Point-by-point responses to the comments & suggestions from the editor

Journal: Geoscientific Model Development

Manuscript ID: EGUSPHERE-2023-1531

Title: "Diagnosing drivers of PM2.5 simulation biases from meteorology, chemical composition, and emission sources using an efficient machine learning method"

Comments from the editor:

I would like to point out an issue regarding the code availability in your manuscript and CMAQ. Currently, for CMAQ, you mention that it is available in a GitHub repository. GitHub repositories are not acceptable for scientific publication or long-term code archival, GitHub itself says it in its webpage, and we mention it in our Code and Data policy. Fortunately, for CMAQ, you have available Zenodo repositories too, such as https://zenodo.org/record/5213949. This way, please, look for the Zenodo repository corresponding to the CMAQ version that you use, post it replying to this comment, and in your manuscript cite it instead of the GitHub. If there is not a Zenodo repository for the version that you have used, you can upload the code yourself and create a new one.

- We appreciate the editor's comment on this manuscript. The CMAQ v5.0.2 code is publicly accessible at https://zenodo.org/record/1079898. We cited it in our manuscript.

Changes in Lines 271-272: "CMAQ is an open-source chemical transport model developed by the US Environmental Protection Agency, which can be downloaded at https://zenodo.org/record/1079898."

Comments from Reviewer #1:

This paper designed an efficient method based on machine learning for diagnosing the drivers of the Community Multiscale Air Quality (CMAQ) model biases in simulating PM2.5 concentrations from three perspectives of meteorology, chemical composition, and emission sources. The authors used source-oriented

CMAQ to diagnose the influences of different emission sources on PM2.5 biases. While the approach presented in the manuscript, particularly the emphasis on identifying biases in the CMAQ simulation of PM2.5 concentration, is innovative and distinct from conventional predictive models, a few issues should be addressed.

- We appreciate the comments from the reviewer, which help improve the manuscript. In the revision, we carefully revised the manuscript based on these comments.

Major comments:

1. Line 102: "Five-fold cross-validation method was used to evaluate the model fit and prediction ability (Browne, 2000)," and Line 143: "The ML models were trained separately for different regions and seasons, and a 5-fold cross-validation was used to measure the model performance (Figure 4)."

Cross-validation is employed primarily for model selection and hyperparameter tuning rather than evaluating the ML performance. To accurately evaluate the model's performance and generalization capabilities, it is essential to test it on a dataset it has never seen during training or validation. In addition, for clarification and completeness, the authors should provide a detailed explanation of why they chose 5-fold cross-validation and how they implemented this methodology in their study.

- Thanks for the comment. We carried out testing for the model used in this study. We randomly selected 20% of the data as the test set, and trained the model using a combination of meteorological, emission, and $PM_{2.5}$ components features, then predicted the simulated bias of the test set and compared it to the true bias ($PM_{2.5}$ from observations minus $PM_{2.5}$ from CMAQ simulations) (Figure S4). The model show an prediction R^2 of 0.68 and RMSE of 17.26 μg/m^3. We have added the corresponding results in Section 3.2. We added the reasons for choosing cross-validation and briefly introduced the cross-validation method that we used in Section 2.3.

Changes in Lines 151-154: "First, 20% of the data (not involved in training) were randomly selected for model evaluation (Figure S4). Training was performed

using a combination of PM$_{2.5}$ components, meteorological, and emission features. The model showed an prediction R^2 of 0.68 and RMSE of 17.26 μg/m^3."

Changes in Lines 107-111: "Cross-validation (CV) is an effective model validation method to prevent overfitting (Browne, 2000). To improve computational efficiency and enlarge the test dataset size, five-fold CV method was used to evaluate the model performance. The dataset was randomly divided into five parts, one was taken in turn as a test and the rest was used for training, which was repeated five times, and then the mean coefficient of determination (R^2) and the root mean square error (RMSE) were calculated."

2. Line 221: "In addition, the main objective of this study was diagnosing the contributors to CMAQ simulation biases using machine learning, therefore we did not pursue a very good model performance."

If the model's efficacy is insufficient, the interpretations and conclusions that can be drawn from it may be weakened. It is crucial to ensure that the model's predictions or interpretations are at least reasonably accurate. In addition, when using a specific model such as LightGBM, it would be beneficial to provide justification or evidence for why it outperforms other models such as Random Forest (Line 219) in this study. Such a justification can lend more credibility to the findings and insights derived from the model.

- Thanks for the comment. We use machine learning to capture the relationship between simulation bias and input variables, rather than for prediction. Since meteorology or emissions can only partially explain the simulation bias, a poor R^2 is justified when fitting the model with only meteorology or emissions variables. R^2 here indicates how well the input variables explain the results, and a low R^2 indicates a minor influence of the input variables to the simulation bias. We added the corresponding description in Section 3.3.

The LightGBM model is an optimized Gradient Boosting Decision Tree (GBDT). Compared to XGBoost, a widely used GBDT, LightGBM uses Histogram's decision tree algorithm along with Gradient-based One-Side Sampling (GOSS), which

saves memory and computation time (Ke et al., 2017). Three tree-based models, Random Forest, XGBoost, and LightGBM, were compared in our previous study (Wang et al., 2023). We found that using the same input data and hyperparameters, LightGBM is as accurate as XGBoost, but faster and less overfitting (the difference in accuracy between training and testing), so here we chose the LightGBM model for simulation bias diagnosing. We added the reasons for choosing LightGBM in Section 2.3.

Changes in Lines 242-246: "In addition, the main objective of this study was diagnosing the contributors to CMAQ simulation biases using machine learning, rather than for prediction. Since meteorology or emissions can only partially explain the simulation bias, a low R^2 is justified when fitting the model with only meteorology or emissions variables (Figure 4), which indicates a minor influence of the input variables to the simulation bias."

Changes in Lines 93-99: "The LightGBM model is an optimized Gradient Boosting Decision Tree (GBDT) (Ke et al., 2017), and has shown accurate performance in many fields (Wei et al., 2021; Yan et al., 2021; Sun et al., 2020; Liang et al., 2020). Compared to XGBoost, a widely used GBDT, LightGBM uses Histogram's decision tree algorithm along with Gradient-based One-Side Sampling (GOSS), which saves memory and computation time (Ke et al., 2017). Three tree-based models, Random Forest, XGBoost, and LightGBM, were compared in our previous study (Wang et al., 2023). Using the same input data and hyperparameters, LightGBM is as accurate as XGBoost, but faster and less overfitting (the difference in accuracy between training and testing), so the lightGBM model was used to diagnose $PM_{2.5}$ simulation biases in this study."

Minor Comments:

1. The study area for this manuscript is China; therefore, "in China" should be added to the title.

- Thanks for the comment. We have changed the title to specify the study area of

China.

Changes in the title: "Diagnosing drivers of $PM_{2.5}$ simulation biases in China from meteorology, chemical composition, and emission sources using an efficient machine learning method"

2. Line 56: Chemical components constitute PM2.5, so they would be considered as labels. Why are they used as features? Additionally, wouldn't the linear summation of all these chemical components essentially represent PM2.5? If I have misunderstood any aspect, it would be helpful if the author could explain it.

- Thanks for the comment. Indeed, chemical components constitute $PM_{2.5}$. So, the simulation bias of total $PM_{2.5}$ should be attributed to the specific components. Using the components as input features to fit the total simulation bias can tell us which components have a large simulation bias. We have added a corresponding note in the manuscript.

Changes in Lines 157-158: "Using $PM_{2.5}$ components as input features to fit the total simulation bias can tell us which components have a large simulation bias."

3. Line 62: How are "problematic data points" defined?

- Thanks for the comment. We are sorry for the unclear presentation. "problematic data points" here mean extreme value, records of $PM_{2.5}$ exceeds PM_{10}, and days with less than 20-hour records. We have added this description in Section 2.1.

Changes in Lines 62-63: "The daily observations data <0.1 % quantile and >99.9 % quantile, $PM_{2.5}$ exceeds PM_{10}, and days with less than 20 valid hourly records are excluded."

4. Line 78: Table S1 is "Summary of the WRF model variables used in this study." The list of PM2.5 components simulated by CMAQ is not available in Table S1.

- Thanks for the comment. We are sorry for our carelessness. The list of $PM_{2.5}$

components simulated by CMAQ has been added to Table A1.

5. Line 96: Could you please clarify what is meant by "three combinations of input variables"? Does this refer to pairwise combinations of the categories (e.g., "meteorological factors" + "chemical components") or something else?

- Thanks for the comment. Three combinations of input variables mean meteorological factors, chemical components, and emission sources, that is, we trained the ML model for three times with three categories of variables separately. By doing so, the sources of simulation bias are analyzed from three perspectives: meteorological, emission, and components. We have added this description in Section 2.3.

Changes in Lines 101-102: "Three categories of input variables were designed to separately fit the simulation biases: meteorological factors, chemical components, and emission sources."

6. Line 107: It should be "Observed PM2.5 concentrations".

- Thanks for the comment. We have modified the corresponding descriptions and examined the manuscript carefully.

Changes in Lines 116: "Observed $PM_{2.5}$ concentrations were higher in BTH (51.172 μg/m^3) and lower in PRD (28.273 μg/m^3)."

7. Figure 3: The left axis features a stacked bar plot for sectoral contribution (with a maximal y-value of 100%), whereas the right axis represents PM2.5 concentration using a scatter plot. However, the areas in which the scatter points overlap with the bars do not provide clear information, making the use of dual axes potentially misleading.

- Thanks for the comment. We have modified Figure 3 and Figure S3, and replaced the solid circles with hollow circles in the PM scatter plot to make the figure clearer.

8. Figure 4: Some models have a training R2 lower than 0.6. This suggests that these models might be underfitting (please see my "major comments").

- Thanks for the comment. We use machine learning to capture the relationship between simulation bias and input variables, rather than for prediction. Since meteorology or emissions can only partially explain the simulation bias, a poor R^2 is justified when fitting the model with only meteorology or emissions variables. R^2 here indicates how well the input variables explain the results, and a low R^2 in some regions and seasons indicates a minor influence of the input variables to the simulation bias. Also, part of the reason could be attributed to poor model fit due to data quantity and quality. We added the corresponding description in Section 3.3.

Changes in Lines 242-246: "In addition, the main objective of this study was diagnosing the contributors to CMAQ simulation biases using machine learning, rather than for prediction. Since meteorology or emissions can only partially explain the simulation bias, a low R^2 is justified when fitting the model with only meteorology or emissions variables (Figure 4), which indicates a minor influence of the input variables to the simulation bias."

Comments from Reviewer #2:

The article describes an interesting method to determine the origin of the bias of a CTM using a ML algorithm. It applies this method to an interesting case and allows to determine what sector biases come from. This method has the potential to be applied in similar studies. An extensive bibliography is provided, enabling the reader to find more details when necessary. It would be nice to have a comparison of the results of this method, in the case studied, with those of other methods. A short reminder   of the basics of the algorithms used (even if the references provided do that in details) would have been welcome.

- We appreciate the comments from the reviewer, which help us a lot. We compare studies of CTM simulation bias identification in China using different methods. We obtained many consistent conclusions, e.g., systematic underestimation of SOA, significant contribution of primary $PM_{2.5}$ emissions, and inaccurate

simulation of nitrate in winter in Beijing. We added a discussion in Section 3.3.

The LightGBM model is an optimized Gradient Boosting Decision Tree (GBDT). Compared to XGBoost, a widely used GBDT, LightGBM uses Histogram's decision tree algorithm along with Gradient-based One-Side Sampling (GOSS), which saves memory and computation time (Ke et al., 2017). We added the introduction of LightGBM in Section 2.3.

Changes in Lines 230-237: "Huang et al. (2019) used a new reduced-form model coupled with a higher-order decoupled direct method and stochastic response surface model to identify sources of uncertainty in CMAQ simulations. An analysis for the PRD of China in spring 2013 revealed a systematic underestimation of SOA and identified wind speed and primary $PM_{2.5}$ emissions as key sources of uncertainties in $PM_{2.5}$ simulations, which is consistent with the results identified using LightGBM in this study. Aleksankina et al. (2019) identified $PM_{2.5}$ simulation bias in Europe using optimised Latin hypercube sampling and also demonstrated the important impact of primary emissions on $PM_{2.5}$ simulation uncertainties. Liu and Xing (2022) used a fully connected neural network to identify $PM_{2.5}$ simulations biases and found that NO_2 is the main contributor in BTH during heavy polluted events in the winter, which is consistent with the main contribution of nitrate that we found in the BTH (Figure S5)."

Changes in Lines 93-96: "The LightGBM model is an optimized Gradient Boosting Decision Tree (GBDT) (Ke et al., 2017), and has shown accurate performance in many fields (Wei et al., 2021; Yan et al., 2021; Sun et al., 2020; Liang et al., 2020). Compared to XGBoost, a widely used GBDT, LightGBM uses Histogram's decision tree algorithm along with Gradient-based One-Side Sampling (GOSS), which saves memory and computation time (Ke et al., 2017)."

Specific comments

L17: "model biases", bias when the model is compared to observations ? It could be precised.

- Thanks for the comment. We have modified the corresponding descriptions and

examined the manuscript carefully.

Changes in Lines 16-19: "In this study, an efficient method based on machine learning (ML) was designed to diagnose the drivers of the Community Multiscale Air Quality (CMAQ) model biases compared to observations in simulating PM$_{2.5}$ concentrations from three perspectives of meteorology, chemical composition, and emission sources."

L20: even if we are in the abstract, the expressions "fitting ability" should be precised.

- Thanks for the comment. We have modified the expressions and explained it with plain language.

Changes in Lines 19-21: "The ML model can capture the complex relationship between input variables and simulation bias well with a small performance gap between training and validation."

L29: could you remind the reader the definition of the acronyms PPM and SIA/SOA ?

- Thanks for the comment. We have clarified the definition of PPM and SIA/SOA.

Changes in Lines 30-31: "Fine particulate matter (PM$_{2.5}$) is a complex mixture of primary particulate matters (PPM) and secondary inorganic/organic components (SIA/SOA), with adverse effects on public health and ecosystems."

L42: which ML methods ?

- Thanks for the comment. We have added the popular ML methods used before: Random Forest and XGBoost.

Changes in Lines 44-46: "Recently machine learning (ML) methods, like Random Forest and XGBoost, have been widely used in environmental science researches due to their simple structure, fast speed and ability to deal with no-linear relationships"

L41-52 Could you quantifiy the gain in computing resources you achieve using your

ML method to the methods using Monte Carlo or Latin hypercube sampling ? Have you tried these other methods and compared their results with the method described in this article ?

- Thanks for the comment. The LightGBM model used in this study is very fast, taking only a few tens of seconds for a single core to train at a time, and with low memory usage (depending on the size of the training dataset), the size of this training dataset is around 360,000 rows, 57 columns, in 64-bit floating-point format, and only requires about 160 MB of memory, making it ready to run on a laptop computer. However, Monte Carlo-based methods require multiple runs of the chemical transport model for sensitivity testing, which can more accurately identify factors that cause bias, but are computationally demanding and typically cannot be run on personal computers, relying on high-performance multi-core computers.

We compare studies of CTM simulation bias identification in China using different methods. We obtained many consistent conclusions, e.g., systematic underestimation of SOA, significant contribution of primary $PM_{2.5}$ emissions, and inaccurate simulation of nitrate in winter in Beijing. We added a discussion in Section 3.3.

Changes in Lines 230-237: "Huang et al. (2019) used a new reduced-form model coupled with a higher-order decoupled direct method and stochastic response surface model to identify sources of uncertainty in CMAQ simulations. An analysis for the PRD of China in spring 2013 revealed a systematic underestimation of SOA and identified wind speed and primary $PM_{2.5}$ emissions as key sources of uncertainties in $PM_{2.5}$ simulations, which is consistent with the results identified using LightGBM in this study. Aleksankina et al. (2019) identified $PM_{2.5}$ simulation bias in Europe using optimised Latin hypercube sampling and also demonstrated the important impact of primary emissions on $PM_{2.5}$ simulation uncertainties. Liu and Xing (2022) used a fully connected neural network to identify $PM_{2.5}$ simulations biases and found that NO_2 is the main contributor in BTH during heavy polluted events in the winter, which is consistent with the main contribution of nitrate that we found in the BTH (Figure S5)."

L63: could you precise what you mean by "same simulation grid" ? Could this grid be defined.

- Thanks for the comment. CMAQ simulation was conducted with a 36 km horizontal resolution. For areas with a high density of observation sites, such as Beijing, several sites may be located in the same 36km*36km grid, in which case the average of several sites in the same grid will be calculated. We have clarified it.

Changes in Lines 63-65: "For observation sites located on the same CMAQ simulation grid (36 km × 36 km), average $PM_{2.5}$ concentrations of these sites were calculated to compare with CMAQ simulation."

L80 would it be possible to me give more detail on the source apportionment method ?

- Thanks for the comment. PPM from different source sectors are tracked by non-reactive tracers (10^{-5} of the PPM emission rates). The concentrations of PPM from given sources are then calculated by multiplying the tracer with 10^5. The contributions of source sectors to SIA are quantified using specific reactive tagged tracers. Specifically, NO_x, SO_2, and NH_3 from different sources were tracked separately through a series of chemical and physical processes involved in SIA formation. We added the corresponding description in **Section 2.2**.

Changes in Lines 83-86: "PPM from different source sectors are tracked by non-reactive tracers (10^{-5} of the PPM emission rates). The concentrations of PPM from given sources are then calculated by multiplying the tracer with 10^5. The contributions of source sectors to SIA are quantified using specific reactive tagged tracers. Specifically, NO_x, SO_2, and NH_3 from different sources were tracked separately through a series of chemical and physical processes involving in SIA formation."

L90-102: a brief presentation of the algorithms would be interesting (see general comments).

- Thanks for the comment. The LightGBM model is an optimized Gradient Boosting Decision Tree (GBDT) (Ke et al., 2017). Compared to XGBoost, a widely used GBDT, LightGBM uses Histogram's decision tree algorithm along with Gradient-based One-Side Sampling (GOSS), which saves memory and computation time. We added the description of LightGBM in Section 2.3.

Changes in Lines 93-96: "The LightGBM model is an optimized Gradient Boosting Decision Tree (GBDT) (Ke et al., 2017), and has shown accurate performance in many fields (Wei et al., 2021; Yan et al., 2021; Sun et al., 2020; Liang et al., 2020). Compared to XGBoost, a widely used GBDT, LightGBM uses Histogram's decision tree algorithm along with Gradient-based One-Side Sampling (GOSS), which saves memory and computation time (Ke et al., 2017)."

L90 the only citation for this method in the bibliography is from a conference. Why not mentioning

"Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., … Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30, 3146–3154." ?

- Thanks for the comment. We learned about lightGBM method from the conference paper, so we cited it. We have changed to a more formal citation provided by the reviewer.

Changes in Lines 93: "The LightGBM model is an optimized Gradient Boosting Decision Tree (GBDT) (Ke et al., 2017)"

L102 even if it defined in the reference, it would be nice to remind (briefly) what the X-fold cross-validation method consists in.

- Thanks for the comment. We have briefly introduced the cross-validation method that we used in Section 2.3.

Changes in Lines 107-111: "Cross-validation (CV) is an effective model validation method to prevent overfitting (Browne, 2000). To improve computational efficiency and enlarge the test dataset size, five-fold CV method was used to evaluate

the model performance. The dataset was randomly divided into five parts, one was taken in turn as a test and the rest was used for training, which was repeated five times, and then the mean coefficient of determination (R^2) and the root mean square error (RMSE) were calculated."

L135-136: why just analyzing source sectors of SIA and not SOA ?

    - Thanks for the comment. The formation mechanism of SOA is complicated and currently incomplete, and the emission of precursor VOCs has high uncertainty, therefore, we did not track sources of SOA. We have added a corresponding description in Section 2.2.

    Changes in Lines 87-88: "The source of SOA was not traced due to the complex and currently imperfect mechanism of SOA formation and the high uncertainty in the precursor VOCs emissions (Liu et al., 2021; Hu et al., 2017)."

L156: What is imperfect: the pathways or their current knowledge ?

    - Thanks for the comment. We apologize for the lack of clarity, but "imperfect" here refers to the imperfect nitrate mechanism (e.g. non-homogeneous oxidation) in the SAPRC11 mechanism that we used. We have clarified this point

    Changes in Lines 167-170: "Nitrate contribution to simulation bias further implies the inaccuracy of nitrate simulations, which can relate to the imperfect pathways of nitrate production (e.g., non-homogeneous oxidation) in SAPRC11 (that we used) and the uncertainties of nitrate precursor emission inventories in winter (Xu et al., 2019; Zhang et al., 2018; Carter and Heo, 2013)."

L196: could you be more specific about the subsurface conditions ?

    - Thanks for the comment. Here "subsurface conditions" mean the land surface properties, the rate of dry deposition is closely related to land cover type. We have added a corresponding note in Section 3.2.

    Changes in Lines 207-210: "Dry deposition is a critical but highly uncertain sink for aerosols, which depends on the chemical and physical properties of aerosols,

and be influenced by land surface properties and meteorological conditions (Shu et al., 2022). Different land-use types (e.g., vegetation, deserts, and snow) have significantly different abilities to capture particulate matter."

L206: I don't see the values mentioned for R2 and RMSE (0.53 and 20.18) in figure 6. What do they relate to ?

- Thanks for the comment. The values mentioned for R^2 and RMSE are from Figure 4, and we modified the expression.

Changes in Lines 217-218: "In China and five key regions, sectoral sources were able to fit the simulation bias well, with mean R^2 and RMSE of 0.53 % and 20.18 µg/m^3 (Figure 4)."

Technical corrections

L25: "contribution" => "contribution to this bias".

- Thanks for the comment. We have modified the expression accordingly, and checked the manuscript.

Changes in Lines 25-26: "Both primary and secondary inorganic components from residential sources showed the largest contribution to this bias (12.05 % and 12.78 %), implying large uncertainties in this sector."

L64: I would suggest to change the formulation: "Analysis focused on nationwide as well as several interested regions" => "Analysis has been carried out on several regions of interest and on all China."

- Thanks for the comment. We have modified the expression accordingly to make it clearer.

Changes in Lines 65-66: "Analysis has been carried out on several haze-polluted regions and on all China (Figure S1)"

L65: could you justify the choice of the regions ?

- Thanks for the comment. We selected sub-regions according to the severity of

haze pollution. We have modified the expression "interest" to "haze-polluted".

Changes in Lines 65-66: "Analysis has been carried out on several haze-polluted regions and on all China (Figure S1)"

L79: "conducted over" => "carried out"

- Thanks for the comment. We have modified the expression accordingly to make it clearer.

Changes in Lines 79-80: "The CMAQ simulation (36 km×36 km) was carried out in mainland China and surrounding regions in 2019."

L107   "higher" => "highest", "lower" => "lowest"

- - Thanks for the comment. We have modified the expression accordingly and double-checked the manuscript to make sure it is correct.

Changes in Lines 116: "Observed $PM_{2.5}$ concentrations were highest in BTH (51.172 μg/m^3) and lowest in PRD (28.273 μg/m^3)."

L173: "the stationary" => "a stationary"

- Thanks for the comment. We have modified the expression accordingly to make it clearer.

Changes in Lines 183-184: "High pressure systems are connected to a stationary weather, which is unfavorable for $PM_{2.5}$ dispersion."

L175: "the uncertain" => "the uncertainties"

- Thanks for the comment. We are sorry for our carelessness and have modified the expression accordingly and double-checked the manuscript to make sure it is correct.

Changes in Lines 187-188: "Contribution of wind direction in YRD may also related to the uncertainties of WRF simulation."

L177: "Earth's radiation receipts" : I would prefer   "radiation received by the Earth"

- Thanks for the comment and suggestion. We have modified the expression accordingly

Changes in Lines 189-190: "In addition to directly changing the radiation received by the earth through scattering and absorbing"

L181: "shown the dominant" => "showed the dominant".

- Thanks for the comment. We are sorry for our carelessness and have modified the expression accordingly. We double-checked the grammar of the manuscript to make sure it is correct.

Changes in Lines 193-194: "Previous study showed the dominant role of cloud chemistry in aerosol-cloud interactions in southern China"

L183: "the missing" => "the lack"

- Thanks for the comment and suggestion. We have modified the expression accordingly

Changes in Lines 194-195: "Therefore, the influence of cloud cover on simulation biases in YRD can attributed to the lack of aerosol feedback mechanism."

L185: "can associate" => "can be associated"

- Thanks for the comment. We are sorry for our carelessness and have modified the expression accordingly. We double-checked the grammar of the manuscript to make sure it is correct.

Changes in Lines 197: "These factors can be associated with ground-level sand rise and dust emission."

L186: "attributed" => "be attributed"

- Thanks for the comment. We are sorry for our carelessness and have modified the expression accordingly. We double-checked the grammar of the manuscript to make sure it is correct.

Changes in Lines 197: "Underestimation of dust aerosol in NWCHN can be attributed to emission"

L196: "influenced" => "is influenced"

   - Thanks for the comment. We are sorry for our carelessness and have modified the expression accordingly. We double-checked the grammar of the manuscript to make sure it is correct.

   Changes in Lines 207-208: "Dry deposition is a critical but highly uncertain sink for aerosols, which depends on the chemical and physical properties of aerosols, and be influenced by subsurface and meteorological conditions"

L199 " the underestimates" => "an underestimation"

   - Thanks for the comment. We have modified the expression accordingly

   Changes in Lines 211-212: "Recent studies for the United States also showed an underestimation for PM$_{10}$ concentrations."

Figure S4 : difficult to distinguish between the different shades of red/pink. The use of another color scale (with differerent colors) would be clearer.

   - Thanks for the comment. We have modified the Figure S4 (renumbered as Figure S5) with blue-yellow-red color bar to make the figure clearer.

Figure 4: Bottom of figure 4: RSME => RMSE

   - Thanks for the comment. We are sorry for our carelessness and have modified the expression in Figure 4 accordingly.

Figure 5 : difficult to distinguish between the different shades of red/pink. The use of another color scale (with differerent colors) would be clearer.

   - Thanks for the comment. We have modified the Figure 5 with blue-yellow-red color bar to make the figure clearer.

Problem in numbering: we have the same notation for figures (S1,...) and tables (S1,...) It would be better to have a different notation for the tables and for the figures

- Thanks for the comment. We have used different notation for figures (S1,…) and tables (A1,…), and modified the corresponding references in the manuscript.

Reference

Aleksankina, K., Reis, S., Vieno, M., and Heal, M. R.: Advanced methods for uncertainty assessment and global sensitivity analysis of an Eulerian atmospheric chemistry transport model, Atmos. Chem. Phys., 19, 2881-2898, 2019.

Browne, M. W.: Cross-validation methods, Journal of mathematical psychology, 44, 108-132, 2000.

Carter, W. P. and Heo, G.: Development of revised SAPRC aromatics mechanisms, Atmos. Environ., 77, 404-414, 2013.

Hu, J., Wang, P., Ying, Q., Zhang, H., Chen, J., Ge, X., Li, X., Jiang, J., Wang, S., and Zhang, J.: Modeling biogenic and anthropogenic secondary organic aerosol in China, Atmos. Chem. Phys., 17, 77-92, 2017.

Huang, Z., Zheng, J., Ou, J., Zhong, Z., Wu, Y., and Shao, M.: A Feasible Methodological Framework for Uncertainty Analysis and Diagnosis of Atmospheric Chemical Transport Models, Environ. Sci. Technol., 53, 3110-3118, 10.1021/acs.est.8b06326, 2019.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y.: Lightgbm: A highly efficient gradient boosting decision tree, Advances in neural information processing systems, 30, 2017.

Liang, W., Luo, S., Zhao, G., and Wu, H.: Predicting hard rock pillar stability using GBDT, XGBoost, and LightGBM algorithms, Mathematics, 8, 765, 2020.

Liu, J. and Xing, J.: Identifying Contributors to PM2.5 Simulation Biases of Chemical Transport Model Using Fully Connected Neural Networks, Journal of Advances in Modeling Earth Systems, 15, 2022.

Liu, J., Chu, B., Chen, T., Zhong, C., Liu, C., Ma, Q., Ma, J., Zhang, P., and He, H.: Secondary organic aerosol formation potential from ambient air in Beijing: effects of atmospheric oxidation capacity at different pollution levels, Environ. Sci. Technol., 55, 4565-4572, 2021.

Shu, Q., Murphy, B., Schwede, D., Henderson, B. H., Pye, H. O. T., Appel, K. W., Khan, T. R., and Perlinger, J. A.: Improving the particle dry deposition scheme in the CMAQ photochemical modeling system, Atmos. Environ., 289, 119343, https://doi.org/10.1016/j.atmosenv.2022.119343, 2022.

Sun, X., Liu, M., and Sima, Z.: A novel cryptocurrency price trend forecasting model based on LightGBM, Finance Research Letters, 32, 101084, 2020.

Wang, S., Wang, P., Zhang, R., Meng, X., Kan, H., and Zhang, H.: Estimating particulate matter concentrations and meteorological contributions in China during 2000–2020, Chemosphere, 330, 138742, https://doi.org/10.1016/j.chemosphere.2023.138742, 2023.

Wei, J., Li, Z., Pinker, R. T., Wang, J., Sun, L., Xue, W., Li, R., and Cribb, M.: Himawari-8-derived diurnal variations in ground-level PM2.5 pollution across China using the fast space-time Light Gradient Boosting Machine (LightGBM), Atmos. Chem. Phys., 21, 7863-7880, 10.5194/acp-21-7863-2021, 2021.

Xu, Q., Wang, S., Jiang, J., Bhattarai, N., Li, X., Chang, X., Qiu, X., Zheng, M., Hua, Y., and Hao, J.: Nitrate dominates the chemical composition of PM2. 5 during haze event in Beijing, China, Science of the Total Environment, 689, 1293-1303, 2019.

Yan, J., Xu, Y., Cheng, Q., Jiang, S., Wang, Q., Xiao, Y., Ma, C., Yan, J., and Wang, X.: LightGBM: accelerated genomically designed crop breeding through ensemble learning, Genome Biology, 22, 1-24, 2021.

Zhang, R., Sun, X. S., Shi, A. J., Huang, Y. H., Yan, J., Nie, T., Yan, X., and Li, X.: Secondary inorganic aerosols formation during haze episodes at an urban site in Beijing, China, Atmos. Environ., 177, 275-

282, 10.1016/j.atmosenv.2017.12.031, 2018.