Response to reviewer #1

1. Resolved questions: The addition of Figure 2a elucidates the surface cover classification system used in the NLSMC dataset and sufficiently addresses the original concern. The added literature references are a helpful addition. Thank you also for color-coding the text changes.

R: The authors would also like to thank the reviewer for the constructive suggestion.

2. Please show a figure or statistics on how the observed quantities varied over the days evaluated in July, 2012 (wind speed and direction, temperature; it would also be helpful to know when the ground was saturated or if the sealed surface was wet). It was also mentioned that certain periods were excluded when rainfall occurred in the simulation [L 295]; could you please indicate how many days of July, 2012 were excluded? Was this one-month run executed from 01-31 July with a preceding 24-hr spin-up period as mentioned for the focus period in Section 3 [L 215-217]? Could you please also show the corresponding results (full-month analysis) for the urban stations as you've done for the non-urban stations in Table 3?

R: As suggested, Table S1 has been added, showing bias, RMSE and R^2 calculated using WRF-UCM2D-simulated wind data at stations 466880, 466920, 466940 and 467080 by for the month of July 2010. Indeed, with the complex geographic structure of northern Taiwan (Figure 1), bias could be sensitive to wind fields due to local topography or some isolated tall buildings. Results shown in Table S1 and Figure S1 of four CWB standard stations (466880, 466920, 466940 and 467080) with uniform environment in the simulation domain indicate that the model in general captures well the wind field in July, evidenced by the u-component having a higher R^2 (0.57-0.84) than the v-component (0.39-0.58).

As requested, Tables 3 and 4 have been added showing one-month air temperature simulation for non-urban areas and urban areas, respectively with data of 15 days on which rainfall was recorded excluded from the analysis.

Yes, the one-month simulation run was executed from 01-31 July, 2010 with a preceding 24-hr spin-up beginning on 30 June, 2010. The Four-Dimensional Data Assimilation (FDDA) scheme was applied in the coarse domain in this study **(L149-152)**.

	466880		466920		466940		467080	
	U	V	U	V	U	V	U	V
Bais	-0.15	-0.12	-0.10	-0.44	-0.70	0.27	-0.29	0.70
RMSE	1.45	1.09	1.60	1.28	1.96	1.87	1.60	1.86
R^2	0.80	0.40	0.84	0.39	0.65	0.58	0.57	0.49

Table S1. Bias, RMSE and R^2 calculated using WRF-UCM2D-simulated wind data at stations 466880, 466920, 466940 and 467080 for the month of July 2010.

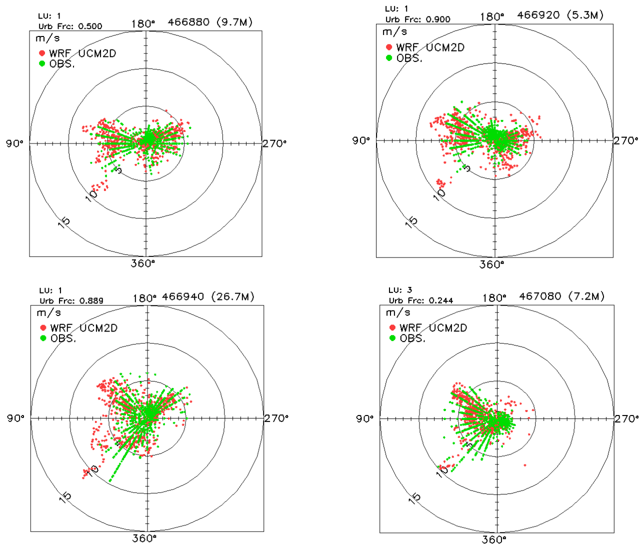

Figure S1. Scatter plots between observed and WRF-UCM2D-simulated temperatures at stations 466880, 466920, 466940 and 467080 for the month of July 2010.

3. One concern I have, particularly for the one-month simulation, is the drastic change in resolution between the boundary values and the parent domain (from around 50 km to 3 km). Are you applying some kind of relaxation zone at the lateral boundaries? Do the horizontal and vertical winds in grids 1 and 2 look realistic relative to available surface analyses or Reanalysis? I would suggest at least commenting on this and any potential impact in the manuscript.

R: The reviewer's concern arose because the original manuscript did not mention that the Four-Dimensional Data Assimilation (FDDA) scheme has been employed to maintain the NCEP GFS data set in domain 1 (coarse domain) for long-term simulation **(L149-152)**. Grid nudging is a major component of the FDDA. The grid-nudging method is specifically three-dimensional analysis nudging, whereby the atmospheric model is nudged towards time- and space-interpolated analyses using a point-by point relaxation term. The model is run with grid-nudging for long periods, e.g., months, to provide a four-dimensional meteorologically self-consistent dataset that also stays on track with the driving analyses. The nudging scheme has proved to be useful for preventing the drift of the regional model away from the large-scale driving field (von Storch et al., 2000; Mabuchi et al., 2002; Miguez-Macho et al., 2004) and for limiting large-scale error growth (Lo et al., 2008).

References:

Lo JCF, Yang ZL, Pielke RA. 2008. Assessment of three dynamical climate downscaling methods using the Weather Research and Forecasting (WRF) model, *J Geophys Res-Atmos*, *113*(D9).

Mabuchi K, Sato Y, Kida H. 2002. Verification of the climatic features of a regional climate model with BAIM, *J Meteorol Soc Jpn*, *80*(4), 621-644.

Miguez-Macho G, Stenchikov GL, Robock A. 2004. Spectral nudging to eliminate the effects of domain position and geometry in regional climate model simulations, *J Geophys Res-Atmos*, *109*(D13).

von Storch H, Langenberg H, Feser F. 2000. A spectral nudging technique for dynamical downscaling purposes, *Mon Weather Rev*, *128*(10), 3664-3673.

4. When describing the features that affect urban climate [starting around L 69], please specifically define the urban heat island (UHI) in the manner in which you intend to identify and analyze it (e.g., difference in 2 m nocturnal minimum temperature at static locations, difference in lowest atmospheric level temperature over time averaged over pre-determined areas, etc.).

R: As suggested, the definition of urban heat island (UHI) has been added (L70-71).

5. What does the diagnostic 2 m temperature [L 348] mean in the urban areas where surface roughness elements exceed this height? I believe in WRF these 2 m and 10 m diagnostic quantities are meant to reflect altitude above the displacement height and are calculated using Monin-Obukhov similarity theory (which is applicable at altitudes well above the height of any roughness elements), which would create a contradiction within the urban canopy. I would suggest using an alternative temperature measure for comparison between urban and non-urban areas.

R: The WRF thermodynamic variables are not directly simulated at 2 m, but diagnosed from values at the land surface and the lowest model layer (Hu et al. 2010). WRF-UCM further considers the forcing from artificial surface (building, roof, wall and road). Building canopy is assumed to resemble aerodynamic roughness, thus implying that the complex urban canopy layer is replaced by a roughness number rendering information regarding the quantities within the canopy layer unavailable (Mirzaei and Haghighat, 2010). Characteristics within the roughness sub-layer vary with a horizontal distance scale determined by inter-element spacing, rather than height and vertical temperature gradient (Arnfield, 2003). In this study, the building height, i.e., the canopy layer, is 6 m**.** The "2-m diagnostic air temperature" is calculated using the energy budget equation in the urban copy model expressed as equation (1) (**Line 388**). The Fsh from Noah LSM is averaged with the sensible heat flux from UCM weighted by the urban fraction to yield the representative Fsh (Chen et al. 2011) **(L382-393)**. The representative skin temperature (Tsk) of a model grid is averaged with the Tsk from Noah LSM and the skin temperature of the artificial surfaces (i.e., roof, wall and road) in the UCM weighted by their respective coverage (Chen et al. 2011). A schematic of the single-layer UCM from Chen et al. (2011) is shown below (Figure S2). Comparisons of temperature simulation performance for urban and non-urban stations are shown in Tables 3 and 4, respectively.

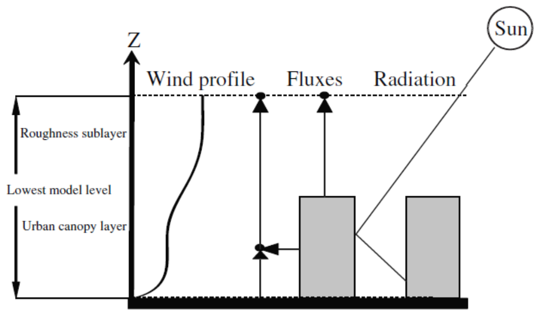

Figure S2. A schematic of the single-layer urban canopy model. (taken from Chen et al. 2011)

References:

Arnfield, A.J., 2003: 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.

Hu X.M., J. W. Nielsen-Gammon, F. Zhang 2010: Evaluation of three planetary boundary layer schemes in the WRF model. *Journal of Applied Meteorology and Climatology*, 49,1831-1844.

Mirzaei P. A., and F. Haghighat 2010: Approaches to study urban heat island-Abilites and limitations, *Building and Environment.* 45,2192-2201.

Technical Corrections

1. The wording in the Abstract [L 30-32] was left unchanged. The second use of this sentence later on [L 89-91] was revised, but is still unclear. I understand the statement "such not only may lead to over- or underestimation [of AH]",and the revision "The simplification . . . underestimation", but I do not understand the meaning of the second part of the sentence ("the temperature difference between urban and non-urban areas has also been neglected") or its connection to the first part. Are both parts referring to AH? Are you implying that the default UCM treatment of AH may lead to local errors in both time (first part) and space (second part)? Or, in the second part, are you referring to the impact of spatially varying urbanized area on model estimation of temperature? I recommend re-writing this sentence (L 30-32 and L 89-91), perhaps dividing into more than one sentence if the topic covers both AH and urban fraction.

R: Both the treatment of AH and urban fraction in the original UCM oversimplify their respective forcing. More detailed explanations have been added to Section 2.2(a) and (b) **(L155-L211)**.

Here we modify the wording in abstract and introduction as following:

L30-32:" Such not only may lead to over- or underestimation of urban fraction and AH in urban and non-urban areas, spatial variation also affects the model-estimated temperature."

L92-95:"Similarly, the UCM assumes the distribution of anthropogenic heat (AH) to be constant and includes only the urban data. Such simplification may lead to over- or underestimation thus affecting the accuracy of model temperature estimations (detailed description in Section 2.2)"

2. The term "model grid net" is used frequently, but I am not clear on what is meant by 'net'. Is this used as a synonym for 'mesh'? The term "model grid" works well also.

R: As suggested, the phrase "model grid net" has been replaced with " model grid" in the revised manuscript. (L161, L184, L191, L192, L307, L339, L372, L374, L375, L376, L397, L400, L401, L402, L406, L408,L468)

3. L 35: Please specify that your variable of interest is temperature

R: The text has been amended. (L36)

4. L 37: Please specify what ‘R’ is.

R: The text has been amended. (L38-39)

5. L 73: “. . .in their study on the urban boundary layer” OR “. . .in their study on urban boundary layers.”

R: The text has been dropped.

6. L 75: “. . .by implementing an urban canopy parameterization. . .”

R: The text has been dropped.

7. L 104: What is meant by ‘scale’? Population? Area?

R: It meant area. The text has been amended. (L108).

8. L 116: Were you using WRF v3 or higher? Check for a later user guide reference (e.g., Skamarock 2008).

R: The WRF version in this study is V3.2.1. (L119).

9. L 159-160: Sentence fragment – missing verb. . .”. . .categorized as rural are (?) totally neglected.” I might suggest separating out this part after the comma into a new sentence.

R: The text has been dropped.

10. L 215: ‘till’ - > ‘until’

R: The text has been amended as suggested. (L219).

11. Section 3a: how is model air temperature derived? Is this the canyon air temperature estimated by the WRF UCM or the lowest temperature of the lowest model level above the urban canopy height (if so, please specify the level and the

mean building height for this class used in the UCM)?

R: Please see answer to Question No. 5

12. Thank you for clarifying what makes a station ‘urban’ versus ‘non-urban’ [L 231-233]. The wording is slightly difficult to follow. May I suggest: “A station is categorized as ‘urban’ if its location in the inner-most model grid has an urban fraction greater than 0.5 and is categorized as ‘non-urban’ if the urban fraction is less than 0.4.“

R: The text has been amended. (L235-236).

13. L 241: Please clarify what is meant by a “heat wave.”

R: The text has been amended. (L245).

14. L 251: “A similar phenomenon. . .”

R: The text has been amended as suggested. (L256).

15. Figure 4: Could you add a third column showing the difference between WRF-UCM and WRF-UCM2D to make it easier to see the magnitude and location of the difference?

R: As suggested, a third column containing Fig. 4(c), (f) and (i) has been added, showing the temperature difference between WRF-UCM2D and WRF-UCM.

16. L 294: “a whole month simulation of July, 2012 was conducted” OR “whole month simulations of July, 2012 were conducted”

R: The text has been amended. (L344-345).

17. L 294: Omit ‘further’

R: The text has been amended as suggested. (L344-345).

18. L 295: “the hourly data” – does this mean all model output? If so

please clarify.

R: The text has been amended (L349).

19. L 295: “. . . in case simulation rainfall occurred. . .” – this makes it sound like data was excluded as a preventative measure. In reality (?), you removed certain times/dates in post-processing where rainfall was present in the model solution, so I suggest this sentence should read something like: “Model data was excluded from analysis for all times where simulated rainfall was found to be present . . . .”

R: The suggested text has been integrated into Section 3(e) (L344-353)

20. L 296-297: Some confusion here due to wording – please re-write. To what are the results in Table 3 are similar? Of which ‘conclusion’ are you speaking?

R: The text has been amended. (L349-359).

21. L 341: ‘focus’ - > ‘focuses’

R: The text has been amended as suggested. (L341).

22. L 350-351: How many points contribute to this average?

R: For better clarity, the text has been amended. (L369-370).

23. L 426: “exchange in the WRF-UCM2D simulation. . .”

R: The text has been amended as suggested. (L445).

Response to reviewer #2

1) the first point regards the simulated temperature the authors used for comparison. The authors did not mention in any part of the paper how they calculated T(2m) and T(10m). What is the elevation of the first grid level? and, what is the average building eight in each grid cell in correspondence of the urban canopy? As the authors well know, one of the main problems in testing urban canopies simulations is the choice of the height used for the comparison. Since the first grid node of WRF is, I presume, well above the canopy, the authors would have used some similarity laws to calculate the simulated temperatures at 2m and 10m. Please clarify. 2) Over the last few years, approaches alternative to the UCM scheme have been developed to simulate urban heat island effect. In spite of this, the authors did not mention any other urban canopy layer scheme. I suggest to add in the Introduction a brief discussion on that issue.

R: (1) In this study, only simulated air temperature at 2-m elevation is employed to compare with the observation **(L233)**. The WRF thermodynamic variables are not directly simulated at 2 m, but diagnosed from values at the land surface and the lowest model layer (Hu et al. 2010).
WRF-UCM further considers the forcing from artificial surface (building, roof, wall and road). Building canopy is assumed to resemble aerodynamic roughness, thus implying that the complex urban canopy layer is replaced by a roughness number rendering information regarding the quantities within the canopy layer unavailable (Mirzaei and Haghighat, 2010). Characteristics within the roughness sub-layer vary with a horizontal distance scale determined by inter-element spacing, rather than height and vertical temperature gradient (Arnfield, 2003). In this study, the building height, i.e., the canopy layer, is 6 m**.** The elevation of the lowest model layer is 25 m in this study. The "2-m diagnostic air temperature" is calculated using the energy budget equation in the urban copy model expressed as equation (1) (**Line 388**). The Fsh from Noah LSM is averaged with the sensible heat flux from UCM weighted by the urban fraction to yield the representative Fsh (Chen et al. 2011) **(L382-393)**. The representative skin temperature (Tsk) of a model grid is averaged with the Tsk from Noah LSM and the skin temperature of the artificial surfaces (i.e., roof, wall and road) in the UCM weighted by their respective coverage (Chen et al. 2011). A schematic of the single-layer UCM from Chen et al. (2011) is shown below (Figure S1).

(2) Thanks for the suggestions. To improve the urban air temperature simulation, WRF (V3.2) has been integrated including (a) A bulk urban parameterization in Noah to represent zero-order effects of urban surface (Liu et al. 2006) (b) Single-layer urban

canopy model by Kusaka et al. (2001) and Kusaka and Kimura (2004) (c) multi-layer urban canopy (BEP) and indoor-outdoor exchange (BEM) models by Martilli et al. (2002). Chen et al. (2011) had reviewed the integration of Weather Research and Forecasting (WRF) model with different urban canopy schemes. (A schematic figure is shown below). We cited in the introduction in this revision **(L75-80)**.

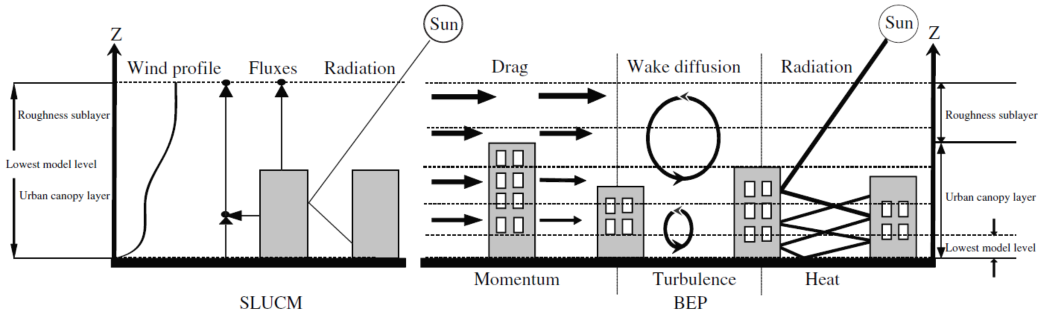

Figure 2. A schematic of the SLUCM (on the left-hand side) and the multi-layer BEP models (on the right-hand side).

(Figure S1 Taken from Chen et al. 2011)

References:

Arnfield, A.J., 2003: 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.

Hu X.M., J. W. Nielsen-Gammon, F. Zhang 2010: Evaluation of three planetary boundary layer schemes in the WRF model. *Journal of Applied Meteorology and Climatology*, 49,1831-1844.

Mirzaei P. A., and F. Haghighat 2010: Approaches to study urban heat island-Abilites and limitations, *Building and Environment.* 45,2192-2201.