Manuscript No.: acp-2022-550

Responses to Reviewer #1

This manuscript investigated future climate change impacts on near-surface O_3 concentrations over Asia from 2020-2100 using a machine learning model along with multisource data. The random forest model was trained based on results from global atmospheric chemical transport model simulations, real-time O_3 observations, and other datasets. Future climate-driven changes in O_3 concentrations were predicted by the trained model together with 18 CMIP6 multi-model simulations under four future scenarios. The paper found that future climate change would aggravate O_3 pollution in Asia and expanded the pollution from North China to South China and extended it into the cold season in a warming future. Overall, this is a good example of machine learning and big data analysis in atmospheric science. The results are of good significance and novelty. The manuscript was well-written and properly organized. Therefore, I recommend the acceptance of the manuscript.

We thank the reviewer for all the insightful comments. Below, please see our point-by-point response (in blue) to the specific comments and suggestions and the changes that have been made to the manuscript, in an effort to take into account all the comments raised here.

General:

1, In this study, the authors trained the machine learning (ML) model using O_3 precursor emissions, but did not consider the role of methane. So, I suggest the authors discuss more about how the results will be affected if the role of methane is included in the ML model prediction.

Response:

The treatment of methane in GEOS-Chem model are different from other precursor emissions. Its concentrations are directly prescribed as a surface boundary condition. Therefore, we did not consider the role of methane in machine learning model. However, the climate influence of methane is considered in the CMIP6 multi-model simulations. Therefore, the impact of future methane concentrations on O_3 via climate change is included in machine learning projections. We have now added a discussion in the revised manuscript. “Although the methane concentrations in GEOS-Chem model are prescribed and its role in the O_3 production is not considered in the ML model, the climate influence of methane is included in the CMIP6 multi-model simulations. Consequently, the impact of future changes in methane on O_3 concentrations via climate change are considered in the future projections.”

Minor:

1. Line 42: Change “primary” to “secondary”.

Response:

Revised as suggested.

2. Lines 67-74: How is performance of model in predicting ozone pollution?

Response:

Xue et al. (2020) and Wei et al. (2022) estimated maximum daily 8 h averaged (MDA8) near-surface O_3 concentrations across China using the random forest model together with multisource data, and the coefficient of determination (R^2) between the observed and predicted O_3 concentrations reached 0.70 and 0.87, respectively. Di et al. (2017) applied the neural network to simulate MDA8 O_3 concentrations in the continental United States with a mean cross-validation R^2 value of 0.76. Su et al. (2020) adopted a support vector machine model to predict summer hourly O_3 concentrations in Nanjing, and the R^2 reached 0.78. Liu et al. (2020) established a nationwide MDA8 O_3 prediction model based on the extreme gradient boosting approach, and reported a R^2 value of 0.78. Liu et al. (2022) used cluster-enhanced ensemble machine learning method to construct a global monthly near-surface O_3 dataset from 2003 to 2019, which showed an excellent overall R^2 of 0.91. We have now added a note at the end of the paragraph as follows: “The abovementioned previous studies utilizing the ML methods showed high computational efficiency and accuracy, with an overall R^2 between the observed and predicted O_3 concentrations in the range of 0.7–0.9.”

3. Line 172: Has this dataset been quality controlled or has been used in scientific studies?

Response:

The Chine Ministry of Ecology and Environment (MEE) categorized a total of 360 key cities by 2020. Each city covers several O_3 monitoring sites, and we average them hourly at city level. Since the observed O_3 concentrations reported by MEE were under the standard conditions of temperature of 273 K and air pressure of 1013 hPa until 31 August 2018, we adjusted the post-August 2018 O_3 concentrations under the ambient state of 298 K and 1013 hPa to that under the reference state.

The dataset has been widely used to examine O_3 pollution over China in previous studies (Li et al., 2020, 2021; Qian et al., 2022). We have now revised the descriptions in the manuscript.

4. Line 217: Please describe more about the four scenarios.

Response:

Thanks for suggestion. We have now added more description about future scenarios that “Monthly meteorological parameters under four different future climate scenarios, including SSP1-2.6, SSP2-4.5, SSP3-7.0 and SSP5-8.5 (a representation of low, intermediate, medium to high, and high forcing levels, respectively), are fed to a ML model to predict near-surface O_3 concentrations.”

5. Line 231: How the meteorological fields be adjusted?

Response:

To minimize the impact of inconsistency in meteorology between CMIP6 models and the GEOS-Chem model, the CMIP6 meteorological variables in 2020–2100 are adjusted by their potential bias, characterized as the difference in their historical climatological mean (2014–2019) and MERRA-2. This adjustment process also removes inconsistency in the initial conditions of meteorological fields used by different CMIP6 models. We have added the descriptions in the manuscript.

6. Usually ground-level O_3 concentrations (225, 353 lines), Surface O_3 concentrations (271 lines), near-Surface O_3 concentrations (24 lines and so on) were included in the article, how do these concentrations differ? If all of them refer to the same meaning, please use the same name.

Response:

Thanks for suggestion. We have now updated all the O_3 concentrations as near-surface O_3 concentrations.

7. As a lot of methods and data are used, I suggest to draw a structural block diagram to more clearly express the structure and ideas of this article, including how to assimilate.

Response:

We thank the reviewer for this suggestion. We have now added a structural block diagram in the revised manuscript.

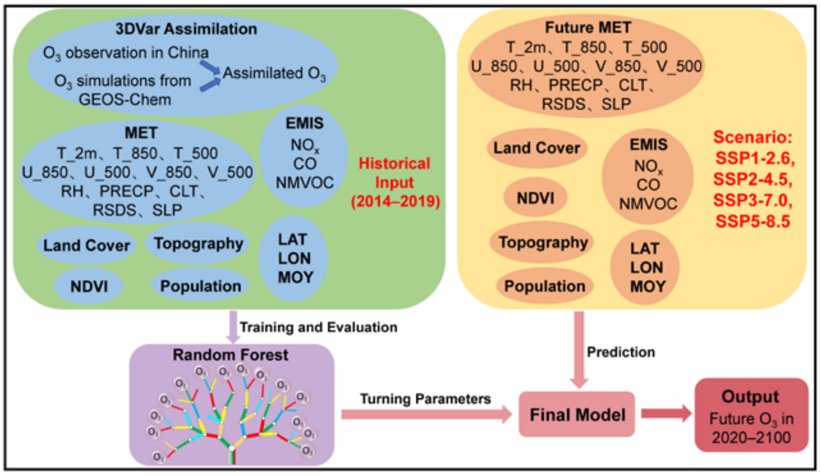

Figure 1. The structure and specific schematics for predicting future near-surface O_3 concentrations under four scenarios based on the machine learning (ML) method.

8. How are Importance scores of independent variables (meteorological parameters, emissions, land use, topography, and population density) used in the ML model for predicting future near-surface O_3 concentrations over Asia obtained?

Response:

In this study, the importance score of each variable is obtained from the training of machine learning model. To retrieve the relative importance scores of each input feature from the Random Forest model for predicting near-surface O_3 over Asia, we use the "feature_importances" attribute in Scikit-learn collection, which called Gini importance. A greater Gini importance implies a greater influence of input feature on the target variable. We have now elaborated the importance scores more in the revised manuscript.

The measurement is based on the calculation of Gini impurity when a variable is chosen to split a node. For each variable, the sum of the Gini decrease across every tree of the forest is accumulated every time that variable is chosen to split a node, and then averaged over all trees of the ensemble, expressed as:

$$\mathrm{GI}(\omega) = \sum_{n=1}^{N} \omega_n (1 - \omega_n) = 1 - \sum_{n=1}^{N} \omega_n^2$$

where *n* represents the number of the categories (*N* = 1, ..., *n*) and ω_n represents the sample weight of each category. The importance of one feature (X_j) on node *m* is that the GI changes before and after node *m* branching as follows:

$$\Delta\mathrm{GI}_{jm} = \mathrm{GI}_m - \mathrm{GI}_l - \mathrm{GI}_r$$

where GI_l and GI_r represent the GI of two new nodes after branching. The importance score for one feature (IS_j) in the then extra trees with *k* trees (*i* = 1, ..., *k*) is calculated as:

$$\mathrm{IS}_j = \sum_{i=1}^{k} \Delta\mathrm{GI}_{ij} = \sum_{i=1}^{k} \sum\nolimits_{m \in M} \Delta\mathrm{GI}_{jm}$$

Where ΔGI_{ij} represents the importance of X_i in the ith tree when the node of feature X_i in decision tree *j* belongs to set *M*. Finally, an additional normalization approach is performed to all obtained importance scores for each feature.

9. The authors could add some discussion and some references. such as:

Li, M., et al. (2021). Rising surface ozone in China from 2013 to 2017: A response to the recent atmospheric warming or pollutant controls? AE, 246. doi:10.1029/2021JD036393, https://doi.org/10.1016/j.scitotenv.2021.150338.

Lu, X., et al. (2020). Rapid Increases in Warm-Season Surface Ozone and Resulting Health Impact in China Since 2013. EST&L, 7, 240-247.

Response:

Thanks for the good suggestion. We have added the references and discussion as follows. “Last but not least, the near-surface O_3 have increased rapidly in China since 2013 owing to both precursor emission changes and atmospheric warming (Li M. et al., 2021), which significantly affect human health (Lu et al., 2020) and also requires further studies.”

References:

Li, K., Jacob, D. J., Shen, L., Lu, X., De Smedt, I., and Liao, H.: Increases in surface ozone pollution in China from 2013 to 2019: anthropogenic and meteorological influences, Atmos. Chem. Phys., 20, 11423–11433, https://doi.org/10.5194/acp-20-11423-2020, 2020.

Li, K., Jacob, D. J., Liao, H., Qiu, Y., Shen, L., Zhai, S., Bates, K. H., Sulprizio, M. P., Song, S., Lu, X., Zhang, Q., Zheng, B., Zhang, Y., Zhang, J., Lee, H. C., and Kuk, S. K.: Ozone pollution in the North China Plain spreading into the late-winter haze season, P. Natl. Acad. Sci. USA, 118, 1–7, https://doi.org/10.1073/pnas.2015797118, 2021.

Li, M., Wang, T., Shu, L., Qu, Y., Xie, M., Liu, J., Wu, H., and Kalsoom, U.: Rising surface ozone in China from 2013 to 2017: A response to the recent atmospheric warming or pollutant controls?, Atmos. Environ., 246, 118130, https://doi.org/10.1016/j.atmosenv.2020.118130, 2021.

Lu, X., Zhang, L., Wang, X., Gao, M., Li, K., Zhang, Y., Yue, X., and Zhang, Y.: Rapid Increases in Warm-Season Surface Ozone and Resulting Health Impact in China since 2013, Environ. Sci. Technol. Lett., 7, 240–247, https://doi.org/10.1021/acs.estlett.0c00171, 2020.

Qian, J., Liao, H., Yang, Y., Li, K., Chen, L., and Zhu, J.: Meteorological influences on daily variation and trend of summertime surface ozone over years of 2015–2020: Quantification for cities in the Yangtze River Delta, Sci. Total Environ., 834, 155107, https://doi.org/10.1016/j.scitotenv.2022.155107, 2022.