

Interactive comment on “Machine learning for observation bias correction with application to dust storm data assimilation” by J. Jianbing et al.

Anonymous Referee #1

Received and published: 12 June 2019

The main purpose of this paper is to explore the impact of observational biases on dust assimilation. These biases exist because surface PM₁₀ contains both dust and non-dust contributions. However, the authors discuss two methods to separate out dust PM₁₀ and then use it in a data assimilation scheme. Novel is their use of a machine learning technique (a neural net called LSTM) to do this. The authors show that, if not corrected, such biases can seriously (and negatively) affect dust forecasts but that use of the LSTM allows for significantly better forecasts. The paper is fairly technical and may be better suited to GMD but is appropriate for ACP too. The paper is well-written, properly organised with good graphics. Its conclusions are warranted by the data presented.

My main comment would be to add more detail to the section on the LSTM bias cor-

C1

rection.

p 2, l 21: please consider adding "Inverting East Asian Dust Emissions of a Severe Winter Dust Event Using the Ensemble Kalman Smoother and Himawari-8 AODs" by Tie Dai, currently in review

p 3, l 25: it is a pity the authors do not compare the correction scheme in Eq (1) also. Maybe it is too much work to do the full assimilation work but at least a comparison of non-dust PM₁₀ estimates should be possible?

p 7, l 5: What sort of uncertainty in emissions results from this: 1%, 10%, 100%? What is the spatial structure of delta f? Does i vary freely from grid-box to grid-box or is some correlation imposed?

p 10, l 10: why are PM₂₅ measurements at nearby (!) sites used? The authors already used PM₂₅ from the AQ network. What sort of error do they think is introduced by using measurements taken at different locations?

p 10, l 20: I would not say the "actually are" non-dust PM₁₀ but rather that they may be assumed to be very close to non-dust PM₁₀

p 10, l 22: "measurements at the nearby sites are considered as the essential input". Maybe the authors can clarify?

p 10, Sect 3.2: since this is arguably the new contribution by this paper, more details seems required. E.g. 1) how large where training and testing datasets; 2) how where they separated?; 3) how where LSTM's hyper-parameters set? 4) Please provide error statistics for training & testing. I am familiar with LSTM use in time-series but in this paper the authors also claim it is useful for spatial variations (p 10, l 1). Was the LSTM adapted to allow for this? It seems random forests might equally be useful. Did the authors consider this? Random forest is easier to train and would also provide feedback on what are the most relevant features for non-dust PM₁₀ prediction.

p 11, l 22: why use a t=1 forecast to estimate non-dust PM₁₀. Can't you train the LSTM

C2

to estimate non-dust PM10 at $t=0$?

p 16, l 1: Would R vary due to bias correction scheme? After all that scheme will affect remaining errors in dust PM10. In particular, I wonder whether the LSTM scheme (which trains on all data throughout a domain) would not lead to off-diagonal elements to be non-zero (i.e. correlations amongst errors at nearby stations)? Please discuss.

p 16, l 5: Can the authors discuss the largest sources for the errors of 10%? Measurement error? non-dust correction? Spatial representation? I see the authors suggest (!) representation is the most important but they give no reason for this. non-dust correction errors are of course available from the LSTM training.

p 16, l 15: "and thus the representation uncertainty" I do not agree with this. The representation error is the difference between a point measurement and a grid average. Here the authors consider variation among point measurements, which is likely to over-estimate representation error. Also, representation error will can show significant variation depending on location of point measurement in grid box, see e.g. Schutgens et al. ACP 2016.

p 16, l 16: Can the authors surmise why relative standard deviation (std/mean) is almost constant for non-dust and dust event? Of course, PM10 levels increase but wouldn't one expect a smoother field in case of a dust plume (long range transport) vs local pollution?

p 16, l 27: Would the authors not expect the representation error in an Urban area to be significantly larger than in the countryside? Also, shouldn't representation error really be calculated for the dust PM10, instead of the total PM10?

Interactive comment on Atmos. Chem. Phys. Discuss., <https://doi.org/10.5194/acp-2019-298>, 2019.