prediction was based on the completely random selection of time and space aspects and can reflect the effect of air quality prediction under various climatic conditions well. The superiority of the LSTM model prediction in the random prediction pattern was more obvious than in the other patterns, which indicates that under irregular conditions, the LSTM model is more suitable for making predictions.

*16. **Comment**: line 24: "Aerosols are a general term" -> "Aerosol is a general term".*

Response: Thank you for your thoughtful insights. We apologize for our mistakes. The correction is as follows (line 25):

Aerosol is a general term for solid and gas particles suspended in air.

*17. **Comment**: Please define all your acronyms on first use, for instance RH, DVV etc.*

Response: Thank you for your thoughtful insights. The correction is as follows (line 12-15):

First, the aerosol optical depth (AOD) in 2017 in Wuhan was obtained based on Moderate Resolution Imaging Spectroradiometer (MODIS) images, with a 1 km spatial resolution, by using the Dense Dark Vegetation (DDV) method. Second, the AOD was corrected by calculating the planetary boundary layer height (PBLH) and relative humidity (RH).

*18. **Comment**: line 74: "intelligent models", please clarify what this means.*

Response: Thank you for your thoughtful insights. We replaced "intelligent models" with "machine learning models". The correction is as follows (line 79-80):

However, with the introduction of new machine learning models, the traditional regression model reflects the inability to balance time, space and random precision.

19.Comment*: line 89: "classic" -> "classical".*

Response: Thank you for your thoughtful insights. The correction is as follows (line 99-101):

Finally, the space and time scales and random $PM_{2.5}/PM_{10}$ predictions were determined and performed, respectively, via the LSTM model, and the prediction results of the LSTM model and other classical models were compared and analyzed.

20. Comment*: line 108: -> "Our environmental monitoring station only monitors data in real-time". I'm also not sure what this sentence is meant to mean. Are you saying that the instrument doesn't provide information about the future? Please clarify.*

Response: Thank you for your thoughtful insights. The supplement is as follows (line 122-123):

The data that our environmental monitoring station can monitor is only real-time data with no predicting subsequent data in advance.

21. Comment*: line 110: "which have", is ambiguous - is this referring to the AOD or the atmospheric aerosols? It seems redundant to say that PM is linked to atmospheric aerosols. Please clarify this sentence.*

Response: Thank you for your thoughtful insights. "Which have" refers to AOD. The correction is as follows (line 123-124):

The AOD which has a great relationship with PM is an important parameter in the study of atmospheric aerosols.

22. Comment*: line 140: "The shortest distance between points exceeds 3 km, and the average distance exceeds 10 km." I don't know what the word "exceeds" means in this context. Can we replace it with "is" instead?*

Response: Thank you for your thoughtful insights. The correction is as follows (line 155):

The shortest distance between points is more than 3 km, and the average distance is about 10 km.

A special thanks to you for your insightful and valuable comments.

We greatly appreciate the editors and reviewers for their helpful work and hope that the corrections will meet your approval. Once again, thank you very much for your valuable and helpful comments and suggestions.

With best wishes

Yours sincerely,

Xueling Wu

Institute of Geophysics and Geomatics, China University of Geosciences

No. 388 Lumo Road, Wuhan 430074, P. R. China

Email: snowforesting@163.com

The corrected tables are as follows:

Table 5 The results and relative error rates of the time pattern predictions.

Measured value	Predicted value				Relative error rate (%)			
	LSTM	BP	SVM	CHAID	LSTM	BP	SVM	CHAID
0.8212	0.6335	0.7786	0.6698	0.4853	22.8604	5.1875	18.4364	40.9036
0.7436	0.5610	0.6961	0.7841	0.4853	24.5491	6.3878	5.4465	34.7364
0.6629	0.7346	0.7074	0.8353	0.6753	10.8069	6.7129	26.0069	1.8706
0.6950	0.7949	0.6850	0.5628	0.6753	14.3746	1.4388	19.0216	2.8345
0.7816	0.7347	0.6871	0.8092	0.5145	5.9982	12.0906	3.5312	34.1735
0.6311	0.7605	0.5864	0.7032	0.6487	20.5089	7.0829	11.4245	2.7888
0.7959	0.7347	0.6870	0.8568	0.6973	7.6931	13.6826	7.6517	12.3885
0.8743	0.8067	0.6474	0.7451	0.6973	7.7307	25.9522	14.7775	20.2448
0.7204	0.6553	0.8106	0.7446	0.8206	9.0291	12.5208	3.3592	13.9089
0.9854	0.7128	0.7154	0.6760	0.8206	27.6610	27.4000	31.3984	16.7242
0.7079	0.7249	0.8321	0.6089	0.7959	2.4048	17.5449	13.9850	12.4311
0.9455	0.7790	0.7064	0.7285	0.7959	17.6108	25.2882	22.9508	15.8223
0.7200	0.4924	0.6692	0.8172	0.6931	31.6131	7.0556	13.5000	3.7361
0.8600	0.6521	0.6192	0.6907	0.6931	24.1694	28.0000	19.6860	19.4070
0.6571	0.6432	0.7011	0.8522	0.5812	2.1242	6.6961	29.6911	11.5508
0.9189	0.7175	0.6195	0.7146	0.5812	21.9150	32.5824	22.2331	36.7505
0.7640	0.7673	0.6549	0.5406	0.7870	0.4291	14.2801	29.2408	3.0105
0.9273	0.7896	0.6354	0.7155	0.7870	14.8513	31.4785	22.8405	15.1299
0.6277	0.4614	0.7308	0.5392	0.6951	26.4993	16.4250	14.0991	10.7376
0.8896	0.6904	0.6685	0.6694	0.7534	22.3909	24.8539	24.7527	15.3103
Mean:					15.7613	16.1330	17.7017	16.2230
Maximum:					31.6111	32.5824	31.3984	40.9036
Minimum:					0.4319	1.4388	3.3592	1.8706

Table 6 The results and relative error rates of the spatial pattern prediction.

Models	LSTM	BP	SVM	CHAID
Mean:	27.9231	34.1333	34.0207	33.7718
Maximum:	178.0639	222.9295	204.7317	230.1367

Minimum:	0.0764	0.1124	0.9026	0.2396

Table 7 The results and relative error rates of the random pattern prediction.

Measured value	Predicted value				Relative error rate (%)			
	LSTM	BP	SVM	CHAID	LSTM	BP	SVM	CHAID
0.5870	0.6031	0.5762	0.6091	0.4928	2.7428	1.8399	3.7649	16.0477
0.6213	0.6581	0.6561	0.6826	0.6795	5.9231	5.6012	9.8664	9.3675
0.9843	0.4662	0.6247	0.6185	0.7422	52.6364	36.5336	37.1635	24.5962
0.8000	0.4198	0.4772	0.5231	0.4928	47.5250	40.3500	34.6125	38.4000
0.4638	0.4654	0.4773	0.5136	0.4928	0.3450	2.9107	10.7374	6.2527
0.7010	0.5762	0.6811	0.6675	0.6795	17.8031	2.8388	4.7789	3.0670
0.2222	0.2470	0.4292	0.3971	0.3737	11.1611	93.1593	78.7129	68.1818
0.5929	0.6418	0.6543	0.6598	0.6795	8.2476	10.3559	11.2835	14.6062
0.9571	0.5875	0.6246	0.6698	0.6164	38.6167	34.7404	30.0178	35.5971
0.7576	0.7095	0.5959	0.6398	0.4928	6.3490	21.3437	15.5491	34.9525
0.6277	0.6368	0.6935	0.6802	0.6795	1.4497	10.4827	8.3639	8.2523
0.8896	0.6508	0.7551	0.7353	0.7422	26.8435	15.1192	17.3449	16.5692
Mean:					18.3036	22.9396	21.8496	22.9909
Maximum:					52.6364	93.1593	78.7129	68.1818
Minimum:					0.3450	1.8399	3.7649	3.0670