Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2019-173-RC1, 2019

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Assessing the impact of Clean Air Action Plan on Air Quality Trends in Beijing Megacity using a machine learning technique" *by* Tuan V. Vu et al.

Anonymous Referee #1

Received and published: 21 March 2019

This manuscript sets out to attribute changes in observed air pollution levels at monitors in Beijing with specific actions taken by the Chinese government under its 2013 Action Plan. The authors apply a Random Forest technique to separate influences of meteorology from daily measurements of multiple measurements. They then attempt to link changes in concentrations of the various pollutants to specific actions on individual source sectors.

The major issue I see with this manuscript is in the lack of detail in model descriptions, evaluations, and data sources, all of which are lacking throughout the manuscript. I've laid out specific concerns below. Overall, a general lack of detail makes it difficult to

fully trust the results and conclusions about the effectiveness of the various control actions.

Specific comments:

Abstract: "improved a novel machine learning-based random forest technique". How?

Line 75: "But they usually gave a poor fitting, suggesting a poor performance of the KZ filter model, or did not allow us to investigate the effect of input variables in neural network models (therefore it is referred as a "black- box" model): A poor fit does not necessarily reflect a poor performance; performance is dictated by the goals of the modeling, whereas fit is a measure of the ability to reproduce training data.

Line 79: Again, "performance" here is not defined. I recommend

Line 79: Should mention the increased propensity of over-fitting with these models for completeness

Line 110: Recommend showing in Figure 1 that you used 70% of the data for training, 30% for model evaluation. In addition, I recommend reading Oreskes et al. (1994) for distinction between evaluation/validation on environmental datasets. Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. Science, 263(5147), 641–646. Line 95: "press." has a period, whereas the other abbreviations do not.

Line 104: it => its

Line 125: With a holdout analysis, there are many comparisons to be made beyond Rˆ2 that tell us more about model fit. Many of the studies cited in the introduction include detailed evaluations, including with slope, intercept, and root mean square error. These should be included at the very least. There may be still other metrics that are informative for the evaluation in this particular application.

Line 128: sample => samples

Line 140-150: Was this a separate random forest model from the initial model described in the “Random Forest (RF) model development” section?

Line 152: This statement (“only either data (MET data) sets were re-sampled”) directly contradicts the statement in the paragraph above.

Lines 162-8: Please state what you are regressing using the Theil-Sen estimator

Lines 207-210: The conclusion that this evidence indicates a robust model requires more exploration. What about the meteorology from 1998-2013 would result in the 2μg m 3 increase in detrended PM2.5 in 2017?

Line $\sim$220: This could also indicate that formation/deposition/reaction of PM10 and NO2 are affected differently than the other pollutants. From the evidence provided, it is difficult to fully embrace the claim that PM10 and NO2 were affected by sources that were not controlled. Figure 2 presents no evidence relating to dust events that I can see.

Line 223: Figure 3 does present differences between urban/rural/suburban, but there is no information on how many sites and their location. I recommend including a map so that distance to roadways/industries/spatial representativeness can be determined

Line 230: This evaluation is difficult to interpret. Are the average WRF-CMAQ values calculated in the same grid cells as the monitors? Presumably, CMAQ modeling used emissions for year 2017 (state this explicitly if so), what about years 2013 and 2016 make them reasonable comparison years for detrended PM2.5?

Line 241-247: For model evaluation, I recommend including the recommended statistics from extensive publication on appropriate evaluation approaches like in Emery et al. 2017, Henneman et al., 2017, and Dennis et al., 2010. Emery, C., Liu, Z., Russell, A., Talat Odman, M., Yarwood, G., & Kumar, N. (2016). Recommendations on Statistics and Benchmarks to Assess Photochemical Model Performance. Journal of the Air & Waste Management Association. Dennis, R., T. Fox, M. Fuentes, A. Gilliland,

S. Hanna, C. Hogrefe, J. Irwin, S.T. Rao, R, Scheffe, K. Schere, D.A. Steyn, and A. Venkatram. 2010. A framework for evaluating regio- nal-scale numerical photochemical modeling systems. J. Environ. Fluid Mech.10:471–89. doi: 10.1007/s10652-009-9163-2. Henneman, L. R., Liu, C., Hu, Y., Mulholland, J. A., & Russell, A. G. (2017). Air quality modeling for accountability research: Operational, dynamic, and diagnostic evaluation. Atmospheric Environment, 166(2017), 551–565.

Line 259: Please define the term “based line”

Line 280: This contradicts the statement above that buffered changes in NO2 are due exclusively to sources that were not controlled

Line 330: Please elaborate on which data would improve this study.

Figure 2: I recommend including separate plots for emissions and concentrations. Plots with two vertical axes can lead to information manipulation (it is not clear, for instance, why an SO2 concentration of 40ppb corresponds to an emissions level of 2 kilotons). It would be useful to include correlations between detrended emissions and concentrations. Further, I recommend extending all vertical axes to values of 0.

Figures S4 and S5 require more description. What are Variable Importance and Variable Interactions?

Where is the emissions data from? What locations?

I recommend moving much of the information on the regulations from the supplement to the main text body.

I recommend using consistent language to refer to the weather normalized concentrations. At points in the manuscript, figures, and tables, these values are referred to as detrended, “Nor.”

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-173, 2019.