

(2018JD029383)
Reviewer comments

**Model evaluation of high-resolution urban climate simulations:
using WRF/Noah LSM/SLUCM model (Version 3.7.1) as a case study**

Recommendation: Accept with major revisions.

Summary comment

In this manuscript the authors aim to propose a methodological framework for the evaluation of urban climate simulations. The framework is outlined and then tested in high-resolution urban climate modelling simulations over an area encompassing two big cities, Shenzhen and Hong Kong. The study addresses an important problem that is often overlooked in the urban climate modeling community: the model evaluation. However, the manuscript comes across as a rather superficial and extensive model evaluation of WRF, rather than as a reference for a new model-evaluation framework (see MC1 to MC3). The paper is also poorly written and several sentences are hard to understand (see MC4). The topic of the paper well fits within the scope of GMD, but I would consider it for publication only after major/substantial revisions are performed in line with the MCs below.

Major comment

1. General MC. This work comes across as a rather descriptive WRF model evaluation rather than as a new model-evaluation framework. The use of PDF and PSS is useful, but I find it is a bit exaggerated to say that a new framework was proposed because these quantities were considered in addition of standard descriptive statistics. This especially in view of the few words spent on the PSS theory in section 2.2 and on the extensive but rather superficial comments made in the model to observation comparison section.
2. Section 2.2: Here I would justify more thoroughly why the authors propose to use a PDF and PSS coefficient when compared to other (perhaps more sophisticated) methods. I would also love to see some physics-based or theoretical derivation for admissible error bounds for given quantities, and a discussion about the strengths and limitations of the proposed framework.
3. I would consider reducing the number of figures and comment more thoroughly.
4. The paper is poorly written and requires substantial revision. Specifically, the authors sometimes use technical terms very loosely (see e.g. mc 8, 9, 10), several sentences are hard to read or understand, and often statements are not supported by proper referencing (see e.g. mc 12, 13). Furthermore, I have encountered several typos and repetitions.

Minor comments

1. P1L27. Consider shortening this sentence.
2. P1L31. Urban climate, and
3. P2L5. “is more sensitive to the inadequacies of the atmospheric model “ -- provide citation to support such a statement

4. P2L6. “and the quality of input data” -- provide citation to support such a statement
5. P2L10-27. I am glad the authors provided evidence from existing literature.
6. P2L30. “wasn’t provided in the previous literatures” -> was not provided.
7. P2L30. “It is especially a research gap in 30 urban climate modelling community to proposing a systematic framework and methods for model evaluation.” – please rephrase
8. P3L32. “Perspectives” -> “periods”?
9. P3L33. Why direct? Please justify the use of each word, it seems to me the English should be improved
10. P7L29-30. Can you expand and justify why this is the case? Why the adjective “natural” ?
11. P733-34. Related to the previous comment: why is it common sense? Please expand..
12. P10L40. The atmospheric model produces the fine atmospheric features 40 which do not exist in the original meteorological data. – I do not understand what the authors are referring to. Please expand.
13. P11L5. What do the authors mean here with “model evaluation”? Please expand.
14. P11L41. As it stands to me it does not come across as a sophisticated technique, but a rather as simple approach to evaluate model performance (i.e., look at departures between PDFs between model and observations).