Response to Reviewer 1

[Cover Letter]

Dear Reviewer,

We appreciate you for spending time to review our paper and providing some valuable comments. It is 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 efforts to address every one of them. However, some revisions may still cannot meet your high standards. The authors welcome further constructive comments if any. We provided the point-by-point response first and will provide the updated version of the paper after proofreading complete.

Sincerely,

Bo Huang, PhD

bohuang@cuhk.edu.hk

Professor, Department of Geography and Resource Management

The Chinese University of Hong Kong

[General Comment] This study evaluates performance of the WRF model in terms of high-resolution urban climate modelling over an area encompassing two big cities, Shenzhen and Hong Kong. The chosen area of Shenzhen is heavily urbanized but only a small part of Hong Kong is urbanized. Perkins skill score is used as a major evaluation method throughout the evaluation. The authors argue that their study has proposed a methodological framework for evaluating model performance in high resolution urban climate simulation. I think this work is useful and has provided some information about high-resolution urban climate modelling applied to south China. I very much appreciate the authors' efforts to pursue this kind of modelling work. However, I feel that the manuscript in the current form cannot be accepted for publication. At a minimum, I would suggest some necessary revisions to make the paper publishable in the journal. But to engender a stronger paper, I feel that more extensive work might have to be done. I will leave it to the editor to decide whether such extensive work is required.

Response: The article is in pertinent response to the increasing presence of ambiguous or careless modelling practices in urban-scale climatology. It intended to state the necessity of

model evaluation of urban-scale climatology modelling, draw attention within the community of urban climate modellers, and be a kick-off in reducing these window-dressing-like modelling practices. Therefore, the purpose of this paper is for reminding the modellers the necessary of model evaluation in the urban climate modelling practices rather than helping the model developer to improve the model. Moreover, the modeller should conduct a systematic model evaluation to establish the trustworthiness of the new findings from an urban climate modelling because the model cannot be verified or validated. Furthermore, we reminded that the modeller should be cautious to conclude a quantitative findings because it is impossible to identify the natural gap, observation bias, and model bias in the difference between observations and its corresponding modelled results. To sum up, we are confident that this paper is important to urban climate modeller community because it points out the pain points which the existence of uncertainties of model affect the trustworthiness of the new findings and it is impossible to identify the uncertainties of model completely.

[Major Comment 1] The introduction should be reformulated with greater care. The authors should survey the literature more thoroughly. Only a few papers are mentioned in the introductory section. I suggest the authors give a good overview of the existing studies on the topic and point out the limitations of the past studies and challenges/ constrains. Identifying a gap or proposing a new method as well as outlining the contributions of the study is also helpful.

Response: We added some new related literatures in Section 1 to emphasize the importance of model evaluation in urban climate modelling, the fact that the modellers paid the less attention in their modelling practices and the values of this paper.

[Major Comment 2] The data and methodology section should be structured in a more logical way. I think the authors could place model description and experiment/model setup before evaluation method. Overall, both section 2 and 3 are a bit confusing. The introduction of the model is lacking. The authors should clearly articulate what has been done and how it has been done. This can aid the readers in understanding the experiment setup/design.

Response: We revised Section 2 for improving clarity and provided more information about model description and setup in Supplementary Material. Moreover, we will submit another paper for describing all details about our urban climate modelling practice which includes the suggestions for modelling process, the atmospheric model design, model setup, primary data processing method, and a quality assurance framework.

[Major Comment 3]

In section 2.1, more details about the new dataset developed by the authors should be offered.

Response: We provided more details about the land surface dataset in Supplementary Material.

The reasons for focusing on the simulations in the year of 2010 should be discussed.

Response: The reason is the data limitation that we only have the land surface data and observation data in 2010. We mentioned it in the revision of the paper.

In section 2.2, more details should be provided as to the four-day segment simulations.

Response: We provided more details about four-day segment in Supplementary Material.

Did the model read in restart files every four days to continue the simulation?

Response: No. Each four-days simulation segment is a separated simulation.

How may a different simulation strategy affect the modelling results?

Response: The different simulation strategies relate with the different spin-up method, which affect the modelling results. We added a small discussion about it in Section 2.2.

In section 2.3, instead of just giving two tables, I think more detailed descriptions of the data should be given. How are the comparisons between model output (grid points) and observations (stations) made? Representativeness of the observations and potential biases should be discussed. The authors should also indicate the reasons for choosing evaluation variables.

Response: We added more details about the modelling variables and observation data in Section 2.3.

[Major Comment 4] In section 2.4, no references are cited regarding the Perkins skill score. Is this a suitable method for this study? There should at least be some discussion. Authors should also discuss whether this method is suitable for all the variables evaluated in the study.

Response: We conducted a small discussion about the evaluation tools in Section 2.4.

[Major Comment 5] In section 3, choosing of the parameterization schemes needs discussion.

Response: We conducted a small discussion of the parameterization schemes in Supplementary Material.

[Major Comment 6] I think the authors should tune down many of their arguments throughout the paper to avoid overstating (e.g., P2L25-26). For example, I don't see any strong methodological framework being discussed and described in the text.

Response: We enhanced the description of methodological framework for supporting our statement. We add a subsection (2.3 A Methodological Framework for Urban Climate Model Evaluation) to describe more details about the methodological framework.

[Major Comment 7] I have the impression that the authors have been too obsessed with 'good results' when evaluating the model's performance. Discussing 'good results' and 'bad results' at the same time, in my opinion, is fair. It's perhaps more important to identify areas for improvements.

Response: This paper intended to state the necessity of model evaluation of urban-scale climatology modelling and provided a methodological framework of model evaluation to help the modeller to establish the trustworthiness of modelling results, and accordingly it focused on the modelling performance rather than the identifying areas for improvements in order to help the model developer improving the model. We added an explanation in Section 1 to emphasize the focus of this paper.

[Major Comment 8] The structure and writing are too repetitive in section 4. This is also true for the figures. The number of figures may be reduced.

Response: We did our best to rewrite Section 4. Moreover, we moved some figures to Supplementary Material for reducing the number of figures in the paper.

While the focus of the paper as stated in the paper is on the urban climate simulation, evaluation seems to be applied to also the vast rural regions. The authors should clarify this.

Response: Yes. The methodological framework of model evaluation also can be applied in the local scale climate simulation wherever in urban or non-urban areas. We added an explanation in Section 5.

I suggest the authors focus on the most important aspects of the urban climate simulation. I would suggest some points (see following) for the authors to consider and they should further develop a better evaluation framework.

Response: Thank you very much for your suggestions. We added a subsection (2.3 A Methodological Framework for Urban Climate Model Evaluation) to describe more details about the methodological framework, which included a graphical presentation of the framework, the grading guidelines for PSS and PDF of the difference, and theoretical explanation to the statistic tools applied in model evaluation.

-Some basic ability of the model such as spatial distribution temperature/precipitation and diurnal cycles of temperature must be assessed.

Response: The diurnal cycles of 2m air temperature had be assessed in Subsection 4.1. We provided more comparisons of meteorological variables between urban and nonurban areas in Supplementary Material.

- The weather and climate variability in the study area is strongly associated with the monsoon flow. So the investigation of the simulation of precipitation and temperature is rather important. Both the spatial distribution (not found in any of the figures in the paper) and temporal variability should be considered.

Response: We agreed that the climate variability in the study area is strongly associated with monsoon flow. However, the monsoon flow is a mesoscale meteorological behaviour, and accordingly, it is not associated the spatial distribution of precipitation and temperature in a local scale region. The spatial distribution of precipitation and temperature is strongly associated with the local land surface attributes. Therefore, we added some discussions in Section 4 about the relationship in the spatial distribution between 2m air temperature and land surface, also the relationship in the spatial distribution between precipitation and land surface. Moreover, we agreed that the seasonal variations in temperature and precipitation are associated with monsoon flow, especially precipitation. Therefore, we added some discussions in Section 4 about the relationship between the monsoon flow and the seasonal variation of precipitation, also between the monsoon flow and the seasonal variation of 2m air temperature.

In particular, the authors may identify some strong urbanization impacts on the precipitation (e.g., precipitation maxima) and temperature (e.g., urban heat island). The model's ability to capture these effects is essential.

Response: It needed the observation data before and after urbanization to evaluate the urbanization impacts on the precipitation and temperature. We cannot provide these evaluations because we don't have these observation data. However, we added some discussions in Section 4 about the relationship in the spatial distribution between 2m air temperature and land surface, also the relationship in the spatial distribution between precipitation and land surface.

In addition, simulation of sea breeze, wind distribution, boundary layer variability, and stability of the atmosphere should be examined.

Response: We agreed that the land-sea breeze exists in the coastal city, and accordingly we provided a discussion about the modelled land-sea breeze in Section 4. These modelled meteorological features (boundary layer variability and atmospheric stability) cannot be examined by the observation due to the unavailability of its corresponding observation data. It is meaningless that the modelled meteorological features examined without comparison with the observation. Therefore, we didn't provide the examination on the these two meteorological features.

The impact of urbanization on the air quality may also be discussed.

Response: This study focused on providing a methodological framework of urban climate model evaluation. The impact of urbanization on the air quality is another big topic which is out of the research scope in this study. Therefore, we didn't provide an discussion on it in this paper.

- The evaluation can be done separately for different seasons. The evaluation should focus on the most important aspects of urban climate/weather.

Response: Actually, the figures included the information of the monthly variations. We provided more details about the seasonal variations of the evaluation in Section 4.

- The scientific value can be enhanced if the authors can demonstrate how the model behaves in simulating the extreme precipitation events or heat wave/cold surge events, and How and to what extent these events may be related to the urbanization.

Response: Thank you very much for your suggestions. However, our study focused on reminding the urban climate modeller the importance of model evaluation in establishing the trustworthiness of modelling results and provided a methodological framework of model evaluation, and accordingly we didn't put too much effort on model performance on simulating the extreme events. In the revision, we added some discussions about the performance on simulation the extreme events on Sections 4 and 5.

- The model's performance between different regions in the study area and between rural and urban regions can also be compared.

Response: Thank you very much for your suggestions. We added more figures about the model's performance in urban and non-urban areas in Supplementary Material.

[Major Comment 9] The figures can be better designed and drawn. Captions of the figures should provide more information. The language could also be improved.

Response: Thank you very much for your suggestions. We did our best to improve the language and the captions of the figures.

[Minor Comment] Minor comments:

The authors should check carefully the use of words and sentences throughout the paper. I suggest some serious edits/revisions. I list only some of the examples. P1L15: add 'have' before paid. P1L26-29: Please split the long sentence. P1L37: place 'into account' immediately after 'take'.

Response: Thank you very much for your suggestions. We did our best to check the paper, corrected the language errors and rewrote the long sentences for improving the readability.