Response to Reviewers

[Cover Letter]

Dear Editor,

We appreciate you and the reviewers for your precious time in reviewing our paper and providing valuable comments. It was your valuable and insightful comments that led to possible improvements in the current version. The authors have carefully considered the comments and tried our best to address every one of them. We hope the manuscript after careful revisions meet your high standards. The authors welcome further constructive comments if any.

Below we provide the point-by-point responses. All modifications in the manuscript have been highlighted in red.

Sincerely,

Bo Huang, PhD

bohuang@cuhk.edu.hk

Professor, Department of Geography and Resource Management

The Chinese University of Hong Kong

Response to Reviewer 1

[General Comment] The manuscript has been well improved, and I think most of my comments have been addressed with either new analysis or necessary discussions.

Response: Thank you very much.

Some minor revisions for the authors to consider:

[Minor Comment 1] P1L20-21 "...ambiguous or arbitrary to some extent, which may still lead to some seemingly new findings, but these findings may be scientifically misleading." Please consider rewriting this sentence.

Response: Thanks for your kind reminders. We revised the sentence as follows:

"Studies without systematic model evaluations, being ambiguous or arbitrary to some extent, may still lead to some seemingly-new but scientifically-misleading findings."[Pg1, Ln18-19]

[Minor Comment 2] P1L23 "urbanising city of Shenzhen, China...". Does the study area also include Hong Kong?

Response: Thanks for your kind reminders. We revised the sentence as follows[Pg1, Ln22]:

"To tackle these challenges, this article proposes a methodological framework for the model evaluation of high-resolution urban climate simulations and demonstrates its effectiveness with a case study in the area of Shenzhen and Hong Kong, China."

[Minor Comment 3] P2L30 Change "wasn't" to "has not been". Please also check through the text. For example, P11L19 "isn't".

Response: Thank you very much for the reminder. We have made revisions accordingly.

[Minor Comment 4] Please trimming Sect. 5.3.

Response: We moved Subsection 5.3 to Supplementary Material [Section S9].

Response to Reviewer 2

[General Comment] The revision highlights the importance of model evaluation and provides a reasonable approach for evaluating the results from urban climate simulations. Overall, the revision is well organized and clearly presented. In particular, Figure 1 is very nice and helpful to understand the framework. In my opinion this paper is worth to be published in GMD.

Response: Thank you very much for your previous comments that helped us improve this manuscript.

Minor suggestions:

[Comment 1] In Table 1, to use PDF & PSS, or probability density functions & Perkins Skill Score, instead of probability density functions & PSS;

Response: Revised accordingly.

[Comment 2] It would be better to put Table 4 & Figure 2-21 in the supplementary material.

Response: Revised accordingly.

Response to Reviewer 3

[General Comment] The overall intention of this submission is a good one. In general I agree that more attention should be paid to model evaluation in urban climate modelling work. However, this submission is relatively formulaic and it misses a number of key nuances that render the overall model assessment somewhat unconvincing and moreover unlikely to be applied by other researchers. I appreciate the authors' intent and believe more rigorous model evaluation is critically needed in the field. However, this submission does not currently approach GMD standards in my view.

Response: Thank you very much for agreeing with us to the intention of this manuscript. We have read your comments carefully and tried our best to address them one by one, especially in terms of providing a more rigorous model evaluation. We hope that the manuscript has been improved towards GMD standards after this revision.

[Comment 1] English grammar requires improvement throughout.

Response: We went through the entire manuscript to eliminate grammatical mistakes.

[Comment 2] There are too many tables. In general the presentation/communication requires improvement.

Response: Thank you for the nice reminder. We combined Tables 2 and 3 into one table (Table 2). Moreover, we moved Subsection 2.4 to Supplementary Material [Section S5] to cut down the number of tables.

[Comment 3] Figure captions are in general insufficiently detailed. For example, it is not clearly indicated whether central tendencies and variation in several figures are spatial or temporal.

Response: Thank you for your nice reminder. We revised most of the figure captions to make them clearer.

[Comment 4] P3L27-33: There is no mention of how spatial patterns are evaluated here. In general, it is not fully clear how spatial patterns produced by urban climate modelling at the city-regional scale is evaluated, which is a prime intention of these models.

Response: Thanks for your question.

At the spatial dimension, climatological studies usually focus on three scales, including local (less than 10^4 km^2), regional (from 10^4 to 10^7 km^2) and global (greater than 10^7 km^2) scales (Intergovernmental Panel on Climate Change, 2012). The similarity between the modelled and observed variables' spatial patterns is indeed a major content in model evaluation of

regional and global climate, especially, the spatial difference of precipitation belt and atmospheric circulation. In the previous literature, there were not many papers on the methods of spatial pattern comparison for local climate simulation, which is a research gap that we focused on.

In urban climatology, the most important spatial pattern of meteorological variables is the difference between in urban and non-urban areas. Therefore, at the spatial dimension, we evaluated the model using the temporal comparison of spatial variation in the whole year in urban and non-urban areas.

We added these contents into Subsections 2.2 [Pg4, Ln11-13], 4.1 [Pg6, Ln28 – Pg 7, Ln2; Pg10, Ln15] , 4.3 [Pg11, Ln27-29; Pg12, Ln3-4], 4.5 [Pg12, Ln24-27; Pg12, Ln34]and 5.4 [Pg15, Ln15-20]. Moreover, we added figures about the comparison of meteorological variables between urban and non-urban areas into the Supplementary Material (Figures S7, S8, S14, S15, S19, S20, S24, and S25).

[Comment 5] P4L9: 2m air temperature is diagnosed in WRF-SLUCM, and I believe that surface temperature is as well. 10-m wind is also diagnosed.

Response: Thank you very much for pointing this out. We revised the sentence as follows:

Pg4, Ln16-17: " Moreover, we chose 7 meteorological variables for the comparison, including 2-meter air temperature, surface temperature, 10-meter wind at u direction, 10-meters wind at v direction, accumulated total cumulus precipitation, accumulated total grid precipitation and 2-meter relative humidity, because these variables are the critical variables in the prognostic and diagnostic equations in the NWP model."[Pg5, Ln3-7]

[Comment 6] Table 3: TSK should not be compared to a MODIS land surface temperature product. TSK includes the emission of all building walls, whereas a satellite does not view longwave emission from any walls (nadir view), or only from walls of select orientations (off-nadir view). Many authors make this mistake; if this is to be a rigorous model evaluation paper that sets an example, it should not be made here. Instead, you must extract the surface temperatures of roads, roofs, and rural area and weight them yourself for each grid square. This is not necessarily trivial and requires adding an "h" to the appropriate variables in the WRF Registry as well as outputting the rural surface temperature (which may require adding a new variable and re-running the simulations).

Response: The reviewer missed a piece of important information - the scan angle of the MODIS satellite is ± 55°. At the end of a scan line, the scanning angle being 55° means that the land surface temperature retrieved from MODIS imagery is not an area-weighted mean value temperatures of roads, roofs, and rural area within a particular pixel. Actually, a MODIS land surface temperature is a result of the inverse calculation based on the longwave radiation through the atmosphere received by satellite according to the theory of blackbody. A MODIS land surface temperature is a manifestation of the surface synthetic radiation brightness temperature. In addition, in the land surface process, TSK is calculated iteratively

according to the energy balance which involves longwave radiation, shortwave radiation, sensible heat, and latent heat, and accordingly, the final TSK value is also a manifestation of the surface synthetic radiation brightness temperature. Although there are some differences between TSK and the brightness temperatures observed by satellites, they describe relatively similar physical quantities. Therefore, we used TSK to compare with MODIS land surface temperature. We also added more explanations into Subsection 2.3 [Pg5, Ln12-19].

[Comment 7] Figure 3: These results will depend entirely on how many measurement stations are available and included, will they not?

Response: We don't think so. Figure 3 is a figure of descriptive statistic. The results may only be distorted if the number of measurement stations is too small. Our statistical results were based on data from 57 stations, so we don't think these results will be very sensitive to the number of measurement stations included.

[Comment 8] P12L2-5: I disagree. I think they are both useful. In particular, with the use of the Perkins Skill Score, a climatological approach appears to be taken here. What is also important for urban climate modelling is that the temporal evolution of each variable is predicted accurately, particularly at diurnal time scales. For example, a measure of covariance between measurements and observation that includes many temporal data points within each cycle of forcing is important (e.g., hourly measurement-model comparison of air temperature would capture the covariance between measured and modelled diurnal variation).

Response: Thanks for your comment. We agree with you that temporal evolution is an important dimension of meteorological processes and should be included in the modeling evaluation. We did conduct comparisons of the temporal evolution, including comparisons of diurnal patterns in Figures 3, and comparisons of monthly patterns in Figure 2 and Figures S6, S13, S18 and S23 of Supplementary Material.

[Comment 9] Ultimately, I think model evaluation will always have a subjective element. How is the authors' method designed to be applicable to all regions, conditions, simulation durations and resolutions, etc?

Response: Thanks for your comment. Our study just focused on evaluating urban climate modelling. The purposes of model evaluations to different scale, conditions, simulation durations and resolutions is the same, which is to establish the trustworthiness of the modelling results. In the proposed framework, we demonstrate tools to compare both the descriptive statistics and the statistical distributions of the observed and simulated meteorological variables. We also demonstrate how to use the tools to compare the same meteorological attributes at different temporal aggregation levels, including daily and monthly/seasonal. For sure that case-specific adaptations have to be made before the proposed methods be applied to other regions, conditions, and resolutions, which is the same

for most models and methods. The simulation model itself has to be calibrated, with an enormous amount of efforts, to different areas as well. Of course, the proposed model evaluation methods are not perfect, but we are confident that our practice can serve well as a reminder and a guide that improves the urgently-lacking model evaluation in current practices of high-resolution urban climate modelling.

Response to Reviewer 4

[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.

Response: Thank you for your comments. We have gone through your comments carefully and tried our best to address them one by one. We hope the manuscript has been improved accordingly.

[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.

Response: Thank you very much for the comment. We didn't find a systematic methodological framework for urban climate model evaluation in previous literature. The PSS theory is a well-known method and really direct and easy to understand, so instead of explaining in details, we pointed the readers to Perkins et al. (2007) for more details.

[Major comment 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.

Response: Thanks for the comment. This paper is intended to suggest urban climate modelers conduct systematic model evaluations in urban-scale climatology modeling and

provide them a practical methodological framework. We think that the practical and easy-understanding methods are better than complicated and nonintuitive ones. Of course, our methodological framework is not perfect, we have specified the limitations in the discussions section which need to be improved by other scientists in the future.

[Major comment 3] I would consider reducing the number of figures and comment more thoroughly.

Response: To reduce the number of figures, we restructured Section 4. Moreover, we also moved many redundant figures to Supplementary Material and only kept few example figures in the manuscript. Please refer to the revised manuscript.

[Major comment 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.

Response: Thank you very much for the comment. We did our best to correct these mistakes.

[Minor comment 1] P1L27. Consider shortening this sentence.

Response: Thank you very much for nice reminder. We revised this sentence as follows:

"Recently, studies on urban climate have received growing attention. It is forecasted that there will be 66% of the world's population living in the urban area by 2050 (United Nations, 2014). The fundamental well-being of the urban population, such as their comfort and health, is directly and significantly affected by urban meteorological conditions, such as temperature, wind speed, and air pollution."[Pg1, Ln25-28]

[Minor comment 2] P1L31. Urban climate, and

Response: Thanks for your nice reminder. We provided the following citations to support this statement.

Dale, V. H.: The relationship between land-use change and climate change. Ecological applications, 7(3), 753-769, 1997.

Kalnay, E., & Cai, M.: Impact of urbanization and land-use change on climate. Nature, 423(6939), 528, 2003.

[Minor comment 3] P2L5. “is more sensitive to the inadequacies of the atmospheric model “ -- provide citation to support such a statement.

Response: Thanks for your nice reminder. We provided the following citation [Pg2, Ln6] as a support:

Warner, T. T.: Quality assurance in atmospheric modelling. B. Am. Meteorol. Soc., 92(12), 1601-1610, 2011.

[Minor comment 4] P2L6. “and the quality of input data” -- provide citation to support such a statement

Response: Thanks for your nice reminder. We provided the following citation [Pg2, Ln7] as a support:

Bruyère, C. L., Done, J. M., Holland, G. J., & Fredrick, S.: Bias corrections of global models for regional climate simulations of high-impact weather. Climate Dynamics, 43(7-8), 1847-1856, 2014.

[Minor comment 5] P2L10-27. I am glad the authors provided evidence from existing literature.

Response: Thanks for the comment.

[Minor comment 6] P2L30. “wasn’t provided in the previous literatures” -> was not provided.

Response: Revised accordingly.

[Minor comment 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

Response: Revised accordingly.

[Minor comment 8] P3l32. “Perspectives” -> “periods”?

Response: Thank you very much for your nice reminder. We considered what is the appropriate word for this context again. Finally, we changed it to “resolution” [Pg3, Ln30].

[Minor comment 9] P3L33. Why direct? Please justify the use of each word, it seems to me the English should be improved

Response: Thank you very much for your nice reminder. We revised the sentence as follows:

"In doing so, our instinct can decide whether the modelled results could replicate the temporal and spatial patterns in the observations or not." [Pg3, Ln31-32]

[Minor comment 10] P7L29-30. Can you expand and justify why this is the case? Why the adjective "natural"?

Response: We already provided an explanation in Subsection 4.1 as follows:

"In fact, the difference includes not only the modelling bias but also an essential difference between a 1-km grid spatial average value and a value of a point located in this grid. Moreover, the observation always locates in an open area, and thus, the observed 2-meters air temperature is the temperature of a point in the open area. The modelled 2-meters air temperature is a mean temperature of a 1-km grid which always includes some vegetation covered areas. It is a common sense that the point air temperature in the open area is always higher than its corresponding 1-km grid mean air temperature in the summertime."

Moreover, we reconsidered if the meaning of the term "natural gap" meets the context of this manuscript. Finally, we replaced it with "essential difference". The essential difference refers to the fact that model outcomes from the simulation models are average values of a grid, while the observations are point-based which only measures the meteorological conditions around the location of the monitoring station.

[Minor comment 11] P7L33-34. Related to the previous comment: why is it common sense? Please expand.

Response: Thank you. It's a good question. An observation station is always located in an open area – an area without the coverage of trees - and so the measured 2-meter temperature at an observation station is always higher than the modelled temperature since the modelled temperature is calculated as a mean value over a grid with vegetation coverage. We revised the sentence as follows:

"In the summertime, the point air temperature in the open area without coverage of trees is always higher than its corresponding mean air temperature of a 1-km grid with some vegetation coverage." [Pg10, Ln10-12]

[Minor comment 12] P10L40. The atmospheric model produces the fine atmospheric features which do not exist in the original meteorological data. – I do not understand what the authors are referring to. Please expand.

Response: Thank you very much for the question. Indeed, this sentence is not so easy to understand. We changed it as follows:

"The fine-scale details are constructed by a limited area atmospheric model which consists of physical components driven by the lateral boundary conditions of coarse-scale meteorological data and land surface forcing data (Lo et al., 2008; Hong et al., 2014). However, these details do not exist in the coarse-scale meteorological data (Hong et al., 2014)." [Pg14, Ln17-20]

[Minor comment 13] P11L5. What do the authors mean here with "model evaluation"? Please expand.

Response: It is a good question. Model evaluation refers to comparisons between the modelled variables and its corresponding observed ones. After modelling, a model evaluation should be conducted for establishing the trustworthiness to the modelling results because of the model incompleteness caused by the approximations and assumptions in the scientific mechanisms of the model even if the model was configured appropriately.

We added these contents into Section 1[Pg1, Ln34-37].

[Minor comment 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).

Response: Thanks for the comment. This paper was intended to remind (again) urban climate modelers of the necessity of conducting systematic model evaluations in urban-scale climatology modelling and reduce these ambiguous or arbitrary modelling practices. We also provided the urban climate modellers a practical methodological framework for model evaluation which we can not find in the previous literature. We replaced "sophisticated methodological framework" with "practical methodological framework" [Pg14, Ln41] in this sentence.