Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2020-243-SC1, 2021

Open Access Earth System Science Data Discussions

Interactive comment on "Long-term trends of ambient nitrate (NO_3^-) concentrations across China based on ensemble machine-learning models" *by* Rui Li et al.

John Wilson

john_whu@yopmail.com

Received and published: 18 January 2021

In this study, the authors provided an ensemble model by stacking RF, GBDT, and XGBoost to acquire monthly ambient nitrate concentrations over China. Generally, the topic of this study is very interesting since national-scale products of ambient chemical components are of great importance. However, the adoption of datasets in this paper is not convincing. To be specific, the spatial distribution of ground sites (only 32) is very sparse, which means that they do not cover most of the study area. How could the authors ensure the accuracy of the whole study area using these ground truths? I wonder how to validate the result in the regions without ground measure-

ments, such as Tibet. Such regions are numerous in this study. Besides, GEOS FP (http://wiki.seas.harvard.edu/geos-chem/index.php/GEOS-FP) can provide global 3-hour ambient nitrate concentrations at a similar spatial resolution. What is the main contribution of this study compared to GEOS FP? The authors need to justify the above issues in detail. Some minor comments are listed below. 1. Section 3: Why did the authors select these three machine learning methods for stacking? What if the authors only chose two of them? 2. Fig. 2: I notice that this flowchart is very similar to those in the authors' previous publications (e.g., Developing a novel hybrid model for the estimation of surface 8h ozone (O3) across the remote Tibetan Plateau during 2005–2018). Maybe a new style would be better. 3. Line 206: The parameters for RF, GBDT, and XGBoost are not given. Please provide them. 4. Fig. 3: XGBoost shows the worst performance, which is unusual. The authors need to provide some discussions. Did this happen in other literatures? 5. Fig. 5: Some point-shaped high values exist in the results (e.g, Northern China), which look like noises. Is this spatial distribution correct?

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2020-243, 2020.