

Interactive comment on “PM2.5 / PM10 Ratio Prediction Based on a Long Short-term Memory Neural Network in Wuhan, China” by Xueling Wu et al.

David Ham

david.ham@imperial.ac.uk

Received and published: 25 October 2019

This is a comment concerning the compliance of this manuscript with GMD standards for code and data availability. The issues raised here must be addressed before a revised manuscript can be published.

This is a slightly unusual manuscript by GMD standards, because the model is a trained neural net using off-the shelf technology provided by Keras. There is absolutely no issue with this scientifically, but from a code and data availability perspective, it is necessary to read our rules in the spirit in which they are intended: which is that all of the data needed to reproduce the manuscript needs to be archived.

C1

The issues are:

1. Preprocessing code is missing. The manuscript details extensive preprocessing which was applied to the data. The code and data presented needs to represent all stages of the work, not just the final neural net. Please include the raw data and the code which preprocesses it to create the training set.
2. Evaluation code is missing. The code used to conduct the evaluation experiments and in particular to produce tables 5, 6, and 7. The interested reader should be able to run a script and have the evaluation occur, resulting in the data in those tables. This is a critical part of the reproducibility of the paper.
3. Missing spreadsheets. The Python script provided loads two Excel sheets of data. These sheets are not in the archive, so users can't actually repeat your work. The spreadsheets need to be included (I presume they contains the same data as the supplied PDF, but that doesn't really help someone wanting to run the code).
4. The work is critically dependent on the LSTM implementation in Keras. It is therefore necessary to specify the exact version of Keras which was used to undertake the work, in case a different version produces different results.
5. Don't use URLs to refer to data on Mendeley. The record on Mendeley even tells you how to cite them. You need to put a reference to the data in your bibliography, including the DOI, and cite that from the code and data section.