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.

General comment:

Based on surface observation, satellite product, meteorological data, land use types and other covariates, this research has developed a monthly NO3- dataset at 0.25° resolution over China during 2005-2015, using ensemble machine-learning models. The long-term NO3- dataset is valuable for the air pollution control work in China. Compared with previous products, this new method also shows better performance in predicting accuracy and inspires peers to utilize interdisciplinary approaches to solve environmental issues. However, I suggest some modifications are necessary before being accepted. My comments are as below:

1. More attention is needed to the details in scientific writing. For example, the abbreviation should have an explanation when it appeared at the first time, NNDMN (Line 127), ERA-Interim (Line 150)ïijŇAOD (Line 283) etc. Please check the manuscript carefully. Besides, U/V wind components are accurately latitudinal and meridional wind components in Line 153.

Figure 3, the name of the color bar is missing.

2. L134, The detection limit of particulate NO3 concentration over China is said to be 0.05 μg/m3 which is unlikely to be true. The authors needs to check for it.

3. Line 248-253 and Figure S3, what is the purpose to discuss the relationship between observed NO3- concentration and other parameters using Pearson correlation analysis?

3. In section 2.3, more details are needed about the method to assimilate the socioeconomic data (GDP, population etc.) every five years to the seasonal or monthly resolution.

3. Line 165, the importance values have been applied to select the independent variables to do the NO3- prediction in this research. The results of the importance values from the ensemble model and the selected variables are expected in this manuscript or in the supplement. In the same way, the regression coefficients (A, B, C, mentioned in Line 210) determined by the MLR model are expected, too. Because they are crucial parameters of the ensemble model.

4. To validate the excellent prediction performance of the ensemble model, detailed information about the observed data in Figure 4 are suggested to be labeled, such as sampling site, month, year etc. Data from sites far from the selected training sites and covering key areas will be more convincing and preferred.

5. A comparison also to chemical transport model results such as GEOS-FP (http://wiki.seas.harvard.edu/geos-chem/index.php/GEOS-FP) modeled nitrate will be

very valuable.

C3