Answer to reviewers

Author comments are in black and reviewer comments in blue.

Anonymous Referee #1

Received and published: 3 March 2016
"general comments"
This paper adresses interesting results on the scope of the presented topic. The methodology is clearly described and the results support the reached conclusions. From my point of view, the manuscript has the quality required to be published after some minor technical corrections. However, I missed some more details on the analysis of results showed in figure 6 (see "specific comments").

Authors would like to thank RC1 for this positive review.

"specific comments"
Line 97-98: I disagree with this argument: "the surrounding topography is flat, which does not favour temperature inversions during night time hours". In fact, flat terrain favours the development of a nocturnal stable layer. However, since inversions are more related on atmospheric stability, I suggest to change this argument. Perhaps you just can comment that there are no significant orographic objects between the two stations.

The sentence has been re-written following the reviewer's suggestion. The new sentence reads as follows: The rural station is located in a delta, and therefore the surrounding topography is flat, with no relevant orographic objects between the two stations.

Line 239-240: Do you think that the LST overestimation of both UrbClim and WRF can be related to the fact that the cloudy nights are not considered after filtering the satellite images as it is described in chapter 2.2? In other words, to compare both models against satellite, the same days have been considered to compute the LST averaged during nightime hours for the warm season (figure 6)?

The days (and the times of the day) considered in the averages are the same in the model data and in the observations. The 8-day averages flagged as containing cloudy days in MODIS were masked before computing the averages in both sides. We devoted a lot of effort to review the code that carries out this calculation, which is not trivial, in order to be sure that the comparison was performed under the same conditions and after checking for errors in the codes. Thus, the cause of the LST overestimation must be a different one. According to some calculations we did, different emissivity used by MODIS and the models can explain up to 50% of the bias. The remaining 50% may be related to soil parameters or small differences in the definition of the surface in the urban environment. As the LST is a very sensitive variable, there are many possible causes that could potentially explain the bias, and therefore we prefer to investigate in depth this interesting issue in another article.

Line 287: I think it is important to emphasize that the first conclusion is obtained after analysing observational data and valid only for the warm season.

This is emphasized in the new version of the manuscript: *"The average UHI in the city of Barcelona during the warm season (May-September) reaches 2.5ºC at night."*

Line 93: style. Substitute "...at the city centre of the city..." for "...at the city center of Barcelona..."

The sentence has been modified as suggested.

Line 146: typing error. Substitute "...important give the..." for "...important given the..."

The typo has been corrected.

Line 154: typing error. Substitute "...nested domains 1b..." for "...nested domains (figure 1b)..."

The typo has been corrected.

Line 242: typing error. Remove "advectionAs"

The word “advection” has been removed, thanks.

Anonymous Referee #2

The authors would like to thank again the referee for the thorough review. The main comments were already addressed in the interactive discussion. Some of these answers are here reproduced for clarity, together with references to the changes implemented in the manuscript to address the reviewer's concerns.

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.

We would like to note that UrbClim is not only an offline urban canopy model (UCM), but it parametrizes the planetary boundary layer (PBL) and performs a simplified simulation of the lower part of the atmosphere (see section 2.3 of the manuscript, and De Ridder et al. 2015). In order to stress these points, the following changes were introduced in the manuscript:

- The title has been changed, and now it reads: *“Advantages of using a fast urban boundary layer model as compared to a full mesoscale model to simulate the urban heat island of Barcelona“*

- In the abstract, UrbClim is more specifically defined as an “urban boundary layer” model, and not just as an urban climate model.

- In the introduction, it is now explained that UrbClim simulates both the PBL and the surface: *“Here we show how, by using a simplified model that only accounts for the Planetary Boundary Layer (PBL) and the surface physics, it is possible to reach resolutions of 250 m with affordable computational resources”.*

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 extremely difficult to afford. Note that the computational cost is in turn the very reason of 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 UrbClim 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. Instead, we show that UrbClim, which is a simplified model that only accounts for the Planetary Boundary Layer (PBL) and the surface physics, simulates the UHI of Barcelona with comparable skill to a full mesoscale model. We think that this is a strong practical result, which justifies the use of UrbClim in for the study of specific aspects of the urban climate.

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). The following paragraph has been added to the manuscript to explain this point (L 204-276):

"From these results, it is clear than the higher resolution of the ECMWF forecast model respect to ERA-Interim (16 km vs 70 km) is greatly improving the performance of UrbClim. It is however unclear how to understand the comparison of WRF with UrbClim. On one hand, UC-ERA and WRF, which are both nested in the same reanalyses, display overall similar scores in temperatures, and WRF does better in the wind speed. However, WRF carries on a full dynamical downscaling up to 1.1 km resolution so, in principle, it should be able to achieve an accuracy similar to UC-FC. But the UrbClim run nested in the ECMWF forecast model does show slightly better scores than WRF. Given the large number of factors involved, it is difficult to find an explanation to this result in physical terms. In general, WRF is more biased than UC-FC (figures 4, 5 and 6) and than the ECMWF forecast itself (not shown). It produces too cold temperatures and slightly too high wind speed during the day. As WRF is very customizable, it may be possible to remove these biases with a more careful configuration. However, WRF biases in the wind speed have been proven difficult to correct, and research is yet ongoing in this line (García-Díez et al., 2015; Lorente-Plazas et al., 2016)."

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 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 having to analyse in depth 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 have rewritten several parts of the manuscript in order to clarify the points commented above. The article does not claim anymore that UrbClim is "better" than WRF in terms of performance, given that it is not the aim of the paper. But at least we think that results here shown prove that the performance of UrbClim has similar skill than that of WRF, under comparable conditions. To prove that, we show that UrbClim and WRF, both driven with Era-Interim, have similar skill in the simulation of the temporal and spatial patterns of the urban climate in Barcelona. As an additional result, we also show that the performance of UrbClim can largely improve when the Integrated Forecast System is used as a driving simulation, although this is not part of the direct comparison with WRF.

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 thank the referee for this suggestion, we however think that additional experiments are not necessary, given that the direct comparison between UrbClim and WFC is shown by means of the use of ERA-Interim as the driving simulation. Instead, we have rewritten parts of the paper and clarified the points explained above. In this way, we do not claim that UrbClim outperforms the RCM, but instead we show that they have comparable skill. As an additional result, we also show that the performance of UrbClim can largely improve when the Integrated Forecast System is used as a driving simulation, although this is not part of the direct comparison with WRF.

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.

The objective of the paper is to show that, despite the simplifications of UrbClim, its simulation is not worse than the RCM, and therefore, given its lower computational requirements, it can be used as an alternative tool to model some urban climate processes, in this case, in the city of Barcelona. Nonetheless, it is not our aim to perform a detailed study of the dynamics of the whole troposphere, given that UrbClim is a simplified model of the lower troposphere.

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 think that the paper is original and relevant due to the geographical location of the study, the complex climate here studied (with a strong sea-breeze system), and the length of the simulation (five months). We think that it is also an important contribution worth publishing because it enables a robust evaluation and comparison of UrbClim (with two driving models) and WRF.

The reviewer mentioned some issues with our interpretation of the results that have been addressed and clarified in the new version of manuscript.

Answer to general comments:

1. The authors mention internal variability in multiple occasions, but it is unclear from their discussion what they mean by internal variability. It is also unclear why internal variability is regarded as intrinsically negative (L161-165). I understand that the authors preference is to avoid model departures from the boundary conditions in terms of the large scale (or mesoscale) conditions and I agree that in that sense, UrbClim does not generate any “internal variability” (L178), but in a broader sense, it is required that models produce some internal variability so they produce results the are different from the boundary conditions (added information). In any case, the authors need to be more explicit about what they mean by “internal variability”. The authors also mention internal variability to justify the resolution jump between the boundary conditions and the UrbClim resolution (L121-123). I don’t quite understand the sentence and why it is acceptable to have such a difference between resolutions. UrbClim is effectively downscaling a single grid point of ERA-Interim and therefore forced everywhere with the same conditions. This is something that compromises the representativeness of the results.

We agree with the reviewer, and therefore we have improved the discussion about internal variability throughout the paper. We believe this clarification is key for the improvement of the interpretation of the results. The modifications are:

Section 2.3, (The UrbClim Model)

“Mesoscale models are tied to their driving models by the boundary conditions. Yet, they develop internal variability (Giorgi and Xunquiang, 2000). In the case of UrbClim, the small size of the domain, and the simplicity of the atmospheric component, greatly reduce the internal variability.

The model can be seen as a "wind tunnel". This can be a limitation for the capability of the model to add value to the coarser resolution boundary data, especially for variables like wind.

On the other hand, the small internal variability has also advantages, as the stability of the simulation is not compromised by the difference in the resolutions of the UrbClim and driving models, as it normally occurs with conventional mesoscale models. Nonetheless, this resolution jump can sometimes affect the quality of the simulation if the driving model does not accurately reproduce the local climate. This balance between the internal variability and the computational power will be key for the interpretation of the results."

2. What does UrbClim do better than WRF-SLUCM? It is necessary that the authors provide a better description of the models and a reasoning of which processes UrbClim might be improving. It is necessary to describe how are the city characteristics seen by the models, what are the differences in the two models (e.g. extension, density, types of building, vegetation cover. . .) Are they different in each of the models? What are the differences how UrbClim deals with urban and rural areas? And WRF-SLUCM?

In the new version of the manuscript we added more details about the configuration of WRF and UrbClim, which help to understand the basic differences between both models. Also note that we included two new figures (figures 2 and 3) with the land cover used by UrbClim and WRF, mapped from CORINE in both cases.

UrbClim and WRF-SLUCM are described in detail in the references provided. Additionally, a new paragraph has been introduced at the end of section 2.3, briefly describing the differences between the urban canopy models of WRF and UrbClim.

Regarding the comparison between UrbClim and WRF-SLUCM, 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. However, in the new manuscript, we do not claim anymore that UrbClim is better than WRF, but instead we say that they show comparable skill. The abstract now says: *"The results show that, generally, the performance of the simple model on reproducing the Urban Heat Island is comparable to the mesoscale model.*

3. The extension of UrbClim domain is roughly 25 km by 25 km. UrbClim is essentially an offline model driven by 1-grid point from ERA-Interim and 4 grid points in ERA-FC. It is obviously much faster than the RCM (considering the number of grid points, including the vertical, in each model provides an idea of the considerably lower computational requirements of UrbClim. This is a very important feature, but it should be considered that the area simulated is much smaller and the applications more limited.

This is already considered, and it is highlighted in the next version of the manuscript. However, we would like to note that the high computational cost of very high resolution RCMs is an important limitation for many applications related to the urban climate. Namely, those needing multi-decadal simulations with large ensembles of models like CMIP5 models in order to account for uncertainty. Note also that, while the applications of UrbClim are certainly more limited than RCMs, a large number of urban climate studies focus precisely on the heat stress and the UHI, so they fall into the scope of UrbClim.

4. The comparison is unfair in many different ways. UC-FC and WRF cannot be compared in any way (They are forced by different mesoscale conditions). In the abstract, the author anticipate a better performance of UC-FC, but in my opinion this result does not prove that UC outperforms the RCM. The authors are evaluating at the same time the performance of ERA-Interim and ERA-FC,

and UrbClim and WRF. This comparison would only be acceptable if WRF is driven by ERA-FC too. I would suggest the authors consider other experiments to adequately assess the performance of the models. In addition to WRF driven by ERA-FC, UrbClim could be driven by WRF outputs and test if it improves WRF estimates.

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, as their explanation and interpretation has been improved following reviewer's guidelines. The abstract has been 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.

Nevertheless, in the new version of the manuscript is is clarified that UC-ERA is the simulation truly comparable with WRF. UC-ERA is now included in the figure showing the minimum temperature climatology, and is the one used in the LST comparison (figures 7 and 8 in the new manuscript), instead of UC-FC. The skill and bias of UC-ERA is clearly comparable to WRF, so the message of the new version of the paper is well supported.

(there is no point 5)

6. The question that arises after reading the manuscript is, why not using statistical downscaling, which is even faster, allows for large ensembles and will reproduce present climate results much better as the are purposely calibrated. Why would an user opt for UrbClim over statistical downscaling? I would guess the answer is the spatial coverage of UrbClim.

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. In the particular case of Barcelona, the scarcity of station data would be a strong limiting factor for statistical downscaling.

Specific comments

L44 Regional climate models have multiple applications, not just downscaling climate projections. Perhaps: “RCMs are limited area models used to downscale climate change projections from coarse resolution Global Circulation Models as well as other applications."

The sentence has been modified as suggested.

L53 Please specify which of the Urban Canopy options if the authors want to keep the following sentence. “This parameterisation. . .”. Otherwise, specify only in the methods. Also, if the authors refer to Single-Layer Urban Canopy Model, it was developed by Kusaka et al. (2001). Chen et al. (2011) describe the implementation in WRF. I also miss some references to work done with WRF/SLUCM for future projections (e.g. Georgescu et al. 2013, Argueso et al. 2015, Kusaka et al. 2012).

A new sentence has been included to clarify that the parameterization used is the Single-Layer

Urban Canopy Model. Kusaka et al. (2001) is now cited, and we also mention three papers about future projections.

L64 Add ", especially" after urban pollution. Urban pollution exacerbates sensitivity to adverse conditions is for all population segments.

The sentence has been modified as suggested.

L70 (FIGURE 1) I would be necessary to have some sort of reference of the urban extension in Figure 1. For those of us not familiar with the region it is difficult to locate the city, and urban areas in general, in the map. It is a major feature of this study and should be shown.

Instead of doing this, two new figures with the land use classes used by UrbClim and WRF where included in the paper (figures 2 and 3). This is related to requests by reviewer#3 and we believe that it helps to understand the interpretation of the results.

L93 How representative is the temperature in the roof of a building of the temperature at levels that matter for population? How comparable is that to the temperatures provided by both models at 2 m .

This is an accurate assumption. See this paragraph of De Ridder et al (2015) and the references:

"This homogeneous mixing assumption in the urban canopy layer is supported by several studies. Nakamura and Oke (1988) measured very slight air temperature gradients only through most of the urban canopy layer, as long as the considered location was not too close to a lateral surface. In their urban canopy model, Erell and Williamson (2006), make this same assumption. Also, measurements acquired in a street canyon in Basel (Switzerland) at different times of the day (Rotach et al., 2005) show little vertical temperature variation, and show that the top-of-canopy temperature is fairly representative of values lower down in the canopy. Finally, in a numerical experiment conducted with a computational fluid dynamics model on an idealized street canyon, Solazzo and Britter (2007) found the canyon air temperature to be spatially nearly uniform, apart from a thin near-wall thermal boundary layer."

Erell, E., Williamson, T., 2006. Simulating air temperature in an urban street canyon in all weather conditions using measured data at a reference meteorological station. Int. J. Climatol. 26, 1671–1694.

Nakamura, Y., Oke, T.R., 1988. Wind, temperature and stability conditions in an E–W oriented urban canyon. Atmos. Environ. 22,2691–2700.

Rotach, M.W., Vogt, R., Bernhofer, C., Batchvarova, E., Christen, A., Clappier, A., Feddersen, B., Gryning, S.-E., Martucci, G., Mayer, H., Mitev, V., Oke, T.R., Parlow, E., Richner, H., Roth, M., Roulet, Y.-A., Ruffieux, D., Salmond, J.A., Schatzmann, M., Voogt, J.A., 2005. BUBBLE – an urban boundary layer meteorology project. Theor. Appl. Climatol. 81, 231–261.

Solazzo, E., Britter, R.E., 2007. Transfer processes in a simulated urban street canyon. Bound.-Layer Meteorol. 124, 43–60.

L97-99 I disagree with this statement. Flat areas are indeed characteristics of temperature inversions, particularly near the coast. Please revise.
The sentence has been re-written following the referee's #1 suggestion. The new sentence reads as

follows: *"The rural station is located in a delta, and therefore the surrounding topography is flat, with no relevant orographic objects between the two stations"*.

L112-113 In the estimate of missing values, did the authors considered all land points within the UrbClim domain? Why 14%?

All land points within the UrbClim domain were considered. The 14% threshold was chosen after inspection of the data, seeking for a compromise between the threshold and the number of days of data finally considered. Lower thresholds lead to the rejection of most of the days.

L116 What to the authors mean by "main features of the urban climate"? Are there hydrological variables (e.g. precipitation, evaporation).

In the model output there is evaporation, together with the latent and sensible heat fluxes, but no precipitation (it is read from the driving model). The sentence has been modified to make it more specific: *"The UrbClim model is designed to simulate the temperature and heat-stress fields at a city scale requiring the minimum amount of computational power, so that it is possible to perform long runs at a resolution of hundreds of metres."*

L132. In which ways? If the extension is not described, this sentence can be removed.

The description has been expanded as requested by referee #3.

L164-165 It is not clearly described how this configuration deals with soil variables. Are they obtained from ERA-Interim at every reinitialisation? If so, NOAH LSM is constantly trying to balance the information from ERA-Interim LSM and they are not necessarily compatible (the even don't share the same layers). A similar question arises for UrbClim in terms of how it sees soil temperature and moisture.

In the case of UrbClim, the simulation is initiated from ERA-Interim and run continuously. In the case of WRF, all the fields, including soil moisture and temperature, are restarted daily, leaving 12h of spin-up. The soil variables are interpolated from the ERA-Interim soil levels to the WRF and UrbClim soil levels by the respective preprocessors.

As the reviewer says, the re-forecast running scheme used to run WRF can in principle introduce a bias if the soil parameterizations are very different. However, the long time it takes to do a full soil spin-up (about one year) means that the soil state related model drift is negligible in short 36 hours simulations, so the physical consistency is granted. In the case of this paper, WRF cold bias during daytime hours seems more related to the overestimation of the wind speed and thus the sea breeze, than to a soil-moisture and evaporation overestimation. Also, note that some studies have found that the soil spin-up can sometimes reinforce the bias rather than removing it (see the comparison between REFOR and MPE-G in García-Díez et al., (2015)). Yet, we agree with the reviewer in that this topic is something worth to be studied with more detail in future experiments.

García-Díez, Markel, Jesús Fernández, and Robert Vautard. 2015. "An RCM Multi-Physics Ensemble over Europe: Multi-Variable Evaluation to Avoid Error Compensation." *Climate Dynamics*, February, 1–16. doi:10.1007/s00382-015-2529-x.

Table 1 How is the variance ratio calculated? What does it represent?

Variance ratio is the quotient between the variances of the model and the observation, i.e var(model)/var(observation). Values above 1 indicate that the model produces too much variability, and values below 1 the opposite.

L170-174 Do these values in UC-ERA ultimately depend on information provided by a single grid point from ERA-Interim? The resolution of ERA-Interim is equivalent to ∼70 km, while the extension of the UrbClim domain is ∼25 km by 25 km.

Essentially yes, that is why we use the analogy of UrbClim being a sort of "wind tunnel", as the boundary conditions are the same in both sides of the domain, but it does use also a pressure gradient term, computed with the surrounding gridpoints.

L179-180. It is not correct to say that the "results can be interpreted as a comparison between Urban Canopy+PBL models driven by" different boundary conditions. In WRF there is a two-way interaction, the results from Urban Canopy+PBL are fed back into the dynamical core and influences local circulation (and potentially larger scale circulation).

Please see below.

L181-182 is especially concerning since the authors disregard that this comparison is mixing multiple things, the mesoscale information to begin with. I agree that at these scales (15km to 1km) and for variables such as temperature, there is little benefit in increasing resolution, but this cannot be inferred from the authors results.

This paragraph (L179-182) has been removed in the new version of the manuscript. It has been replaced by a new paragraph at the end of section 3.1. In this new paragraph, we discuss the difficulty of comparing WRF with UC-ERA and UC-FC, given the scope and resolution differences, in the line of the comments of the referee #2 and the answers given in the present document.

L197-198 Is this a merit of FC over ERA-I rather than UrbClim?

The improvement between UC-ERA and UC-FC is obviously a merit of FC, but the small bias of UC-FC is also a merit of UrbClim. Note that FC still does not resolve the city, and is more biased than UC-FC when comparing it to the station data (it does not see the UHI). See the figure below. Thus, we see that provided with the accurate wind of FC, UrbClim is doing its part of the job very well.

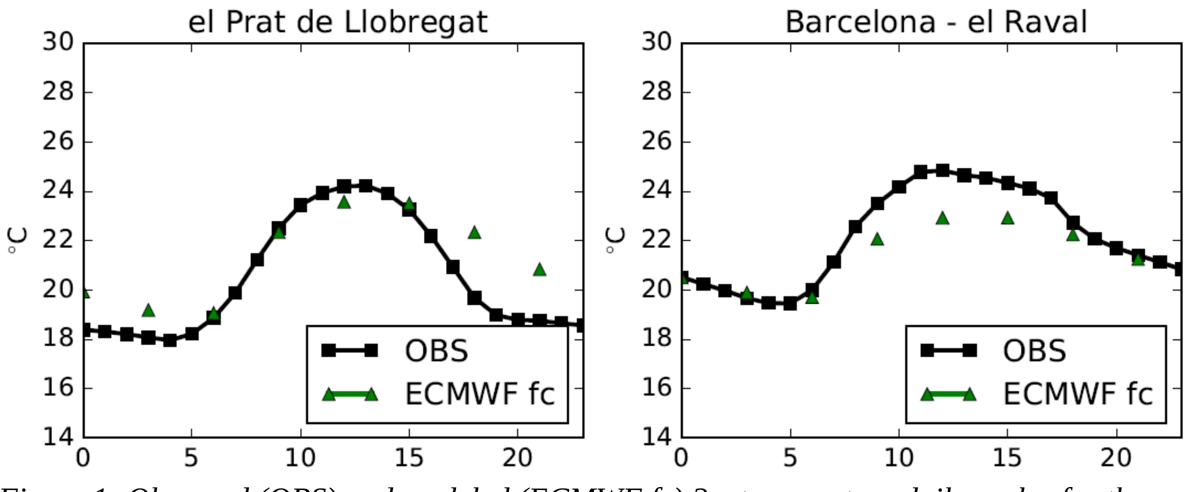


Figure 1: Observed (OBS) and modeled (ECMWF fc) 2m temperature daily cycles for the rural (left) and the urban (right) stations. The model is the ECMWF forecast model with aprox. 15 km resolution.

L200-205 It is obvious that if the model was assigning urban land use to the rural location, then for the purpose of this study they are not comparable by any means. I do not think this needs to be mention in the text. Instead, I would say that the closest grid point with a particular land cover (and specify the land cover) was selected to compare with the rural station, and perhaps say the distance between the grid point and the station.

The paragraph has been re-written, but yet the process is explained with some detail, to take into account also referee #3 comments. Also, a figure with the location of the station, WRF land use and the chosen gridpoints has been added as a supplementary figure, as requested by the reviewer #3.

L 206-207 I would say that the bias occurs throughout the day except in the evening (16H-00H)

The sentence has been changed as suggested.

L 210 This is not surprising. As the authors suggest, at 70 km the intensity of the sea breeze is often underestimated. But also, only information from 1 ERA-Interim grid point is provided to UrbClim.

We agree with the referee, although as mentioned above, also the pressure gradient is used, involving information from the surrounding gridpoints. However, the pressure gradient related to the sea-breeze in ERA-Interim is too small.

Figure 3. A physical explanation of why sea breeze is weaker in urban points would be desirable. It cannot simply be increased drag because the urbanised areas along the coast act as a barrier for the rural station too. Furthermore, the urban station is located in a roof, so depending on the surrounding buildings and the height of the station, the drag could be negligible for that station. In this case, the comparison tell us about the boundary conditions, not necessarily the UrbClim. It could be well the case that UrbClim is doing a fantastic job in both cases, but ERA-FC information is more accurate (or more comparable to those scales).

The shadow casted by the urbanized areas in the wind speed field is very short-lived, as the vertical mixing transports linear momentum from upper levels quickly after the flow leaves the urbanized area. Both WRF and UC-FC show weaker winds in the urban location consistently with the observations. Also, the Urban-Rural wind speed difference is smaller in UC-ERA because the wind is way weaker in general. Regarding the location of the urban station in a roof, this could be an issue if the building was higher than the surroundings, but it is not, and in thus representative of wind speed in the urban canopy.

L221-228 It is unclear what the authors want to illustrate with figure 4. What is the contribution to the paper? Isn’t this day-to-day variability highly influence by the boundary conditions?

It was somewhat surprising to find that the UC-ERA run was getting the average UHI daily cycle right despite miss-representing the sea-breeze cycle. In the figure we show that, despite getting the average values right, this simulation is suffering from large errors in some days. This is fixed both in UC-FC by using higher resolution forcing, and in WRF by doing an actually complete dynamical downscaling. This figure also shows that the day-to-day variability of the UHI of Barcelona is significant, being able to reproduce it in the simulations is a way to improve heat stress forecasts in the city.

L 230-238 (Description of figure 5) This is basically qualitative. It does offer finer detail because it has finer resolution, whether this is correct remains unclear, even after comparison with relatively low resolution MODIS (Figure 6), where a more quantitative measure is provided.

The goal here is to show that both UrbClim and WRF produce consistent spatial patterns despite being very different approaches, and also to show the extra detail that the 250 m resolution can offer. It is true that this detail cannot be properly evaluated because of the lack of observations, but it is plausible and based on physics.

L 239-240 (and onwards) Among the multiple descriptions of UHI, two are widely used. The skin temperature (or surface temperature) UHI and the screen level temperature UHI. Although linked, they involved completely different processes. Indeed, skin temperature UHI is generally positive at day and night, but the screen level temperature UHI is often negative during the daytime. The authors should capture this in their discussion.

A new sentence has been introduced at the end of the section to mention this: *"Finally, as mentioned in the introduction, LST and Surface Air Temperature (SAT) Urban Heat Islands are not equivalent, and are driven by different phenomena. Thus, it is also possible that models that reproduce the SAT UHI correctly produce a biased LST UHI."*

L 241 I understand that results from Zhou et al. 2015 contrast with the authors finding rather than agreeing.

Yes, this is correct, we admit that the "also" used in the manuscript was confusing. In the new version it has been removed, and the sentence now reads: "*Other studies (Zhou et al., 2015) found small errors when comparing MODIS and UrbClim LST...*". Thanks.

L 249-250 If the authors provide these confidence bounds, both values are exactly the same (not just statistically insignificant). Not only the confidence bounds overlap but the both include the estimate from the other model.

This is mentioned in the text: "Thus, UrbClim correlation is higher, but the difference is not statistically significant, as the confidence bounds overlap."

L 300 But UrbClim does not provide rainfall at all, or does it?

That is correct. That is why WRF provides more detail. In the new version of the manuscript a more detailed description of UrbClim shortcomings is included.

Technical corrections

L34-36 Please revise sentence, how it links to the previous one. Please revise use of commas.

We followed the suggestion, removing a comma, a linking word and re-arranging the sentences:

"*They found that the relative contribution of these factors depends on the local background climate of the city and on the time of the day. In general, during daytime, convection efficiency and evapotranspiration are the main drivers of the UHI, while heat storage is the most relevant during* *the night. Zhao et al. (2014) used satellite retrieved land surface temperatures, but these can defer from screen level temperatures. Other authors (Arnfield, 2003), highlight the complexity of the problem of measuring the UHI, because of the difficulty of getting observations of the urban climate with enough detail and reliability.*"

L37 Replace "defer" with "differ"

This mistake has been corrected as suggested.

L38-39 Please revise use of commas.

A comma has been removed: *"Furthermore, the complexity of the urban surface, featuring anisotropy and vertical surfaces, makes it complicated to sample by satellites (Voogt and Oke, 1998)."*

L41 The end of this sentence is unclear. Please rewrite. (The last half of this paragraph needs to be clarified)

It has been divided into two sentences to make it clearer: "*These difficulties with the observations increase the value of numerical simulations, that can produce detailed fields which are not observable. At the same time, the lack of observations hampers the evaluation of these simulations."*

L59 Remove "now" (?)

The sentence now reads: *"Taking into account that the Mediterranean countries are currently more vulnerable to environmental summer conditions than other European societies, the larger magnitude of the projected temperature increase is expected to become a major challenge for public health (Ostro et al. 2012)"*.

Ostro B, Barrera-Gómez J, Ballester J, Basagaña X, Sunyer J. The impact of future summer temperature on public health in Barcelona and Catalonia, Spain. International Journal of Biometeorology 56, 1135-1144 (2012).

L93 Replace "Km" with "km"

This mistake has been corrected as suggested.

L152 Remove comma after sub-modules. Missing verb before available?

One particularity of this model is that i has a large amount of parameterization schemes, dynamical options and sub-modules, available to the user to choose among them.

L242 Please remove "advection"

This mistake has been corrected as suggested.

Anonymous Referee #3

The authors would like to thank referee #3 for the useful review.

The authors compare three different types of model runs for Barcelona. The details of the models are not provided. A summary Table which compares the key features (model characteristics, run resolutions, etc) and could include the computational resources difference, and key performance differences would be a useful addition. This could be cited throughout the paper (methods, results) to allow the reader to be clear how the benefits/costs are arrived at.

As requested by the reviewer, a table has been added to the paper (table 2) with relevant information about the model configuration (domain size, time step) and the results of the benchmarking experiment, and it is cited in the text. Model performance against observations has not been included in the table, because of the difficulty of finding single numbers representative of the model skill. This could lead to misunderstandings because, as referee #2 pointed out, the interpretation of the results in terms of “which model is the best” is not straightforward. Please note also that the models are described with great detail in the references.

More details are needed on the measurements and processing of the evaluation data; the implications of the study period selected (clear). The comment (L200) concerning the gridpoints and the land use for the evaluation data needs to be made clearer or justified. It appears a better result is being selected – rather than understanding if there is a larger issue.

In the new version of the manuscript, more detail has been introduced in the description of the evaluation data in section 2.1. This includes new supplementary figures that show clearly how the choice of the representative gridpoints in the WRF grid is justified.

All figure captions should be standalone. Add additional material/text to these.

The captions have been extended, following the reviewer's request.

1. L5 use the term evaluated not validated (and equivalent throughout)

All the forms of the verb “validate” have been replaced by “evaluate” throughout the paper.

2. L8 including not using

The sentence was left unchanged, as in WRF using an UCM or not is left at choice of the users.

3. L18 use the ‘most well-known’ rather than ‘main’

The sentence has been modified following the correction.

4. L36 – reword

Referee #2 also commented on this part. The revised sentences are now: “*They found that the relative contribution of these factors depends on the local background climate of the city and on the time of the day. In general, during daytime, convection efficiency and evapotranspiration are the main drivers of the UHI, while heat storage is the most relevant during the night. Zhao et al. (2014) used satellite retrieved land surface temperatures, but these can defer from screen level temperatures. Other authors (Arnfield, 2003), highlight the complexity of the problem of measuring the UHI, because of the difficulty of getting observations of the urban climate with enough detail and reliability.*”

5. L41 – see point (1) (repeated through text)

As mentioned, the correction of pint (1) has been applied in all the paper.

6. L47 250 m, not 450m (change throughout)

This correction has been throughout the paper too.

7. L63 use 13.7 not 13,7 notation (correct throughout)

The mistake has been corrected throughout the sentence.

8. L65 reword

The sentence now reads: This larger sensitivity to environmental conditions is exacerbated by urban pollution especially in old people living in cities with pre-existing or chronic cardiovascular and respiratory diseases (McMichel et al., 2006).

9. L70 – Figure not figure

This mistake has been corrected in new version of the manuscript.

10. L85,86 4, 7 – numbers less than 10 write in full

These numbers have been written in full as requested.

11. L88/9 – what height and exposure? How high is the sensor? Be clear about samples and averages.

The stations follow the WMO standards. Apart from that, we lack more precise metadata about sensor height.

12. L92 – Cereal fields, so changing height through the course of the year

The cereal fields surround the station, but the station location itself is not cereal, but grass and is regularly maintained, which includes cutting the grass.

13. L93 – how high above the roof? What is the height of the building?

See the new information added in the manuscript.

14. L94 Km should be km

This mistake has been corrected in new version of the manuscript.

15. All maps need scales.

The software used for producing the maps does not allow to plot scales when using the regular lon-lat (Plate Carree) projection. Instead, the axes have been labeled with latitudes and longitudes, which add information both about the scale and location of the maps.

16. L105 on – what correction used for emissivity? Between areas/urban etc

We have used the Land Surface Temperature data as provided by the MODIS team. A detailed version of the algorithm is available in Wan (2008).

Wan, Z. 2008. "New Refinements and Validation of the MODIS Land-Surface Temperature/Emissivity Products." *Remote Sensing of Environment* 112 (1): 59–74. doi:10.1016/j.rse.2006.06.026.

17. L110 be clear that selection of no cloudy days introduces a bias to certain meteorological conditions

A new sentence has been introduced to be clear about this: *"This introduces a bias in to certain meteorological conditions (clear-sky days), but it is unavoidable."*

18. L160 cite chapter authors, not the book

The reference has been modified following the suggestion.

19. Table 1 link to Figure 1 (stations); Define Variance ratio or cite reference

A link to figure 1 has been introduced in the caption of table 1, as suggested. Also, a new sentence has been added to the caption, to define the scores, including the variance ratio: "*The scores are: Mean bias (model - observed), Root Mean Squared Error (RMSE), and variance ratio (variance of the model divided by variance of the observation).*"

20. L118 meters -> UK or USA English?

It has been modified to "metres" to be consistent with UK English.

21. L170 standard scores or metrics – reference

The word "standard" has been removed, as there are not"standard" scores in the literature. While the scores we used are very typical, we think it is not fully accurate to describe them as "standard".

22. L170 2 m

The sentence has been modified as suggested.

23. Figure 2 – indicate in the caption where codes for key are explained. Captions should be standalone

A new sentence has been introduced in the caption following the advice of the referee: *"The UC-ERA, UC-FC and WRF legend codes are defined in section 2, while OBS is the observation."*

24. L185 Check in all places oC; or express in terms of K and remove o. In some places reversed ◦ C (e.g. Table 1)

The temperature units have been checked and corrected throughout all the manuscript.

25. L189 – Garcia

The reference has been corrected as suggested.

26. L193 - Note importance of land cover. What do they represent in terms of Local Climate Zones?

There are no significant differences in the local climate of the stations, apart from the effect of urbanization. Furthermore, the new figures give much more information about the land uses.

27. L193 cite these previous studies

This sentences refers to the (Moreno-Garcia, 1994) paper cited above.

28. L200 on – But don't you need to check all grids now? Land use? Advection? Etc. what

Advection of the UHI over the rural location is a plausible hypothesis under certain conditions given that it is close to a urbanized area. This is especially true in WRF, in which, due to its lower resolution, the rural gridpoint is close to a urban gridpoint. However, the advection of the UHI seems to be very small in Barcelona, according to the models, and the sea breeze does not blow from the city to the rural station. But, bearing all these considerations in mind, we must note that the final goal of the study is not a perfect measurement of the UHI in Barcelona, which would require a measurement campaign, but the evaluation of the model. If there is an influence of the UHI in the rural location through advection, this should be simulated by the models.

29. Figure 3 – be explicit about UHI – temperature difference

Both in former figures 2 and 3 (which are figures 4 and 5 in the new manuscript), the title of the right panel has been changed from "UHI" to Urban – Rural to be more explicit.

30. L 222 (e.g. Figure 4) rather than which are here depicted in . . ..

The sentence has been modified as suggested and it now reads as follows: "*It is interesting to highlight the day-to-day variability of the observed and simulated times series for the month of May (figure 5)* ."

31. L231 Figure 5 (introduce space)

This mistake has been corrected in new version of the manuscript.

32. Figure 4 – Significant figures! Relabel X-axis no need for May 2011 on all as in caption

The figure has been edited as suggested.

33. L234 as above space between number and units

This mistake has been corrected, thanks.

34. L240 as above evaluation rather than validation

This has been corrected through all the manuscript.

35. L 242 – typos near delete .advection

The typo has been corrected.

36. L256 – be more explicit about long spin-up. Some suggest for certain models 10-20 years are needed to get soil moisture characteristics correct.

They suggest this, but do not show the proof.

37. Line 319 data were not was

The sentences has been modified as suggested by the reviewer.

M. Mohan comment

Authors want to thank M. Mohan for this comprehensive review.

General Comments:
The paper addresses advantages of using a fast Urban Climate Model (UrbClim) over a full scale mesoscale model (WRF) by using a case study depicting UHI simulation for the city of Barcelona. Urban Climate Model is relatively of high resolution that is driven by ERA interim reanalysis data or GFM of ECMWF. The model has earlier been validated for other European cities and now being implemented for Barcelona and utilized for climate projections. The model is demonstrated to be computationally efficient with higher resolution than a mesoscale model. However, as mentioned in the title, the model (UrbClim) is not a urban canopy model as canopy features are not included in the true sense and the mesoscale model WRF does not include WRF- UCM. Hence the title needs to be appropriately modified with direct usage of UrbClim Though UHI is derived from the temperature differences, the paper presntlly lacks the robustness of estimating UHI. Further, model validation shall include meteorological parameters such as Wind speed and PBL etc. to demonstrate the efficacy of the two models. It would be interesting to see the performance of all the key meteorological parameters from the two models to examine the comparable performance and infer of the desired efficiency of UrbClim and this model comparison alone is adequate to make this point.

The comments are elaborated below:

1. Line 90 onwards: Station number 6 is located on a rooftop at 33m while Station number 5 which is a rural station in located in a delta. The thermal properties of land surface and a land use predominently reflecting a water body is drastically different and the fact that no inversion is observed over rural station (considered here as reference station for UHI) would lead to erroneous representation of estimated UHI of the city.

The objection here is unclear. Note that the river forming the delta is quite small, and the water bodies do not prevail over the location of the rural station at all. The land use is dry land, cultivated mainly with cereal but also other species, with artificial irrigation (but not flooded). Also, the sentence mentioning inversion in the paper was confusing, and it has been replaced by: *“The rural station is located in a delta, and therefore the surrounding topography is flat, with no relevant* *orographic objects between the two stations"*. The terrain is also very similar in both rural and urban stations: coastal plain with small or no slope.

2. Figure 1 a and 1 b: Figure 1 b should also depict the rural and urban stations. Both figures should depict coordinates (lat/long) using ArcGIS or another appropriate software. It appears that Urban and rural stations are not within the same nest and with similar resolution. This might affect the results somewhat. This needs to be clarified.

All the maps of the paper include now coordinates in their axis. Also note that the urban and rural stations are in the same nest in the WRF domain. In figure 1b, the black contour represents the UrbClim domain. Thus, it is not possible to plot the stations in figure 1b, because the region covered by the map is too large.

3. Brief description of incorporation of urban canopy model (physics/key eq.) in UrbClim model shall be included to explain the science and for comparison with mesoscale model to understand level of simplification.

The details are provided in the references. As the scope the paper is very applied, we think that discussing details about the physics and the parameterizations is out of place.

What are the time step for running these two models? A climate model run needs larger time step and a weather model like WRF requires lower time step to run.

Information about the time steps is now included in the paper. For UrbClim is 20s for the land surface module, but the atmospheric scheme is solved with a variable time step, using the Courant-Friedrichs-Lévy stability criterion to determine the longest possible stable time step.

If this difference is large between the two models, UrbClim will obviously be computationally much more efficient. It would be more appropriate to compare UrbClim with another Climate model with similar or comparable time step resolution for depicting it's efficiency for predicting average temperatures from May to Sept. 2011. Essentially, this efficiency is one of the major claim by the authors in this paper and that is not shown by considering two dissimilar type of models in terms of phenomenal applicability of UHI. Further by including other meteorological parameters such as wind speed, fluxes and PBL and demonstrating the comparative model performance would prove the point more effectively.

The new version of the manuscript is clear about the different scope and applicability of WRF and UrbClim, and the goal is not to replace WRF with UrbClim is all applications. UrbClim is focused on reproducing the UHI and the heat stress, so this defines the variables used in the study. WRF is being used in many studies focused on the UHI that could use a faster and simpler model as UrbClim, and our goal is to show that.

Also, note that the main reason of the performance difference is not the time step or the number of gridpoints, but the simplicity of the dynamical core, and that UrbClim does run less parameterizations (there are no cumulus, radiation and microphysics parameterizations). Thus, the paper does not try to compare the computational efficiency of two similar models, but of two very different approaches to simulate the same thing (the UHI of Barcelona).

4. UHI requires detailed description of measurement sites including station pictures with surroundings.

Pictures of the measurement sites are available in the web page of the Catalan meteorological

Service:
- urban: http://meteo.cat/observacions/xema/dades?codi=X4&dia=2016-05-26T13:30Z
- rural: http://meteo.cat/observacions/xema/dades?codi=XL&dia=2016-05-26T13:30Z

2 station data will not be sufficient for robustness. Thus, it is suggested that model implementation and efficacy claim is limited to temperature predictions and other met. parameters shall also be included for this purpose.

We disagree with this statement. Two well placed and maintained stations can provide a good representation of the UHI. Furthermore, the study is focused in the UHI, and therefore the evaluation of other meteorological parameters is out of the scope.

Further, authors claim that there are 11 met. stations in the domain; however data for only 2 stations is being used to depict UHI. Were the chosen stations showed the maximum UHI? How about UHI for other stations ? Could they reflect justifiable trends vis-a- vis their LULC? Most of the study would use 25 or more stations for UHI(Mohan et al., 2013:, Assessment of urban heat island effect for different land use–land cover from micrometeorological measurements and remote sensing data for megacity Delhi, Theoretical and Applied Climatology, 112, 647-658. DOI 10.1007/s00704-012-0758-z).

The other possible station pairs to be used to measure the UHI have different problems: Are too far away, located at very different height above sea level, separated by topographical barriers, too close to the sea, or to the airport. The station pair used was chosen after a careful analysis of all the data.

On line 190, authors mention that " The measurement of the UHI with only two points has some limitations, as it may be sensitive to very local features such as the land use in the vicinity of the stations. However, the representativeness of these points has been carefully checked with high resolution satellite images". No details are provided as to how it has been checked.

They have been checked by means of high resolution google earth images, together with the pictures provided in the web page of the meteorological service (find more details in previous comments).

In addition, the agreement with previous studies increases our confidence in the results here presented. No details of any previous studies provided here.

A previous study is referenced (Moreno-García et al. 1994).

Satellite data would represent LST and not air temperature while UHI estimated and shown are based on air temperature.

Yes, this was already explained in the original version of the manuscript. LST is the only possible way to evaluate the spatial pattern of the simulations.

5. Does the same trend in UHI was seen based on LST also? It is not clear. UHI phenomena occurs on a diurnal scale for which WRF-UCM is applied (Bhati and Mohan, 2015; TAC; doi:10.1007/s00704-015-1589-5).

RMSE of temperature of about 2 deg. is acceptable WMOas per guidelines on shorter time scales (of an hour or day as demonstrated in this study. Therefore results can be sufficiently robust for temperatures averaged over 5 months but the differences of 2.5 deg. as in UHI may not be ; hence model application seems appropriate for temperatures and not for UHI.

Unfortunately, we are not sure we completely understand what the comment is trying to refer to.

Morerover, different heights of urban and rural stations will further add to this uncertainty.

As mentioned in the original version of the manuscript, the height difference can account for a difference of 0.15-0.25 degrees depending on the gradient considered. Also see the answer to referee #2 regarding the representativeness of temperatures measured in roofs (page 6 of this document).

6. Simulations are carried out from May to Sept, 2011. It is not clear for statistical paired analysis what temporal and spatial resolution and how many data points are used?

This information is now included in the text and in table 2.

7. It will be good to include the monthly variations of UHI considering seasonality in May to Sept. data and examine the trends ? Similarly for the air temperatures and LST as well the monthly variations could be included.

This interesting suggestion from the reviewer will be addressed in a future article, in which much longer simulations, of the order of 30 years, will be performed in order to analyze the climatology of the city. In this article, however, we simply wanted to perform an initial evaluation of the model, for the particular case of the city of Barcelona.

8. Based on model simulations the spatial variation of UHI needs to be studied. It shall be shown whether spatial variation shows maximum difference at the two selected points and other places in the domain are depicting lower UHI so that urban -rural contrast can be deciphered.

Unfortunately, there are no observations of the spatial pattern of the 2m temperature. Figure 7 depicts the spatial pattern found in the simulations, which shows that the two selected stations are representative of the maximum difference between the city centre and the surrounding areas.

9. The title mentions fast Urban Canopy Model while the abstract and text categorises this as Urban Climate Model and the mesoscale model used is WRF and not WRF-UCM. As per the text, WRF includes USGS LULC and no mention of WRF UCM is made. WRF has tremendous scope of improvement by using recent LULC other than USGS and including Urban Canopy as demonstrated by Bhati and Mohan (2015). Thus urban canopy model in the title may be replaced with urban climate model.

The title has been modified in the new version of the manuscript to account for this, and now it reads: *"Advantages of using a fast urban boundary layer model as compared to a full mesoscale model to simulate the urban heat island of Barcelona."* which we believe is more accurate. As mentioned in the original version of the manuscript, the LULC used in all the simulations in the paper is the European CORINE dataset, which is considerably more accurate than the USGS dataset included in WRF.

Manuscript prepared for Geosci. Model Dev.
with version 2015/04/24 7.83 Copernicus papers of the LaTeX class copernicus.cls.
Date: 10 June 2016

Advantages of using a fast urban ~~canopy~~ boundary layer model as compared to a full mesoscale model to simulate the urban heat island of Barcelona

M. García-Díez *[1], D. Lauwaet[2], H. Hooyberghs[2], J. Ballester[1], K. De Ridder[2], and X. Rodó[1, 3]

[1]Institut Català de Ciències del Clima, Barcelona, Catalonia, Spain
[2]VITO, Antwerp, Belgium
[3]Institució Catalana de Recerca i estudis Avançats (ICREA), Barcelona

Correspondence to: Markel García-Díez (markel.garcia@ic3.cat)

Abstract. As most of the population lives in urban environments, the simulation of the urban climate has become a key problem in the framework of the climate change impact assessment. However, the high computational power required by these simulations is a severe limitation. Here we present a study on the performance of ~~a Urban Climate Model (UrbClim),~~ UrbClim, a urban boundary layer model designed to be several orders of magnitude faster than a ~~full-fledge~~ full-fledged mesoscale model. The simulations are ~~validated~~ evaluated with station data and ~~with~~ land surface temperature observations ~~retrieved by satellites.~~ from satellites, focusing on the Urban Heat Island.

To explore the advantages of using a simple model like UrbClim, the results are compared with a simulation carried out with a state-of-the-art mesoscale model, the Weather Research and Forecasting model, ~~using an Urban Canopy model. The effect of using different driving data is explored too, by using both~~ which includes an urban canopy model.

This comparison is performed with driving data from relatively low resolution reanalysis data (70 km)~~and~~. In addition, the effect of using different driving data is also explored in the case of UrbClim, with data from a higher resolution forecast model (15 km).

The results show that ~~, generally,~~ the performance of the Urban Heat Island in the simple model is ~~comparable to or better than the mesoscale model. The exception are~~ generally comparable to the one in the mesoscale model when driven with reanalysis data. The only exception is found in the winds and the day-to-day correlation~~in the reanalysis driven run~~, but these problems ~~disappear when taking~~ in UrbClim disappear when we consider the boundary conditions from the higher resolution forecast model.

*markel.garcia@ic3.cat

1 Introduction

According to the United Nations, more than 50% of the world population lives in cities, and this percentage is expected to increase in the coming decades. The urban environment is known to modify the local climate in several different ways. The ~~main one~~ most well-known is the so-called Urban Heat Island (UHI) effect, that consists on temperatures being several degrees higher over the urban area with respect to its rural surroundings. Due to anthropogenic climate change, the frequency of heat waves is expected to undergo a widespread increase (Meehl and Tebaldi, 2004) in the following decades. This raises concerns about the vulnerability of people living in urban areas.

Although the magnitude of the UHI effect does not necessarily increase due to global warming (Lauwaet et al., 2015), it has been shown to be large enough to pose significant impacts. The most important ones are human health, through heat stress (Gabriel and Endlicher, 2011; Dousset et al., 2011) and energy consumption (Kikegawa et al., 2006; Kolokotroni et al., 2012).

The physical causes of the UHI effect were enumerated by Oke (1982), but the relative contribution of each one is still discussed. Zhao et al. (2014) used satellite observations and a model simulation to calculate the relative contribution of the different causes of the Urban Heat Island in 65 cities of North America. They considered contributions from modifications in the radiative balance, evaporation, convection efficiency, heat storage and anthropogenic heat sources. They found that the relative contribution of these factors depends on the local background climate of the city ~~,~~ and on the time of the day. In general, during daytime, convection efficiency and evapotranspiration are the main drivers of the UHI, while heat storage is the most relevant during the night. ~~However, other authors (Arnfield, 2003) , highlight~~ Zhao et al. (2014) used satellite retrieved land surface temperatures, but these can differ from screen level temperatures. Other authors (Arnfield, 2003) highlighted the complexity of the problem ~~, that is very related to the difficulty of getting observations of the~~ of measuring the UHI, given that it is difficult to monitor the urban climate with enough detail and reliability. ~~Zhao et al. (2014) used satellite retrieved land surface temperatures, but these can defer from screen level temperatures.~~ Furthermore, the complexity of the urban surface, featuring anisotropy ~~,~~ and vertical surfaces, makes it complicated to sample by satellites (Voogt and Oke, 1998). These difficulties with the observations increase the value of numerical simulations, that can produce detailed fields which are not observable~~but, obviously, hamper the validation of~~. At the same time, the ~~models~~lack of observations hampers the evaluation of these simulations.

Recently, ~~Urban Canopy~~ urban canopy models or parameterizations have been included in many Regional Climate Models (RCMs) (see. for example Huszar et al. (2014) and Ching (2013)). ~~The~~ RCMs are limited area models ~~which are usedto downscale the~~ used, among many other application, to dynamically downscale climate change projections from ~~the~~ coarse resolution Global Circulation Models. Nevertheless, computational power limitations do not allow RCMs to reach the level of resolution that is required to resolve most of the cities. Here we show how, by using a simplified model that only accounts for the Planetary Boundary Layer (PBL) and the surface physics, it is possible to reach resolutions of ~~250m~~ 250 m with affordable computational resources. This model, called UrbClim, has been developed by De Ridder et al. (2015), and is described in section 2.3. UrbClim has been already ~~validated~~ evaluated in a few European cities (De Ridder et al., 2015; Zhou et al., 2015; Lauwaet et al., 2016) and used to generate ~~Climate Change~~ climate change projections (Lauwaet et al., 2015). Note that UrbClim has a different, more specific scope than the RCMs, being focused on the fast and computationally light simulation of the UHI and the heat stress in the urban environment, so that it can be easily transferred between cities.

In the present paper, we ~~validate~~ evaluate UrbClim over the city of Barcelona, and compare it with a standard mesoscale model, the Weather Research and Forecast model (WRF), using a Urban Canopy parameterization. ~~This parameterization was developed by (Chen et al., 2011) and~~ Namely, the Single Layear Urban Canopy Model (SLUCM) was used, which was developed by Kusaka et al. (2001) and coupled to WRF by (Chen et al., 2011). It has been verified in several studies (Lee et al., 2011), and used for future climate change projections (Argüeso et al., 2015; Georgescu et al., 2014; Kusaka et al., 201

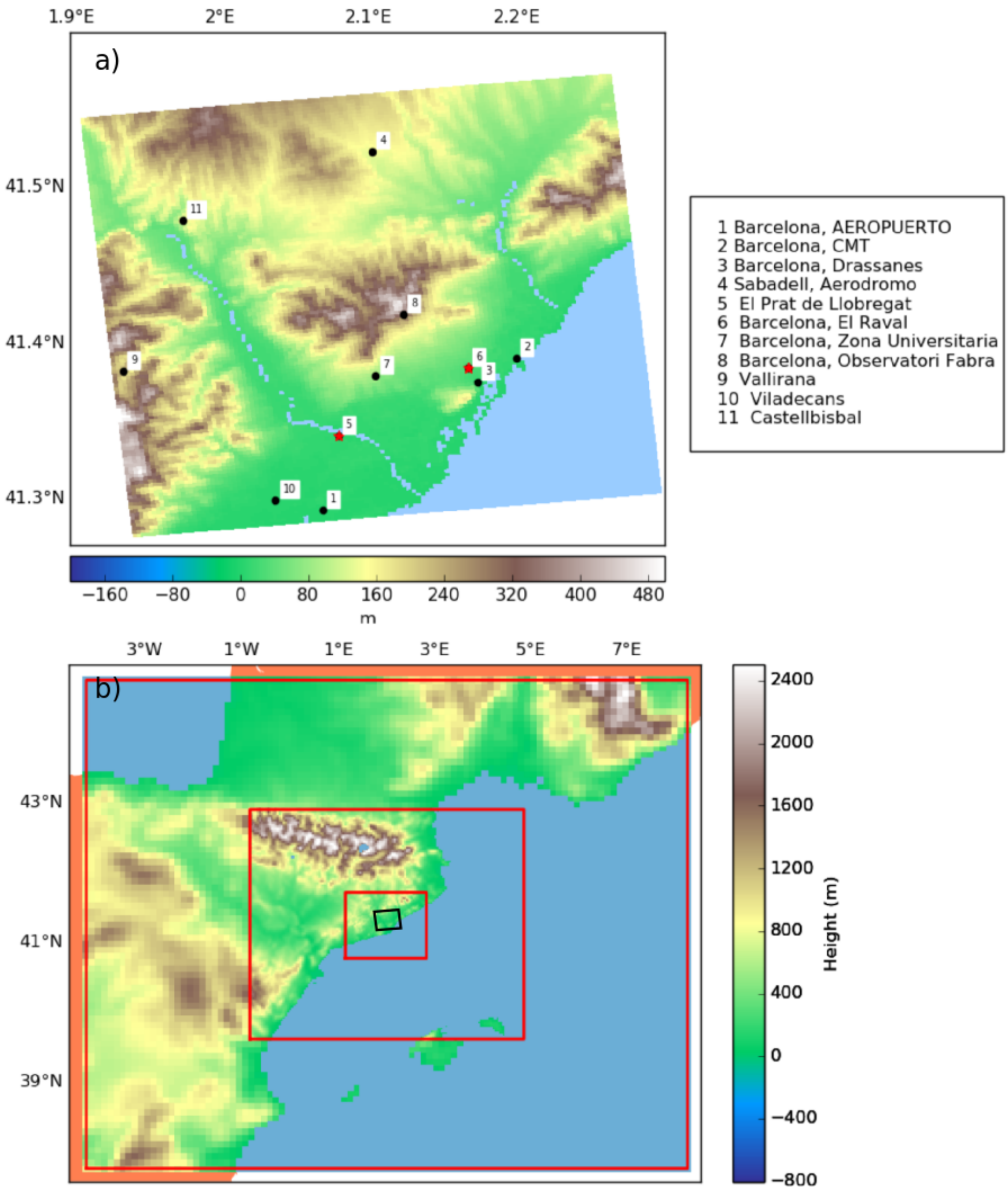
Barcelona is located in the Euro-Mediterranean ~~area. This area~~ region, which has been defined as a primary climate change hot spot (Giorgi, 2006), as it emerges as an especially responsive area to climate change, with more frequent, longer and harsher summer heat waves ~~(Meehl and Tebaldi, 2004; Ballester et al., 2009; ?, 2010b).~~ ~~In addition to~~ (Meehl and Tebaldi, 2004; Ballester et al., 2009, 2010a, b).

Taking into account that the Mediterranean countries are currently more vulnerable to environmental summer conditions than other European societies, the larger magnitude of the projected temperature increase ~~, the Mediterranean countries are already now more vulnerable to environmental summer conditions~~is expected to become a major challenge for public health in summer (**?**). For example, the negative effects of the record-breaking 2003 heat wave in central and southern Europe were particularly damaging in the Euro-Mediterranean arch (Robine et al., 2008). The seasonal mortality excesses were indeed similar in Spain (~~13,7~~13.7%), France (~~11,8~~11.8%) and Italy (~~11,6~~11.6%), although temperature anomalies were ~~at least twice as large~~ twice larger in France than in the southern countries (Ballester et al., 2011).

This larger sensitivity to environmental conditions is exacerbated by urban pollution especially in old people living in cities with pre-existing or chronic cardiovascular and respiratory diseases (McMichael et al., 2006).

Taking into account all these considerations, the city of Barcelona emerges as a particularly vulnerable area within the continent. Barcelona is located in northeastern Spain, surrounded by the Mediterranean Sea in the south and east, a small ~~500m~~ 500 m mountain range in the northwest, and two rivers in the southwest and northeast (figure 1, and see also figure 2 for the location of the urbanized area). Its Mediterranean climate (Csa in the Köppen classification) is shaped in summer by the local wind breeze regime, whose diurnal evolution exhibits a clockwise rotation from southerlies in the morning to winds blowing roughly parallel to the southwest-northeast shoreline in the late afternoon (Redaño et al., 1991).

Figure 1. a) Topography of the UrbClim domain and locations of the meteorological stations~~used~~. The two stations used as ~~references to compute~~ a reference of the ~~Urban-Rural difference~~ urban and rural climates are highlighted with red stars. b) Three WRF domain edges (red squares) and UrbClim domain edges (black contour), together with the topography of the WRF domains with 10, 3.3 and 1.1 km resolution.



The main goals of the paper are:

– Evaluation of a UrbClim simulation of the urban climate ~~of~~ in the city of Barcelona~~with the UrbClim model, by comparing it with~~, driven by reanalysis data at 70 km resolution, against station and satellite data.

– Comparison, both in terms of model skill and computational resources, of this run against a benchmark simulation performed with a state-of-the art mesoscale model, driven with the same reanalysis data.

– Analysis of the sensitivity of the UrbClim simulation to the boundary conditions, comparing ~~two simulations nested in global datasets with different horizontal resolution (70 km and~~ the original run against a simulation driven by a higher resolution forecast dataset (15 km).

– ~~Comparison of the UrbClim simulations with a benchmark simulation carried out with a state-of-the art mesoscale model, focusing both on model skill and computational resources demanded.~~

2 Data and Methodology

2.1 Surface stations

As a first approach in the evaluation of the model performance, we have used data from a set of meteorological stations, ~~4~~ four of them belonging to the Spanish Meteorological Agency (AEMET) and ~~7~~ seven to the Catalan Meteorological Service (SMC). All are well maintained automatic stations, that deliver meteorological data with 10 or 20 min frequency. In the present work, only hourly data were used. The locations of these stations, as well as their names, are displayed in figure 1a, together with the topography.

Station number 5 (El Prat de Llobregat) is chosen to be representative of a rural location near the city. This station is ~~located in the middle of cereal fields ,~~ surrounded by cereal fields (figure 2), located 300 m from the Llobregat river and 650 m from the closest urban area. Station number 6 (el Raval) is instead chosen as the reference urban station. This station is located on the roof of a building, at the city ~~centre of the city~~center of Barcelona, 8.5 ~~Km~~ km away from the rural station. Pictures for the locations of both el Raval[2] and el Prat[3] stations are available on the internet site of SMC. These two points are almost the closest possible rural-urban points located at a similar height. The rural station is at 8 m above sea level, while the urban station is at 33 m, which can account for a difference of 0.15-0.25 °C difference in a standard atmospheric profile. The rural station is located in a delta, and therefore the surrounding topography is flat, ~~which does not favour temperature inversions during night time hours~~with no relevant orographic objects between the two stations. Thus, the differences between these two stations are considered to be representative of the UHI effect in the city of Barcelona.

[2]http://www.meteo.cat/observacions/xema/dades?codi=X4
[3]http://www.meteo.cat/observacions/xema/dades?codi=XL

2.2 Satellite data

The spatial pattern of the simulations is evaluated through data from the MODerate Resolution Imaging Spectoradiometer (MODIS) of the National Aeronautics and Space Administration (NASA) of the United States. Following previous works (Schwarz et al., 2011; Zhou et al., 2015), MODIS datasets MOD11A2 and MYD11A2 (version 5) were downloaded and processed. These correspond to the Terra and Aqua satellites respectively, and are 8-day aggregations of the daily MOD11A1 and MYD11A1 datasets, using only the clear-sky days. The variable considered is Land Surface Temperature (LST), which is derived from the infrared radiance and emissivity estimated from land cover types. A more detailed description of the algorithms is available in Wan (2008).

The LST data were processed considering only the data flagged as “good quality, not necessary to examine more detailed QA” in the Quality Flag provided with the data, and with no cloudy days during the 8 day period. This introduces a bias to certain meteorological conditions (clear-sky days), which is unavoidable. MODIS and UrbClim LST data were interpolated to a 0.01 deg. regular grid for direct comparison. Finally, only the images with less than 14% missing values were used (this percentage does not include the data over the sea which are always missing). This process left a total of 15 values for most gridpoints (supplementary figure 1) over the whole period.

2.3 The UrbClim model

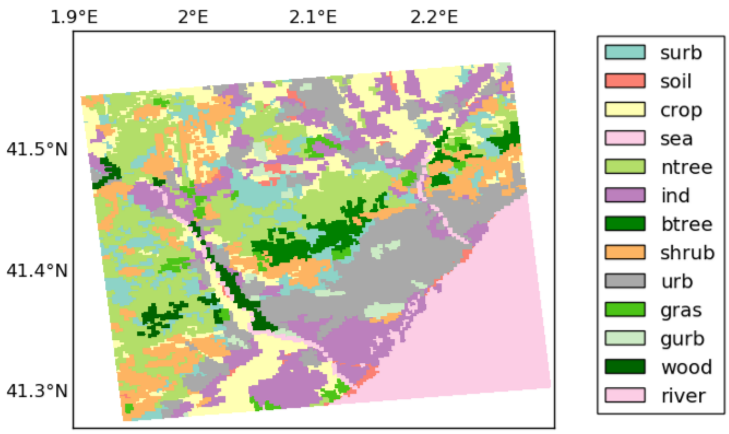
The UrbClim model is designed to ~~reproduce the main features of the urban climate~~ simulate the temperature and heat-stress fields at a city scale requiring the minimum amount of computational power, so that it is possible to perform long runs at a resolution of hundreds of ~~meters~~metres. A detailed description of the model is available in De Ridder et al. (2015).

UrbClim models the lower 3 km of the atmosphere, and consists of a 3-D boundary layer model and land-surface scheme with urban physics. The boundary data needs to be read from a lower resolution model. ~~Given that UrbClimgenerates very small internal variability, the~~

Mesoscale models are tied to their driving models by the boundary conditions. Yet, they develop internal variability (Giorgi and Bi, 2000). In the case of UrbClim, the small size of the domain, and the simplicity of the atmospheric component, greatly reduce the internal variability. The model can be therefore seen as a "wind tunnel". This can be a limitation for the capability of the model to add value to the coarser resolution boundary data, especially for some variables like wind.

On the other hand, the small internal variability has also advantages, as the stability of the simulation is not compromised by the difference in the resolutions of the UrbClim and driving models, as it normally occurs with conventional mesoscale models. Nonetheless, this resolution jump can sometimes affect the quality of the simulation if the driving model does not accurately reproduce the local climate. This trade-off between the internal variability and the computational efficiency will be key for the interpretation of the results in this study.

Figure 2. Distribution of the land use types used in the UrbClim simulations, which were derived from the CORINE dataset.



The land use data, ~~that~~ which are needed to represent the surface properties, are taken from the CORINE dataset~~(http://www.eea.europa.eu/publications/COR0-landcover~~[4]). This dataset is publicly available online, and was produced by the European Environmental Agency at a resolution of 100 m. Figure 2 shows the distribution of the land use classes used by UrbClim.

The land-surface scheme is a standard soil-vegetation-atmosphere model based on ~~De Ridder et al. (1997) with some extensions. In the original scheme,~~ (Ridder and Schayes, 1997), which was extended to fully account for the urban canopy~~was represented as a simple impermeable slab. In the updated version ,~~ . This updated version is described in detail in (De Ridder et al., 2015)~~, this is extended in many ways. However, it can still be considered a simple urban canopy model compared with other existing approaches.~~ . The 3D boundary layer model represents a simplified atmosphere by using the conservation equations for the horizontal momentum, potential temperature, specific humidity and mass. The turbulent vertical diffusion is represented following Hong and Pan (1996). The urban physics use a urban slab, together with a parameterization of the inverse Stanton number. This simple approach is justified in De Ridder et al. (2015), because the heat coefficients can be taken from real-world experiments, rather than the scale experiments that the more detailed urban canopy models use to get the transfer coefficients of walls, roofs and roads.

In contrast, the Single-Layer Urban Canopy Model (SLUCM) included in WRF represents simple symmetrical street canyons with infinite length (Kusaka et al., 2001). It is important to note that the goal of the present study is not no compare this UCM with the one used by UrbClim, but the whole WRF-SUCM modelling system. Thus, the differences found between WRF and UrbClim are not necessarily related to the different approximations used to represent the Urban Canopy.

[4]http://www.eea.europa.eu/publications/COR0-landcover

2.4 Experimental setup

UrbClim

2.3 The UrbClim simulations cover the five warmest months of year 2011, i.e. from May to September. The domain is represented by a horizontal grid with 121x121 points at a resolution of 250 m, with 19 vertical levels ~~up to 3000 m~~ within the 3 lower km of the troposphere (figure 1a). The driving model data is updated every 3 hours. Two simulations have been studied, labeled as UC-ERA and UC-FC. The former is driven by the ERA-Interim reanalyses (Dee et al., 2011), while the latter is driven by the Integrated Forecast System (IFS) version 37r2 global forecast model of the European Centre for Medium-Range Weather Forecasts (ECMWF). In 2011, this model ran with a spectral resolution of T1279 (≃15 km), in contrast with the T255 (≃70 km) of ERA-Interim. Thus, it is able to provide more local details, which can be important ~~give~~ given the aforementioned mesoscale-driven weather of Barcelona.

WRF

The Weather Research and Forecast model is an open-source, non-hydrostatic limited area model (Skamarock et al., 2008). Thanks to its availability, it has a large community of users. These contribute to the development of WRF, which is leaded by the National Center for Atmospheric Research (NCAR). One particularity of this model is that it has a large amount of parameterization schemes, dynamical options ~~,~~ and sub-modules, available to the user to choose among them. These options are set up in a namelist file that must be edited for each simulation. ~~The version of WRF used is the 3.6.1.~~

In the present~~work, WRF~~ , we used the version 3.6.1 of WRF, which was configured to run in three nested domains (figure 1b~~, with~~), with 40 vertical levels and horizontal resolutions of ~~9, 3 and 1 kmand 40 vertical levels. The 250m~~ 10, 3.3 and 1.1 km. The 250 m of UrbClim were not reached because the computational cost was not affordable. However, the simulations were carefully configured to make them comparable with UrbClim: ~~They~~ they were nested in the same dataset (ERA-Interim) and used the same land use (CORINE). WRF land use is ~~by default taken~~ taken by default from the United States Geological Survey (USGS) dataset. Thus, the CORINE land classes were mapped to the USGS 33 classes following table 7.1 of Chrysoulakis et al. (2014). The resulting land use class map is shown in figure 3

Despite being nested in a reanalysis, the regional models tend to generate their own internal variability. While this is necessary to the RCM to add value, it is not convenient to let the model to drift too much from the reanalyses, as these incorporate observations and are accurate descriptions of past atmospheric states. There are two approaches to solve this: using nudging, or restarting the model frequently. In this case, based on the experience of previous works (Menendez et al., 2014; García-Díez et al., 2015), daily 36 hours simulations have been carried out and concatenated, leaving

Figure 3. Distribution of the land use types used in the WRF simulation. They were mapped from the CORINE dataset to the USGS classes.

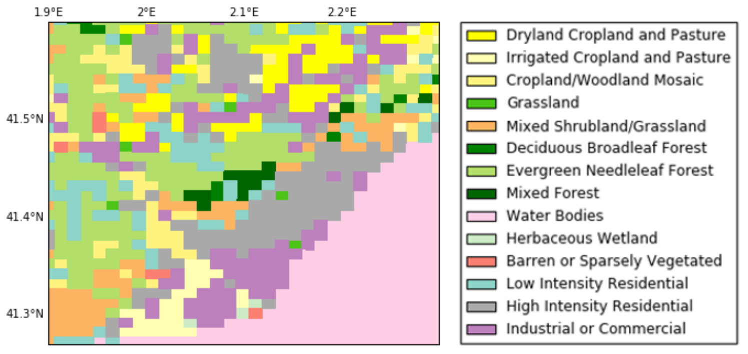


Table 1. Scores of daily mean 2 ~~meter~~ metre temperature for the UC-ERA, UC-FC and WRF simulations and the 11 stations depicted in figure 1. The scores are: Mean bias (model - observed), Root Mean Squared Error (RMSE), and variance ratio (i.e. the variance of the model divided by the variance of the station).

Station	Bias (°C)			RMSE (°C)			Variance ratio		
	UC-ERA	UC-FC	WRF	UC-ERA	UC-FC	WRF	UC-ERA	UC-FC	WRF
1	1.4	0.1	-0.5	2.0	0.9	0.9	1.1	0.8	0.7
2	2.0	0.7	-0.8	1.9	0.8	0.7	1.6	1.1	0.9
3	1.6	0.3	-0.8	1.9	0.8	0.7	1.5	1.1	0.9
4	2.1	1.1	1.1	2.1	1.2	1.1	1.4	1.0	1.0
5	1.1	0.2	-0.3	1.7	0.6	0.7	1.3	1.1	0.9
6	1.2	0.1	-1.0	1.8	0.7	0.6	1.5	1.1	0.9
7	1.8	0.5	0.6	1.8	0.7	0.8	1.4	1.0	0.9
8	0.6	-0.7	0.6	1.8	1.1	0.9	1.1	0.8	0.8
9	0.5	-0.3	-0.2	1.6	0.7	0.7	1.1	0.9	0.9
10	1.2	-0.2	0.3	1.7	0.6	0.7	1.4	1.0	0.9
11	0.3	-0.5	-0.0	1.7	0.8	0.8	1.2	0.9	0.9

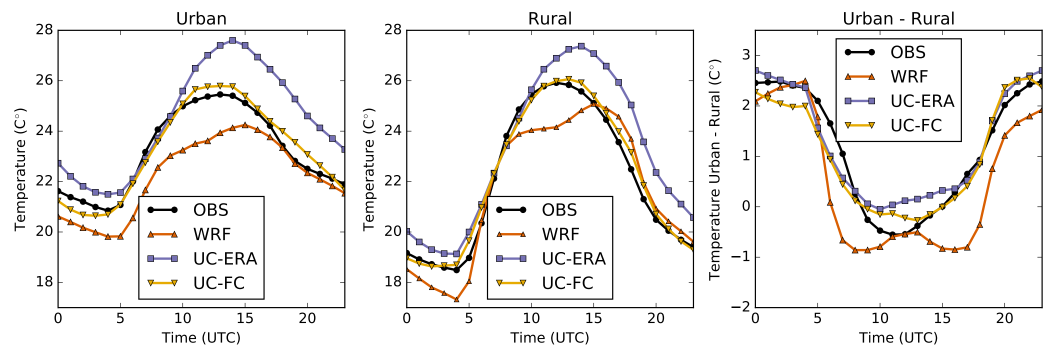
12 hours as spin-up. These simulations cover the same time span as UrbClim, from May to September 2011. Thus, 153 individual simulations have been carried out. To handle them, the WRF4G framework (Fernández-Quiruelas et al., 2015) has been used.

3 Results

3.1 Time series

Table 1 shows ~~the standard~~ scores of daily mean 2 ~~meter~~ m temperature for the UC-ERA, UC-FC and WRF simulations and the 11 stations. The largest errors are found in UC-ERA, which generally overestimates daily temperatures by up to +2°C in some stations. This overestimation is associated

Figure 4. Average daily temperature cycle in the urban (left) and rural (middle) stations. The difference urban ~~-~~ minus rural is shown in the panel on the right. The "UC-ERA", "UC-FC" and "WRF" legend codes are defined in section 2, while "OBS" represents the observation.



with the misrepresentation of the sea breeze, which has larger effect on maximum temperatures (see below).

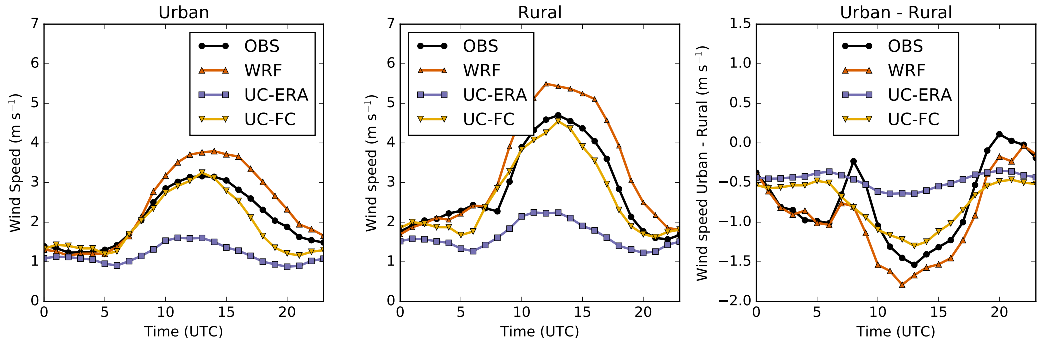
UC-ERA also overestimates the day-to-day variability, having higher Root Mean Square Error (RMSE) than the other runs. Instead, UC-FC and WRF show similar, smaller scores, which indicate the good performance of these simulations.

~~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).~~

Figure 4 shows the average daily cycles for the urban and rural stations, as well as their difference. The average magnitude of the UHI during the night is found to be 2.5°C, which is large enough to have direct impacts on human health during heat wave episodes (Ye et al., 2012). During daytime hours, the UHI is found to decrease down to -0.5°C. Note that this is in very close agreement with the values derived from observational data in (Moreno-García, 1994), despite it used two different reference points. The measurement of the UHI with only two points has some limitations, as it may be sensitive to very local features such as the land use in the vicinity of the stations. However, the representativeness of these points has been carefully checked with high resolution satellite images. In addition, the agreement with previous studies increases our confidence in the results here presented.

UC-ERA tends to overestimate temperatures in both stations after 10 UTC and particularly during daytime hours, but errors in both stations cancel each other, and therefore the UHI magnitude is generally well represented with biases smaller than 0.5°C. The UHI average daily cycle is similar

Figure 5. Same as figure 4, but for wind speed.



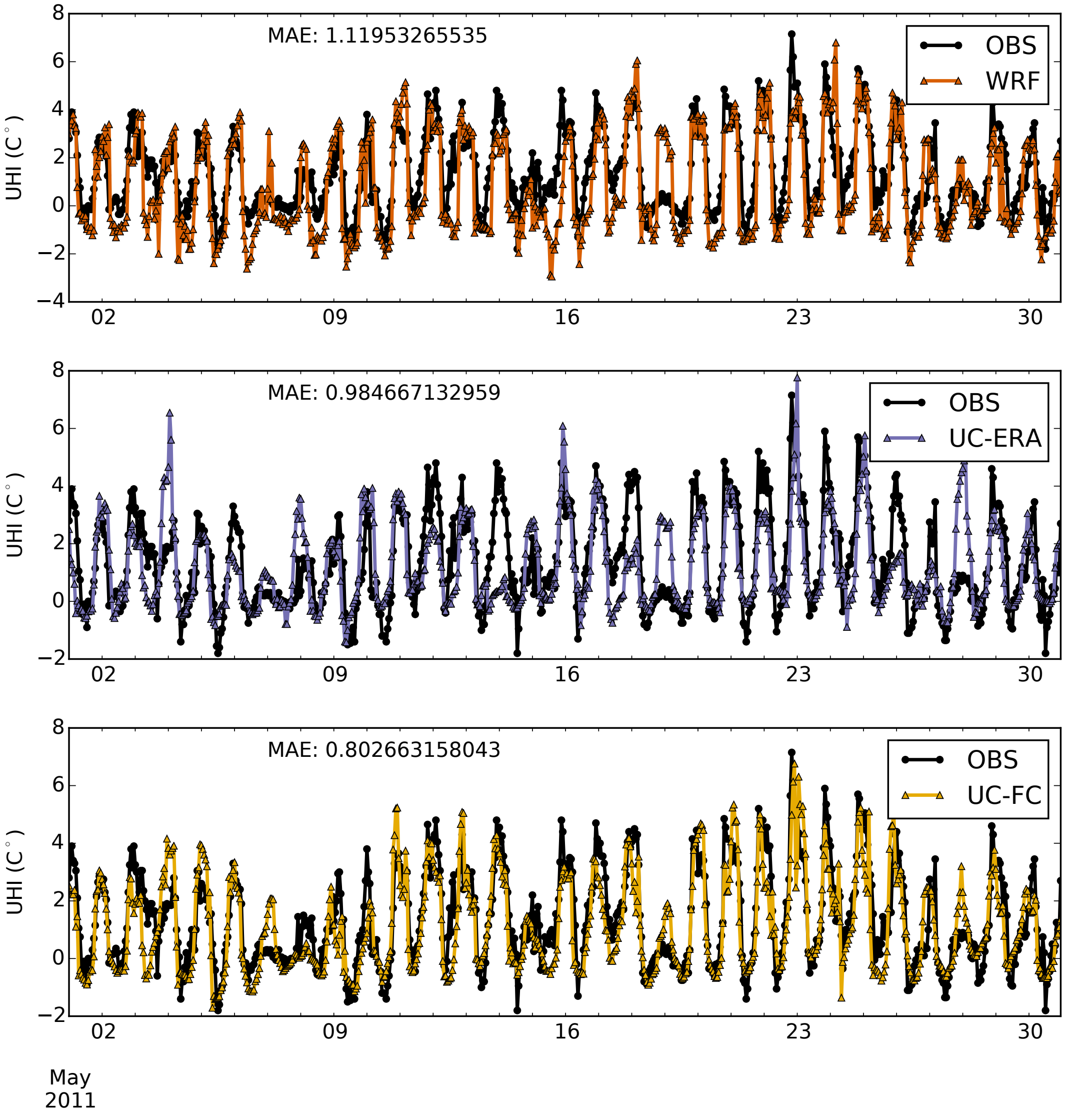
in UC-FC and UC-ERA, but UC-FC does not show any warm bias, and accurately reproduces the observed temperatures of the individual stations.

In the case of WRF, we initially considered the nearest gridpoint to the rural and urban stations, and biases in the three panels were found to be clearly larger than those in UrbClim (not shown). This problem was found to be related to the land use of the gridpoints, which were not representative of the land use of the stations. Indeed, the gridpoint representing the rural station was found to be classified as urban in the land cover map used by WRF. In order to address this problem, we considered a more representative, adjacent gridpoint to represent the rural station, which is used throughout the paper (see supplementary figures 1 and 2 for the details).

Results show that biases in WRF for the individual stations are large and negative ~~during the morning hours, before 15 UTC,~~ throughout the day except in the evening, but comparable in magnitude to those in UC-ERA. In addition, although the UHI at noon is correctly reproduced by WRF, ~~it exhibits large bias maxima of~~ its bias is clearly larger at 7 (-1.5° ~~and -1° at 7 and~~ C) and 17 UTC ~~, respectively~~ (-1°C).

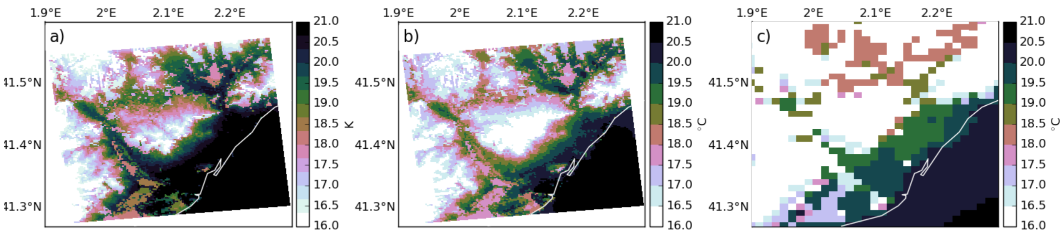
Regarding the wind speed (figure 5), the intensity of the sea breeze is clearly underestimated in ~~the run driven by the ERA-Interim reanalysis~~UC-ERA, with a bias of up to -2.5 $\mathrm{m\,s^{-1}}$ at noon in the rural station. This problem is likely to be related with the coarse resolution of the ERA-Interim driving ~~run~~dataset, which is not able to resolve the sharp daytime, thermally-driven pressure gradient between the continent and the sea. The lack of sea breeze in turn explains the nearly constant daily cycle of the rural minus urban difference in wind speed in UC-ERA. The wind regime is clearly better reproduced in the other simulations. UC-FC accurately reproduces the daily wind cycle in both the urban and rural stations, while WRF overestimates the wind speed by up to 1 $\mathrm{m\,s^{-1}}$ during daytime hours. Regarding the urban minus rural difference, WRF is the model that better captures the hourly evolution of the wind speed. UC-FC ~~does catch well~~ correctly simulates the overall magnitude of the difference, but ~~without reproducing~~ it is not able to reproduce the secondary minima and maxima ~~of~~ at 6 and 8 UTC.

Figure 6. Hourly time series of the 2 meter temperature difference between the urban ~~minus~~ and the rural locations for May ~~2011.~~2011, for WRF (top), UC-ERA (middle) and UF-FC (bottom).



It is interesting to highlight the day-to-day variability of the observed and simulated times series ~~, which are here depicted in figure 6~~ for the month of May (figure 6). The whole period is not shown for clarity, but the same conclusions are applicable for the other months. The daily evolution ~~in~~ of the UHI is well represented in UC-FC and WRF, while biases of the order of up to 4~~°C~~ °C at noon are found in UC-ERA for some specific days. However, the largest Mean Absolute Error (MAE) is found ~~for~~ in WRF (1.11~~°C~~°C), due to the systematic underestimation of the UHI during daytime hours (figure 4). The best MAE score is found in UC-FC (0.80~~°C~~°C), which shows regular skill with almost no large errors in specific days.

Figure 7. Daily minimum temperature averaged over the period May-September 2011 in ~~UC-FC (left~~a) UC-ERA, b) UC-FC and ~~WRF (right~~c) WRF.



From these results, it is clear that the performance of UrbClim is largely improved as a result of the higher resolution of the ECMWF forecast model compared to ERA-Interim (15 km vs. 70 km). It is however unclear how to understand the comparison of WRF with UrbClim. On one hand, UC-ERA and WRF, which are both nested in the same reanalyses, display comparable scores regarding the magnitude of the UHI, althoguh WRF better represents the wind speed and some spatio-temporal features of temperature.

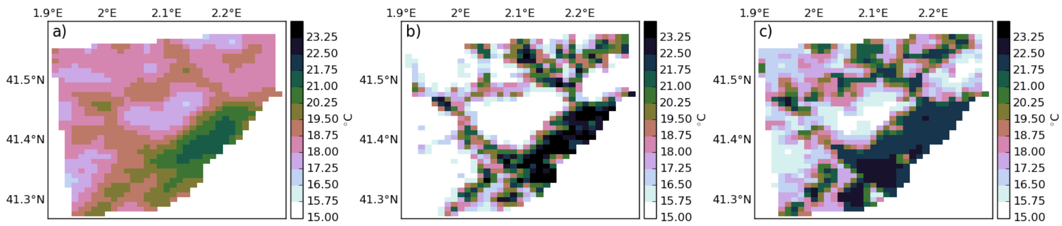
However, WRF performs a full dynamical downscaling down to a resolution of 1.1 km, and therefore, in principle it should be able to achieve an accuracy similar to UC-FC. But, as we have shown, UC-FC exhibits better scores than WRF. Given the large number of factors involved, it is difficult to find an explanation to this result in physical terms. In general, WRF is more biased than UC-FC (figures 4, 5 and 6) and than the ECMWF forecast itself (not shown). It tends to underestimate temperatures and to overestimate the wind speed during the day. As WRF is very customizable, it could be possible, in principle, to find a configuration that removes these biases. However, WRF biases in wind speed are found to be difficult to correct, and research is yet ongoing in this line (García-Díez et al., 2015; Lorente-Plazas et al., 2016).

3.2 Spatial pattern

The evaluation of the spatial variability simulated by the urban climate model is a challenging issue due to the lack of reliable, high-resolution observations. Figure 7 shows the average daily minimum temperatures for ~~the~~ UC-ERA, UC-FC and WRF, for the 5 months considered. Although both models are able to resolve the main features of the UHI of Barcelona, the surrounding cities and the airport, the UrbClim run at a resolution of ~~250m~~ 250 m provides much more detailed information, e.g. a clear representation of the hill to the southwest of the city centre. UC-ERA is generally warmer than UC-FC, due to the mis-representation of winds, but the spatial patterns are very similar. In addition, the temperature difference is similar in the urban and rural sites.

Unfortunately, the scarcity of surface observations did not allow us to evaluate the spatial patterns at the screen level, and therefore we evaluated the spatial variability of the model by analysing the MODIS satellite LST, as described in section 2.2.

Figure 8. Land surface temperature averaged during nighttime hours over the period May-September 2011 in MODIS (left), ~~UC-FC~~ UC-ERA (center) and WRF (right).



During the night, both ~~UrbClim~~ UC-ERA and WRF have been found to overestimate LST over urban areas~~and, thus,~~, and therefore also the LST UHI (figure 8). This is surprising, given the small error found in the ~~validation~~ evaluation of the screen level UHI. Other studies (Zhou et al., 2015) ~~also~~ found small errors when comparing MODIS and UrbClim LST in and around the city of London. ~~advectionAs~~ As mentioned in the introduction, measuring LST over urbanised areas is challenging, due to the uncertainties associated with the measurement of both the radiation and the emissivity. The bias outside the urban areas is found to be ~~small~~ relatively small in WRF (figure 8), and slightly negative in UC-ERA, while the spatial patterns are reasonably similar between the models and the satellite data. Determination of emissivity over urban areas is notoriously difficult and subject to ~~a large uncertainty~~large uncertainties, which could explain at least part of model deviation ~~for~~ of LST. The spatial Pearson correlations between the observed and simulated fields are ~~0.74~~0.77±~~0.06 for UC-FC and 0.69~~0.025 for UC-ERA and 0.70±~~0.07~~ 0.03 for WRF, where the confidence bounds were computed with bootstrapping (1000 samples). Thus, ~~UrbClim correlation~~ the spatial correlation in UC-ERA is higher, ~~but~~ and the difference is ~~not statistically significant, as the confidence bounds overlap.~~ statistically significant. However, taking into account the above-mentioned differences in the bias, we conclude that the performance of the spatial pattern in WRF and UC-ERA are comparable.

It is worth mentioning that the MODIS LST appears to have an effective resolution coarser than 1 km, given that the spatial patterns are smooth and do not resolve many detailed features. Finally, as mentioned in the introduction, LST and Surface Air Temperature (SAT) Urban Heat Islands are not equivalent, and are driven by different phenomena. Thus, it is also possible that models that reproduce the SAT UHI correctly generate at the same time a biased LST UHI.

3.3 Computational resources

In this section, the computational resources ~~consumed~~ required by UrbClim and WRF are compared. The comparison is not fully trivial because UrbClim does not currently support running in parallel~~.~~ ~~This~~, which can be seen as an important drawback. However, UrbClim does not require a long spin-up, and therefore the simulations can be parallelised just by splitting the time period in subperiods and run the corresponding simulations simultaneously in different machines or nodes.

For a direct comparison, both models were run in the local cluster of the Institut Català de Ciències del Clima (IC3), while the main UrbClim runs used in the paper were carried out in the VITO cluster. The IC3 cluster is made of 48 homogeneous server blades, having each of them two "quad core" processors, 48GB of memory, 146GB of disk space and fast network interconnect (Infiniband). The blade model is Sun Blade X6270 (see http://www.oracle.com/us/products/servers-storage/servers/blades/sun-blade-x6270-m2-ds-080923.pdf for a full description) equipped with Xeon (Nehalem) X5570 processors. ~~With this~~ Results are summarized in table 2. With these settings, a WRF simulation of 36 hours took 2.5 hours to finish (using an average of 10 simulations), including the preprocess carried out with the WRF preprocessor (WPS). This preprocess was run in serial, in 1 core, while WRF was run in 16 cores, this is, two blades. Thus, the total serial equivalent wall-time was 40 hours, assuming perfect scaling (the real value will be somewhat below). WRF was compiled using the Intel fortran compiler version 14.0.1 with the Intel MPI Library for Linux OS, Version 4.1 Update 3.

Regarding UrbClim, for this test, it has been compiled with the same compiler and run in the same cluster. A 36-hours simulation with UrbClim took 0.3 hours to finish (average of 10 simulations) running in one core. Thus, UrbClim running at 250 m resolution is found to be 133 times faster than WRF at 1 km resolution. This enables downscaling large climate change ensembles for a big collection of cities.

Note that the UrbClim speed is not only explained by the smaller number of gridpoints (table 2), but especially because of the simplicity of the dynamical core, and the smaller number of parameterizations, compared to WRF.

4 Conclusions

In the present work, we have evaluated the performance of a boundary-layer urban climate model (UrbClim) for the warm season in the city of Barcelona. We were particularly interested in the study of the urban heat island (UHI) effect, given that it represents a major source of health problems in summer for vulnerable people living in urban environments (e.g. heat stress, temperature-related mortality, pollution, vector-borne diseases). We have ~~analysed the effect of the model resolution in the driving simulation~~ compared this model (UC-ERA~~and UC-FC) , and compared these runs~~) with the output of a regional climate model (WRF)~~.~~ , and analysed the effect of the model resolution in the driving simulation (UC-ERA vs. UC-FC). All these simulation have been evaluated against observations from meteorological stations and satellite data (MODIS), in order to analyse the temporal and spatial variability of the UHI effect, respectively.

The main conclusions of our work can be summarised as follows:

- The average UHI in the city of Barcelona during the warm season (May-September) reaches 2.5~~°C~~ °C at night.

This is relevant for the study of climate impacts, given that it increases the stress to the vulnerable population and for the health care systems under extreme conditions.

- UrbClim correctly reproduces the UHI of Barcelona when it is nested to the coarse dataset of ERA-Interim, with some systematic biases. When it is nested to a higher resolution model (ECMWF IFS), UrbClim additionally reproduces well the temperature evolution of the individual rural and urban stations used for the calculation of the UHI.

- UC-ERA and WRF reproduce the UHI of Barcelona with comparable skill, but WRF ~~reproduces the UHI intensity nearly as well as UrbClim, but it~~ provides less detailed spatial information ~~consuming~~ and it demands much larger computational resources.

- The ~~realism of the~~ spatial pattern of LST is similar in UrbClim and WRF, ~~when it is validated against MODIS data,~~ even though significant biases are found in both models when they are evaluated against MODIS data.

In conclusion, ~~the~~ UrbClim has been found to be well suited for the numerical description of the UHI of Barcelona, providing an accurate description of the temperature field. The choice between UrbClim and WRF for the simulation of the urban environment largely depends on the type of variable and process that is to be analysed. WRF has the advantage of providing a more detailed and complete description of atmospheric winds and rainfall, which ~~may be~~ is required in some applications (e.g. pollutant dispersion~~). Apart from this~~, urban effect in rainfall). On the other hand, UrbClim has been ~~found to be an optimal tool for the numerical description of the~~ proven to be as accurate as WRF on reproducing the UHI of Barcelona during the warm season, and several orders of magnitude faster. This opens the door to the performance of multi-decadal simulations of urban heat stress in a large number of cities at a reasonable computational cost, using multi-scenario and multi-GCM ensembles to account for uncertainty, and testing for urban adaptation scenarios.

We found that, in cities affected by strong mesoscale flows (e.g. sea breeze) such as Barcelona, ~~providing an accurate description of the temperature field thatis generally better than that in WRF.However,~~ it must be taken into account that ~~, if nesting UrbClim in a low resolution model, there will be~~ UrbClim will be subject to inaccuracies caused by the ~~misrepresentation~~ mis-representation of the wind, ~~specifically the sea breeze daily cycle. The sea breeze is important for reproducing the climate of Barcelona in summer, where the influence of mesoscale processes is strong.~~ in case that it is nested in a low resolution model.

Note that this ~~problem is a particularity~~ is a specific problem of Barcelona, ~~as~~ given that it has not been found in other European cities where UrbClim ~~nested in~~ driven by ERA-Interim has been successfully tested (De Ridder et al., 2015; Lauwaet et al., 2016; Zhou et al., 2015).

From these results, it is reasonable to infer that the skill of UrbClim, and probably of other similar urban ~~climate~~ boundary layer 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.

Table 2. Summary of the results of the benchmarking. The number of gridpoints represents the total number of gridpoints in the model domain, which is 121*121*19 in UrbClim, and 100*80*40 + 136*112*40 + 121*97*40 in case of WRF (taking into account the three nested domains).

Model	Number of gridpoints	Horizontal resolution	Time step	Wall-time for 36h
UrbClim	278,179	0.25 km	adaptative[5]	0.3 h
WRF	1,398,760	1.1 km (10x3.3x1.1)	60 s	40 h

Code and data availability

The Urbclim source code is not publicly available. In order to access it, a specific agreement needs to be signed with VITO. Please contact koen.deridder@vito.be for more details. The WRF model is an open source model, and its code is freely available upon registration in http://www2.mmm.ucar.edu/wrf/users/download/get_source.html. Weather station data from the Catalan and Spanish meteorological agencies is available for research purposes upon request in dades@meteo.cat and https://sede.aemet.gob.es/AEMET/es/GestionPeticiones/home respectively. MODIS data ~~was~~ were downloaded from the "Reverb" NASA tool http://reverb.echo.nasa.gov, where it is freely available upon registration. The CORINE land cover is available in the EEA website http://www.eea.europa.eu/publications/COR0-landcover free of charge for both commercial and non-commercial purposes.

Acknowledgements. The work described in this paper has received funding from the European Community's 7th Framework Programme under Grant Agreements No. 308299 (NACLIM) and 308291 (EUPORIAS). ~~It has also received~~ J. Ballester gratefully acknowledges funding from the European Commission through a Marie Curie International Outgoing Fellowship (~~European Commission,~~ project MEMENTO from the ~~355~~ FP7-PEOPLE-2011-IOF call)~~and the Catalan Government Beatriu de Pinós~~ , and from the European Commission and the Catalan Government through a Marie Curie - Beatriu de Pinós Fellowship (project 00068 from the BP-DGR-2014-B call).

[5] 20 s for the soil scheme and adaptative for the atmosphere, using the Courant-Friedrichs-Lévy stability criterion.

References

Argüeso, D., Evans, J. P., Pitman, A. J., and Luca, A. D.: Effects of City Expansion on Heat Stress under Climate Change Conditions, PLOS ONE, 10, e0117 066, doi:10.1371/journal.pone.0117066, 2015.

Arnfield, A. J.: Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island, Int. J. Climatol., 23, 1–26, doi:10.1002/joc.859, 2003.

Ballester, J., Douville, H., and Chauvin, F.: Present-day climatology and projected changes of warm and cold days in the CNRM-CM3 global climate model, Clim Dyn, 32, 35–54, doi:10.1007/s00382-008-0371-0, 2009.

Ballester, J., Giorgi, F., and Rodó, X.: Changes in European temperature extremes can be predicted from changes in PDF central statistics, Climatic Change, 98, 277–284, doi:10.1007/s10584-009-9758-0, 2010a.

Ballester, J., Rodó, X., and Giorgi, F.: Future changes in Central Europe heat waves expected to mostly follow summer mean warming, Clim Dyn, 35, 1191–1205, doi:10.1007/s00382-009-0641-5, 2010b.

Ballester, J., Robine, J.-M., Herrmann, F. R., and Rodó, X.: Long-term projections and acclimatization scenarios of temperature-related mortality in Europe, Nat Commun, 2, 358, doi:10.1038/ncomms1360, 2011.

Chen, F., Kusaka, H., Bornstein, R., Ching, J., Grimmond, C. S. B., Grossman-Clarke, S., Loridan, T., Manning, K. W., Martilli, A., Miao, S., Sailor, D., Salamanca, F. P., Taha, H., Tewari, M., Wang, X., Wyszogrodzki, A. A., and Zhang, C.: The integrated WRF/urban modelling system: development, evaluation, and applications to urban environmental problems, Int. J. Climatol., 31, 273–288, doi:10.1002/joc.2158, 2011.

Ching, J. K. S.: A perspective on urban canopy layer modeling for weather, climate and air quality applications, Urban Climate, 3, 13–39, doi:10.1016/j.uclim.2013.02.001, 2013.

Chrysoulakis, N., de Castro, E. A., and Moors, E. J.: Chapter 7: Meso-scale meteorological models in the urban context, in: Understanding Urban Metabolism: A Tool for Urban Planning, pp. 69–79, Routledge, 2014.

De Ridder, K., Lauwaet, D., and Maiheu, B.: UrbClim – A fast urban boundary layer climate model, Urban Climate, 12, 21–48, doi:10.1016/j.uclim.2015.01.001, 2015.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, 2011.

Dousset, B., Gourmelon, F., Laaidi, K., Zeghnoun, A., Giraudet, E., Bretin, P., Mauri, E., and Vandentorren, S.: Satellite monitoring of summer heat waves in the Paris metropolitan area, Int. J. Climatol., 31, 313–323, doi:10.1002/joc.2222, 2011.

Fernández-Quiruelas, V., Fernández, J., Cofiño, A. S., Blanco, C., García-Díez, M., Magariño, M., Fita, L., and Gutiérrez, J. M.: WRF4G: WRF experiment management made simple, Geosci. Model Dev. Discuss., 8, 6551–6582, doi:10.5194/gmdd-8-6551-2015, 2015.

Gabriel, K. M. and Endlicher, W. R.: Urban and rural mortality rates during heat waves in Berlin and Brandenburg, Germany, Environmental Pollution, 159, 2044–2050, doi:10.1016/j.envpol.2011.01.016, 2011.

García-Díez, M., Fernández, J., San-Martín, D., Herrera, S., and Gutiérrez, J.: Assessing and Improving the Local Added Value of WRF for Wind Downscaling, J. Appl. Meteor. Climatol., doi:10.1175/JAMC-D-14-0150.1, 2015.

Georgescu, M., Morefield, P. E., Bierwagen, B. G., and Weaver, C. P.: Urban adaptation can roll back warming of emerging megapolitan regions, PNAS, 111, 2909–2914, doi:10.1073/pnas.1322280111, 2014.

Giorgi, F.: Climate change hot-spots, Geophysical research letters, 33, 2006.

Giorgi, F. and Bi, X.: A study of internal variability of a regional climate model, J. Geophys. Res., 105, 29 503–29 521, doi:10.1029/2000JD900269, 2000.

Hong, S.-Y. and Pan, H.-L.: Nonlocal Boundary Layer Vertical Diffusion in a Medium-Range Forecast Model, Mon. Wea. Rev., 124, 2322–2339, doi:10.1175/1520-0493(1996)124<2322:NBLVDI>2.0.CO;2, 1996.

Huszar, P., Halenka, T., Belda, M., Zak, M., Sindelarova, K., and Miksovsky, J.: Regional climate model assessment of the urban land-surface forcing over central Europe, Atmospheric Chemistry and Physics, 14, 12 393–12 413, doi:10.5194/acp-14-12393-2014, 2014.

Kikegawa, Y., Genchi, Y., Kondo, H., and Hanaki, K.: Impacts of city-block-scale countermeasures against urban heat-island phenomena upon a building's energy-consumption for air-conditioning, Applied Energy, 83, 649–668, doi:10.1016/j.apenergy.2005.06.001, 2006.

Kolokotroni, M., Ren, X., Davies, M., and Mavrogianni, A.: London's urban heat island: Impact on current and future energy consumption in office buildings, Energy and Buildings, 47, 302–311, doi:10.1016/j.enbuild.2011.12.019, 2012.

Kusaka, H., Kondo, H., Kikegawa, Y., and Kimura, F.: A Simple Single-Layer Urban Canopy Model For Atmospheric Models: Comparison With Multi-Layer And Slab Models, Boundary-Layer Meteorology, 101, 329–358, doi:10.1023/A:1019207923078, 2001.

Kusaka, H., Hara, M., and Takane, Y.: Urban Climate Projection by the WRF Model at 3-km Horizontal Grid Increment: Dynamical Downscaling and Predicting Heat Stress in the 2070's August for Tokyo, Osaka, and Nagoya Metropolises, Journal of the Meteorological Society of Japan. Ser. II, 90B, 47–63, doi:10.2151/jmsj.2012-B04, 2012.

Lauwaet, D., Hooyberghs, H., Maiheu, B., Lefebvre, W., Driesen, G., Van Looy, S., and De Ridder, K.: Detailed Urban Heat Island Projections for Cities Worldwide: Dynamical Downscaling CMIP5 Global Climate Models, Climate, 3, 391–415, doi:10.3390/cli3020391, 2015.

Lauwaet, D., De Ridder, K., Saeed, S., Brisson, E., Chatterjee, F., van Lipzig, N. P. M., Maiheu, B., and Hooyberghs, H.: Assessing the current and future urban heat island of Brussels, Urban Climate, 15, 1–15, doi:10.1016/j.uclim.2015.11.008, 2016.

Lee, S.-H., Kim, S.-W., Angevine, W. M., Bianco, L., McKeen, S. A., Senff, C. J., Trainer, M., Tucker, S. C., and Zamora, R. J.: Evaluation of urban surface parameterizations in the WRF model using measurements during the Texas Air Quality Study 2006 field campaign, Atmos. Chem. Phys., 11, 2127–2143, doi:10.5194/acp-11-2127-2011, 2011.

Lorente-Plazas, R., Jiménez, P. A., Dudhia, J., and Montávez, J. P.: Evaluating and improving the impact of the atmospheric stability and orography on surface winds in the WRF model, Mon. Wea. Rev., doi:10.1175/MWR-D-15-0449.1, 2016.

McMichael, A. J., Woodruff, R. E., and Hales, S.: Climate change and human health: present and future risks, The Lancet, 367, 859–869, doi:10.1016/S0140-6736(06)68079-3, 2006.

Meehl, G. A. and Tebaldi, C.: More Intense, More Frequent, and Longer Lasting Heat Waves in the 21st Century, Science, 305, 994–997, doi:10.1126/science.1098704, 2004.

Menendez, M., García-Díez, M., Fita, L., Fernández, J., Méndez, F. J., and Gutiérrez, J. M.: High-resolution sea wind hindcasts over the Mediterranean area, Clim Dyn, 42, 1857–1872, doi:10.1007/s00382-013-1912-8, 2014.

Moreno-García, M. C.: Intensity and form of the urban heat island in Barcelona, International Journal of Climatology, 14, 705–710, 1994.

Oke, T. R.: The energetic basis of the urban heat island, Q.J.R. Meteorol. Soc., 108, 1–24, doi:10.1002/qj.49710845502, 1982.

Redaño, A., Cruz, J., and Lorente, J.: Main features of the sea-breeze in Barcelona, Meteorl. Atmos. Phys., 46, 175–179, doi:10.1007/BF01027342, 1991.

Ridder, K. D. and Schayes, G.: The IAGL Land Surface Model, J. Appl. Meteor., 36, 167–182, doi:10.1175/1520-0450(1997)036<0167:TILSM>2.0.CO;2, 1997.

Robine, J.-M., Cheung, S. L. K., Le Roy, S., Van Oyen, H., Griffiths, C., Michel, J.-P., and Herrmann, F. R.: Death toll exceeded 70,000 in Europe during the summer of 2003, Comptes Rendus Biologies, 331, 171–178, doi:10.1016/j.crvi.2007.12.001, 2008.

Schwarz, N., Lautenbach, S., and Seppelt, R.: Exploring indicators for quantifying surface urban heat islands of European cities with MODIS land surface temperatures, Remote Sensing of Environment, 115, 3175–3186, doi:10.1016/j.rse.2011.07.003, 2011.

Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Barker, D., Duda, M., Wang, W., and Powers, J.: A description of the Advanced Research WRF Version 3, Tech. rep., NCAR, 2008.

Voogt, J. A. and Oke, T. R.: Effects of urban surface geometry on remotely-sensed surface temperature, International Journal of Remote Sensing, 19, 895–920, doi:10.1080/014311698215784, 1998.

Wan, Z.: New refinements and validation of the MODIS Land-Surface Temperature/Emissivity products, Remote Sensing of Environment, 112, 59–74, doi:10.1016/j.rse.2006.06.026, 2008.

Ye, X., Wolff, R., Yu, W., Vaneckova, P., Pan, X., and Tong, S.: Ambient temperature and morbidity: a review of epidemiological evidence, Environmental health perspectives, 120, 19–28, 2012.

Zhao, L., Lee, X., Smith, R. B., and Oleson, K.: Strong contributions of local background climate to urban heat islands, Nature, 511, 216–219, doi:10.1038/nature13462, 2014.

Zhou, B., Lauwaet, D., Hooyberghs, H., De Ridder, K., Kropp, J. P., and Rybski, D.: Assessing seasonality in the surface urban heat island of London, J. Appl. Meteor. Climatol., doi:10.1175/JAMC-D-15-0041.1, 2015.