The authors would like to thank the referee for this thorough review. Despite being negative about the paper, we believe that the review is useful, as it raises a number of valid points and criticism that will improve a future version of the manuscript once they are taken into account. As we explain below, we believe that the decision of rejecting the manuscript is not justified enough by the reviewer's comments. If the editor approves it, we will take into account all the reviewer's comments in order to proceed with the rewriting of several paragraphs of the manuscript and a more in-depth interpretation of the results. In this message we will answer the main comments of the reviewer. A more complete answer corresponding to the modifications introduced will be provided after the discussion is closed.

This paper is aimed at showing the advantages of an off-line urban canopy model with respect to a regional climate model coupled to an urban canopy model for urban heat island studies.

Note that UrbClim is not exactly an offline urban canopy model (UCM). Apart from an UCM, UrbClim has a planetary boundary layer (PBL) parameterization too, and does carry out a simplified simulation of the lower part of the atmosphere (see section 2.3 of the manuscript and De Ridder et al. 2015). This is an important point that perhaps we did not stress enough in the manuscript, but that we will explain in more detail in the new version.

De Ridder, Koen, Dirk Lauwaet, and Bino Maiheu. 2015. “UrbClim – A Fast Urban Boundary Layer Climate Model.” *Urban Climate* 12 (June): 21–48. doi:10.1016/j.uclim.2015.01.001.

It is undeniable that there is benefit in the use of application-specific models such as UrbClim, particularly in terms of computational costs. In such circumstances, even if they both show similar results, the computational costs associated with the regional climate model make the use of UrbClim very attractive.

The computational cost is not just an added value, but a critical factor, because it allows the performance of multi-decadal simulations in large ensembles of cities at sub-kilometer resolutions. The cost of doing this with a Regional Climate model (RCM) is not possible to afford. Note that the computational cost is also the very reason of the existence of RCMs.

However, in order to justify the use of a standalone urban model, the authors compare completely different tools (i.e. a more like-to-like comparison would be to show benefits of Urb-Clim over the Single-Layer Urban Canopy Model, both running offline with the same boundary conditions) and do not fully acknowledge (only at the very end of the paper) that a regional climate model coupled with an urban model incorporate features that an offline model cannot (e.g. two-way interaction with surrounding circulation such as sea-breeze).

As we pointed out, UrbClim is not exactly an offline UCM. The paper aims to compare UrbClim with a widely used RCM such as WRF, and not to compare UrbClim with the UCM of WRF. The scope of the paper is applied and pragmatic, in the sense that it is not trying to find the best theoretical approach to model the Urban Heat Island (UHI). We try to show that in this case it is not necessary to run the primitive equations in the whole troposphere, together with the parameterizations of radiation, cumulus... etc., as WRF does, in order to reproduce a realistic UHI.

But what is more important, the comparison is to a large extent unfair because the authors claim a better representation of local temperatures when the Urb-Clim is driven by ERA-Forecast, but do not test the regional climate model running with the same large/mesoscale information.

This seems to be a major concern for the reviewer. A fair comparison between two completely different approaches is not fully possible. A regional model like WRF is able to generate his own mesoscale variability so, in principle, it should be able to compete with the UrbClim run nested in ECMWF forecast model, as the large-scale variability is very similar in the two global models used.

In fact, it could be argued that WRF has an advantage here, as its UCM and PBL are running with forcing fields at a 15 times better resolution than those used for UC-FC. These kind of issues make the fair comparison between the two approaches complicated (though the comparison is certainly not fair for the WRF UCM alone, but as mentioned, this is not the goal of the paper). See the following paragraph in the manuscript (L 178-185):

"As UrbClim is hardly able to generate internal variability, these results can be interpreted as a comparison between Urban Canopy+PBL models driven by ERA-Interim (70 km), ECMWF forecast (16 km) and WRF (1 km). Thus, differences in the results show the added value of the higher resolution in the ECMWF forecast model and WRF. However, note that the extra resolution of WRF (about 15 times higher than ECMWF) is not clearly improving the results. This is consistent with previous studies suggesting diminishing returns for added value in this resolution ranges (García-Díez et al., 2015)."

The problem here may be that WRF is more biased than the ECMWF forecast model, which is stunningly precise. So, that is why we say in the conclusions (L310-312): "From these results, it is reasonable to infer that the skill of UrbClim, and probably of other similar urban climate models, is constrained by the performance of the driving model, and particularly for variables that are important for the UHI, this is, wind speed and cloudiness". This holds also for WRF-UCM. If driven by ECMWF forecast model, WRF-UCM may be able to perform as well as UC-FC. But testing this is out of the scope of the paper.

There may be other RCMs showing better results but, as WRF is currently widely used in this kind of studies about the UHI effect, we think that the results of the paper are relevant. On the other hand, results show that relatively high resolution (16 km) forcing fields are needed in order to remove UrbClim biases, so an RCM may still be needed to intermediate between the GCM and UrbClim in cities like Barcelona. This needs to be explained in the manuscript, and we acknowledge that the overall discussion of the results needs to be improved.

Finally, the authors do not provide any explanation of what processes are better represented in UrbClim that make it perform better than the regional climate model?

Addressing biases in climate models, especially in complex models such as WRF, is difficult and seldom evaluation papers in the literature manage to offer a clear explanation of the biases found. Here we focus on the practical problem of producing high resolution simulations of the temperature at a city level, including the UHI. We show that a simplified approach, essentially running the UCM plus the PBL, produces comparable results to an RCM, using much less computational power, and reaching x4 resolution (250m Vs 1 km). We believe that is a result worth publishing, without needing to analyse in deep the biases found.

In my opinion, the starting point is not correctly posed and the authors do not adequately support their conclusions with a rigorous analysis. I agree with the authors that for this particular application, UrbClim might present advantages over a Regional Climate Model coupled with an urban model, but I don't think the authors provided enough evidence for that.

We think that the several parts of the manuscript need to be re-written in order to clarify the points that we commented above. Stating that UrbClim is "better" that the RCM in terms of performance is not the goal of the paper, and perhaps is going too far, given that the reanalysis run does have biases comparable to WRF, or clearly larger in the case of the wind.

In my opinion, the authors could make additional experiments, perform a like-to-like and more in-depth comparison, with possible reasons as to why UrbClim outperforms the RCM.

We believe that additional experiments are not necessary, but a re-writing of some parts of the paper and clarifications of the points explained above. Also, as mentioned above, explaining WRF biases would require a separate study, and it is out of the scope of the present work.

In that case, they should also mention that RCMs are a tool design to conduct atmospheric research and therefore have a wider range of applications, and this is the reason why they are selected over offline and faster models.

This will be mentioned in the new version of the manuscript.

As it is, I am unsure the paper makes a scientific or model development contribution worth publishing. Perhaps including the additional analyses suggested above could lead to a paper that is adding to the current knowledge. In addition to some general comments, I have also suggested some specific and technical comments aimed at improving a future version of the manuscript. Therefore, I would not recommend this paper for its publication at Geoscientific Model Development.

We defend that the study presented in the paper is original and relevant in several ways: Geographical location, interesting climate (with a strong sea-breeze), length of the simulation (five months), that enables a robust evaluation, and comparison of UrbClim (with two driving models) and WRF. Thus, it is a contribution worth publishing. The reviewer mentioned some issues with our interpretation of the results that need to be addressed and clarified in the manuscript.

Answer to general comments:

1. We address this issue, these explanations mentioning the internal variability will be improved in the new version of the manuscript.

2. In the new version of the manuscript we will add more details about WRF configuration and to elaborate more on the differences with UrbClim. However, as mentioned before, model biases are usually hard to explain, so we are unlikely to be able to explain all the results in physical terms, as it is the case of the majority of climate model evaluation papers.

3. This is already considered, and will be highlighted more in the next version of the manuscript. However, note that the high computational cost of very high resolution RCMs is also a severe limitation for many applications.

4. The new experiments proposed by the reviewer would be very interesting. However, we believe that current results are interesting and clear enough for the paper to hold, provided that their explanation and interpretation is improved following reviewer's guidelines. The abstract will be modified too to change the sentence mentioned in the comment. As explained above, WRF is a mesoscale model, and with the domain size used it generates its own daily breeze cycle, as shown in figure 3 of the manuscript, so the comparison with UC-FC is not meaningless, though is very influenced by the small bias of ECMWF forecast model itself.

(there is no point 5)

6. Apart from the spatial coverage, UrbClim allows long multi-decadal adaptation experiments where changes in the city surface parameters can be tested (e.g., the color of the roofs). This would not be possible with a statistical model. The idea is to simulate only the fundamental processes that cause the Urban Heat Island, so the model is lightweight, but still based on physics, and thus allowing sensitivity experiments. Statistical downscaling is also very observation dependent, and has generalization problems when applied to future periods with climatologies far from the calibration period.

Specific comments will also be addressed in the new version of the manuscript, and answered in the final answer that will be submitted once the discussion is closed.

Best regards

The authors