Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-180-AC1, 2019

Geoscientific Model Development Discussions
Open Access

Interactive comment on "PM2.5 / PM10 Ratio Prediction Based on a Long Short-term Memory Neural Network in Wuhan, China" *by* Xueling Wu et al.

Xueling Wu et al.

wuxl@cug.edu.cn

Received and published: 22 November 2019

We would like to thank all of the positive and constructive comments concerning our revised manuscript. These comments are all valuable and helpful for revising and improving our paper. We have carefully studied the comments. Based on the comments, we have made corrections, and in the revised manuscript which in the supplement, all the corrections were marked in red. 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.

Printer-friendly version

Discussion paper

Response: Thank you for pointing this out. We agree with this valuable comment. The correction is as follows (line 12-14):

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 method. Second, the AOD was corrected by calculating the planetary boundary layer height and relative humidity.

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 83-86)): 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 PM10 and PM2.5. Why does the paper only point out aerodynamic particle size of PM2.5, but not PM10? Response: Thank you for your thoughtful insights. We agree with this valuable comment. The correction is as follows (line 27-30):

Particles with an aerodynamic particle size not exceeding 10 μm are called PM10. PM10 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: PM10 is primarily produced by natural processes, such as resuspending local soils, sandstorms, and roadside dust, and various industrial processes. This sentence need significant improvement. PM10 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 27-30): Particles with an aerodynamic particle size not exceeding 10 μm are called PM10. PM10 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 31-33): PM2.5 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 37-39): In addition, since the scattering extinction contribution of PM2.5 particles accounts for 80% of the extinction of the atmosphere, the concentration of PM2.5 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 74-76): 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 52-54): 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 149-151): 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 178-188): 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 neighborhood. 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 neighbor, 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 314-322, 327-330, 333-336, 350-353, 369-373): 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 PM2.5/PM10 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 could produce predictions at such extremely high values. In the other models, there was only one result greater than 0.8 for the prediction data, while the LSTM algorithm had three prediction results higher than 0.8. 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. In this spatial prediction, the accuracy of the prediction result when the PM2.5/PM10 was lower than 0.2 was the lowest, and the accuracy of the prediction result when the PM2.5/PM10 was larger than 0.8 was better than that when the PM2.5/PM10 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.

Please also note the supplement to this comment:
https://www.geosci-model-dev-discuss.net/gmd-2019-180/gmd-2019-180-AC1-supplement.pdf

---

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-180, 2019.