RC1.

General comments

Quantifying outdoor heat stress of urban inhabitants requires consideration of air temperature and humidity, mean radiant temperature (MRT) and wind speed. As noted in the abstract, temperature and relative humidity vary on scales of hundreds of meters but wind speed and MRT vary at the meter scale. As a result, current simulation workflows require use of a mesoscale simulation to provide boundary conditions for time-consuming microscale computational fluid dynamics (CFD) simulations that, in the absence of supercomputer resources, apply only to a specific neighborhood for a short period of time. The authors argue for an alternative approach that relies on an estimation of MRT at several locations in streets of varying orientation, microscale CFD simulations to determine mean wind speed and its variation across a range of urban morphologies, and mesoscale simulation of temperature and relative humidity to estimate thermal-stress indicators and their variation at the sub-grid level of the mesoscale simulation.

While the method is more efficient than applying microscale simulations at large temporal and spatial scales, it is more complex than simple estimates of air temperature and humidity, which existing mesoscale simulations can efficiently provide over long time scales, and more demanding than analysis of remotely estimated surface temperatures. Here the authors argue that heat stress depends on more than air temperature and humidity and that surface temperatures estimated from satellite measurements do not correlate with heat maps generated with multiple input variables.

The proposed approach has merit in meeting the needs of urban planners and designers to assess heat stress at fine spatial scale and expansive time scales on the order of years. As developed by the authors, the workflow has the rigor necessary to produce useful results.

The backbone of the workflow is the BEP-BEM multilayer urban canopy parameterization, developed by one of the authors, that includes a building energy model and is incorporated into the widely used mesoscale Weather Research and Forecasting (WRF) model. The authors parameterize MRT at locations within the geometry and orientation allowed in BEP-BEM. The

parameterization of wind speed is based on a large number of CFD simulations (both Reynolds Averaged Navier-Stokes and Large Eddy Simulations) that span realistic and idealized urban configurations, with wind speed parameterized by the building roof area or, more satisfactorily, vertical wall area, each normalized by the total horizontal area in a grid cell.

From an assumed variation of 1 ^{o}C in air temperature and computed variations in wind speed and MRT, the authors calculated 54 combinations from which they calculated two measures of heat stress and documented the 10th, 50th and 90th percentile values.

The paper effectively communicates the application of the method for a typical heatwave day in Madrid, Spain, with diurnal plots of MRT, air temperature and UTCI and Heat Index for high- and low-density neighborhoods and bar charts of thermal stress categories as estimated by UTCI and Heat Index. The conclusion is a crisp summary of the method and results, establishing the contributions of the research. In all, the paper is coherent and effectively claims a significant addition to the modeling of urban outdoor stress.

Authors: We thank the reviewer for their comments. Below are our answers to the specific comments

Specific comments

1)BEP-BEM is limited to two street orientations, each with the same street width and building height distribution. This limitation defines the considered variation in MRT, which is computed for three positions (sidewalks on opposite sides of the street and street center) for each of the two street orientations. BEP view factors and shading algorithms are used to estimate shortwave reflection and longwave emission and reflection. MRT accounts for shortwave and longwave radiation reaching a pedestrian, weighting radiation received from body surfaces at different orientations. Not stated in the paper is the calculation of surface temperatures, which depends on a heat balance in which absorbed radiation can be emitted as longwave radiation or conducted into building material. Model results are validated by comparison with more detailed simulations made with the measurement-validated TUF-Pedestrian, but only over the designated locations in a street canyon. Modestly more explanation of the physics in the reference model would have better bolstered the asserted confidence in the streamlined methodology developed in the paper.

Authors: An appendix has been added with more details about how the different radiation components used to estimate the mean radiant temperature are computed.

2) Variations over a wider range of geometries derived from the detailed simulation would have determined whether the range of MRT as constrained by the BEP geometry is a reasonable approximation.

*Authors: We agree with reviewer's comment, but we think that this limitation is intrinsically linked with the idealization done by the urban canopy paramterization (UCP) of the urban morphology. Giving that changing the idealized urban morphology used by the UCP is beyond the scope of the article, we consider that the approach presented is able to characterize – at least – the coolest and hottest spots in the grid cell, with reasonable accuracy. This point has been added in the **Limitations** section (4).*

3) Spatial maps at two times of the simulated day make a strong case that air and surface temperatures do not accurately predict UTCI values in hot-spot locations. The display of Heat Index is puzzling, because the paper promised the use of SET and UTCI but does not present simulated values of SET.

Authors: The Heat Index was introduced as an example of an index that can be derived from standard mesoscale model outputs and accounts for moisture in addition to air temperature, but does not take into account MRT. We clarified this in the text.

4) What's missing is a thoughtful discussion of limitations and, perhaps, a more detailed comparison with appropriate ground-truth simulations (with CFD and energy balances) for a single neighborhood. Do the authors think that the MRT model is adequate for all possible building materials and morphologies? What about the impact of trees? Similarly, is the parameterization of wind speed universally applicable or would city planners need to conduct or commission local simulations?

Authors: We thank the reviewer for this suggestion. A discussion of limitations has been added as Sect. 4.

Editorial comments

Line 47. Please replace 1 with one.

Authors: done.

Line 65 uses “autonomy” to characterize inhabitant choice of thermal environment, a different use of the word than in “spatial autonomy,” which refers to the extent, spatial or temporal” a space is thermally comfortable. People have agency to make choices, but spaces do not.

Authors: Please review Nazarian et al. 2019 referenced here for detailed descriptions of the term outdoor thermal comfort autonomy (OTCA) and the spatial and temporal extent of that used in the literature.

Line 71. Please hyphenate “grid average” (proper when followed by a noun).

Authors: done

Line 101. The sky is also a source of shortwave radiation, in which a portion of direct radiation from the Sun is scattered in the atmosphere.

Authors: the source of shortwave radiation in the sky is the sun – in any case we decided to speak only of short and longwave radiation in the sentence that now reads:

For shortwave reflection and longwave emission and reflection, the standard BEP view factor and shading routines (Martilli et al. 2002) are used to estimate the amount of shortwave (direct and diffuse) and longwave radiation reaching a vertical segment 1.80 m tall and located in each of the six positions previously mentioned (Fig. 1, Appendix A).

Line 119. In Equation 1, aK and aL are not defined.

Authors: These are the absorptivity of the pedestrian in the shortwave and longwave, respectively. These definitions have been added in the text.

The text asserts excellent agreement between two models for shortwave radiation loading but the figures in Appendix A show peak differences as much as 100 W/m2.

Authors: considering the strong idealization of the urban morphology used by BEP-BEM we believe that these differences are acceptable. We change the sentence as follows:

“A comparison of the shortwave radiation loading on the pedestrian between the two models reveals very good agreement (Appendix B Fig. B1, B2), considering the highly simplified urban morphology used by BEP-BEM, with biggest errors limited to short periods of time”

Line 147. Please replace “module” with “modulus.”

Authors: done

Line 154. Please consider “Data are considered from over 173 microscale CFD simulations of urban airflow over realistic and idealized urban configurations,…”

Authors: done

Line 177. Please consider deleting the commas that bracket “therefore.”

Authors: done

Line 185. For consistency with other choices of tense, please replace “used” with “use.”

Authors: done

Line 191. Please replace “Where” with “where.”

Authors: done

Line 202. Please consider “that increasing cause severe heat stress….”

Authors: done

Line 203. Please separate “27,9,3” with appropriate spaces.

Authors: done

Line 228. Grid cell or grid point would appear to be better than grid, to describe a specific location.

Authors: done

Lines 231, 232, 244 and 247 use heat stress index, Heat Index, Heat index and heat index; please consider more consistency.

Authors: changed to Heat Index.

Line 263. The time here is stated as 9 UTC while in line 281 it is 09000 UTC. For consistency with the afternoon time of 1600 UTC and the caption to Figure 9, both should be 0900. Please consider starting the sentence with “In the dense region,….”

Authors: done

Line 270. Please consider “…completely different pattern; on the city center at that time of day,….”

Authors: done

Line 309. Please consider “…this has not been done before….”

Authors: done

Line 311. Please delete the comma after “case.”

Authors: done

Line 454. The caption for Figure A1 should define Short 1 and Short 2. In the caption for Figure A2, please replace “a E-W…” with “an E-W….”

Authors: done

Reviewer 2:

This manuscript describes how to calculate outdoor heat stress across a city using the atmospheric mesoscale model, WRF. This manuscript deals with spatial variabilities in mean radiant temperature (MRT) and wind speed in urban canyons in a decisive manner by simplifying their many aspects and assumptions. Originally such spatial variations cannot be resolved in the mesoscale model. The ideas proposed in this manuscript are interesting but simplified too much. Further investigation of this study needs important things below.

Authors: We thank the reviewer for their comments. Below are our answers:

1) The key idea of this study comes from the six-directional weighting method by (Thorsson et al., 2007). But this reference is missing in the manuscript and the description of this method is quite descriptive. This journal is for the code and model, and we expect more detailed information and code description.

Authors: We added the references relevant to the six-directional weighting method. We also added an appendix with more details about the calculation of the radiation components of the Mean Radiant Temperature.

2) The similar problem also goes to wind speed calculation in 2.2.

Authors: The parameterization proposed in section 2.2 is based on a fitting of the data calculated from a large set of CFD simulations, as described in some detail in Sect. 2.2. The data of the CFD simulations are about to be made public, and are the subject of another paper entitled "UrbanTales: A comprehensive dataset of Urban Turbulent Airflow using systematic Large Eddy Simulations".

Additionally, some information is vague. For example, what is the meaning of "close to the pedestrian height (~2.5 m)"? The symbol "~" stands for the approximation and why we need this approximation? So wind speed is at 2.5 m above the road?

Authors: We use the symbol for "approximate" 2.5m above ground, because WRF uses sigma pressure terrain following coordinates, and therefore the thickness of the lowest model level slightly changes with the position (e. g. topography) and time. In the simulations presented the thickness of the lowest model level at sea level and standard pressure is 5m, and – since WRF uses an Arakawa C grid – the wind components are defined at 2.5m.

1) What are the implications and limitations to use spatially averaged wind speed with estimation of MRT at three different locations? We need more considerate discussion on many assumptions and parameter values used in this study.

   Authors: a discussion on the limitations has been added (Section 4).

2) Figure 2 for the model evaluation may not be useful because the proposed model is based on the two-street orientation. We can also argue that parameters in the proposed model is calibrated in some sense to match the results. It will be quite useful if there is comparison between the model and in-situ data in a city.

   Authors: The aim of the comparison shown in Fig. 2 is more verification (check if the model is working properly), rather than validation (check if the model represents the reality). This is why it is important that both BEP-BEM and TUF-3D

have the same urban morphology. The comparison shows that – for the same urban morphology – BEP-BEM gives very similar MRT values to TUF-3D. The limitations of the idealized morphology and the two street orientations are now discussed in section 4.
It is important to mention, also, that TUF-3D has been validated with measurements (Lachapelle et al 2022; Jiang et al. 2023) of MRT for neighborhoods in Phoenix, AZ and Guelph, ON. However, those measurements are at the microscale and cannot be used for validation of the BEP-BEM approach presented in the paper, which aims to assess the variability of MRT within the grid cell of a mesoscale model (order of one km^2). We are not aware of suitable experimental datasets to evaluate the MRT calculated by WRF-Comfort at the neighbourhood scale (i.e., spatial max and min MRT, for example).
In summary, Fig. 2 shows that 1) once the morphology is fixed, the computation of MRT done in BEP-BEM is correct, and 2) with all the limitations linked to the idealized morphology mentioned in section 4, we can expect reasonable values of MRT from BEP-BEM.

3) Please check carefully if description on variables, abbreviation, and indices are well described in the manuscript.

   Authors: done