Reviewer 2

General statement

The paper "Evaluating the radiative fidelity of PALM (v25.04) in high-resolution: impact of diverse urban morphology and vegetation on short-wave radiation" by J. Radović et al. evaluates 3D urban shortwave radiative transfer model of the PALM model system against pyranometer measurements from four sites near Dejvice, Czech Republic. PALM is run in a spinup mode, without 3D atmospheric model. The evaluation is performed over 16 cases (episodes), with downwelling radiation derived from two different WRF setups. The evaluation shows that PALM is able to capture the SW radiative transfer within urban canopy well, but is limited by the accuracy of the prescribed irradiances from WRF as well as the quality of input data prescribing the urban form.

The scope of the work is well defined and fits within the scope of GMD. The main novelty of the work is the comparison of the PALM-modelled radiative fluxes against real-world pyranometer measurements, an useful addition to prior evaluations of PALM's representation of urban canopies. The evaluation is methodologically sound, and highlights both the strengths and limitations of PALM's urban representation adequately. The quality of presentation is generally good, although there is still room for improvement. The paper reports useful findings for researchers working with 3D-resolved urban canopy simulations.

I have some general and specific critical comments as well as suggestions listed below, but I do not think addressing these would require major revisions to the manuscript or any substantial new work. Therefore, I recommend the manuscript to be published in GMD subject to a minor revision that adequately addresses these comments:

Main comments

1. **I would suggest a small rewrite of the Introduction section so that the theoretical and practical background for the work would be clearer. Currently, the emphasis is too much on different pre-existing models given as examples rather than in fundamental modelling approaches. This gives a lot of focus on the models itself, although they are not used in the study. I would suggest turning this the other way around: introduce the different modelling approaches in general, and just shortly list examples of models implementing the approach.**

Answer: *Thank you for pointing out this suggestion. We have rewritten the text throughout the Introduction section to prioritise the modelling methodologies over the specific model names. Please review the "Introduction" section for the changes we implemented.*

2. **I feel that a comprehensive description of the complete evaluation strategy is missing. Instead, the descriptions of the various evaluations and comparisons performed, as well as the reasoning behind them is scattered along the Results section. I suggest adding one as a new subsection to the Methods section, moving all information describing the evaluation (what was done and why) from Results in there. This way, after reading the Methods section, the reader would already have an understanding of how the evaluation was performed and why so, and the Results section could be dedicated purely for reporting the results. Currently, the reader needs to pick these pieces of information while reading through Results.**

Answer: *We thank the reviewer for this constructive feedback. We fully agree that centralising the evaluation strategy enhances the narrative flow and clarity of the manuscript. Following your suggestion, we have restructured this portion of the manuscript. We expanded Section 2.6, "PALM output and observation data processing," to provide more detailed information on data processing.*

3. **I think it would be important to compare whether the averaging time scale has influence on evaluation metrics. Especially during non-clear-sky episodes, the point evaluation with a relatively short temporal averaging can be very sensitive to timing and positioning of single clouds, even if the average radiation over multiple hours (or large spatial area) would be close to truth. In addition to the current hour-by-hour pairwise comparison, I would suggest comparing at least the integrals of daily SW radiation pairwise from the model and the measurements for each of the episodes (and episodes together). The dependency of evaluation metrics on selected averaging time scale could be studied further as well (from 10 min to daily), if authors consider it viable.**

Answer: *We thank the reviewer for this suggestion. We conducted an additional analysis based on daily integrated shortwave radiation (incoming and outgoing) and compared the PALM modelled shortwave radiation with the observed shortwave radiation across all possible combinations (stations, episodes, and all stations and episodes together). The results obtained were consistent with the conclusions drawn from the original hourly evaluation presented in the manuscript. In particular, this new set of results confirmed the model behaviour identified in the hourly comparison, without revealing additional discrepancies or compensating effects at longer averaging timescales. Namely, the model configurations that performed better, i.e., PALM-FU, in the hourly assessment also showed improved performance in the daily integrated quantities, which is physically expected; the same is true for stations, e.g., the HAN station, which showed a strong underestimation. During non-clear-sky conditions, hourly comparisons over short intervals are, of course, influenced by the precise timing and positioning of individual clouds in relation to the measurement site, and when the radiation is averaged over longer periods, such as daily totals, positive and negative deviations compensate for each other, which leads to smaller differences between PALM modelled and observed values. For this reason, this additional analysis did not provide any new or significant information beyond confirming the robustness of the conclusions that we already derived from the hourly analysis. Hence, we also* *expect that extending the analysis to shorter time-averaging intervals, e.g., 10-minute averages, would increase 'noise' in the evaluation metrics rather than yield quantitatively different conclusions about model performance. Therefore, we decided to maintain the primary focus on the hourly evaluation, which more accurately reflects the model's ability to reproduce the temporal variability of surface radiation during the studied episodes. In addition, correlation coefficients were not produced for some subset of analysis because the number of available samples for evaluation was too small for a statistically meaningful correlation analysis. Therefore, for these subsets, evaluation focuses on RMSE and MBE, which are more interpretable. (The integrated shortwave radiation quantities are expressed in MJ m⁻² for consistency with common conventions used in radiation studies).*

1. ***CLEAR SKY:***
   a) ***Statistical evaluation metrics: Integrals of daily SWin and SWout across FU-clear-sky episodes***

Variable	RMSE [MJ/m²]	MBE [MJ/m²]	Correlation
SWin	1.023	-0.032	0.978
SWout	1.683	-1.147	0.768

   b) ***Statistical evaluation metrics: Integrals of daily SWin and SWout for each episode for FU-clear-sky***

Episode (date)	Variable	RMSE [MJ/m²]	MBE [MJ/m²]
e1 (19-20/06/2017)	SWin	0.631	-0.500
e1 (19-20/06/2017)	SWout	0.521	-0.509
e2 (19/07/2017)	SWin	0.761	-0.056
e2 (19/07/2017)	SWout	0.630	-0.215
e3 (7/8/2017)	SWin	1.146	0.618
e3 (7/8/2017)	SWout	0.623	-0.433
e4 (30/06/2018)	SWin	1.536	-1.090
e4 (30/06/2018)	SWout	2.654	-2.168
e5 (02-03/07/2018)	SWin	1.148	-0.489
e5 (02-03/07/2018)	SWout	2.497	-2.021
e6 (11-12/09/2018)	SWin	0.737	-0.364

e6 (11-12/09/2018)	SWout	1.222	-0.962
e7 (9/6/2017)	SWin	1.328	1.213
e7 (9/6/2017)	SWout	0.226	-0.061
e8 (11/6/201)	SWin	0.383	-0.210
e8 (11/6/201)	SWout	0.309	-0.282
e11 (7/7/2018)	SWin	0.872	-0.020
e11 (7/7/2018)	SWout	2.336	-1.868
e13 (24/07/2018)	SWin	0.935	-0.110
e13 (24/07/2018)	SWout	2.168	-1.769
e14 (1/8/2018)	SWin	1.058	0.502
e14 (1/8/2018)	SWout	2.284	-1.904
e15 (17/08/2018)	SWin	1.332	0.804
e15 (17/08/2018)	SWout	1.700	-1.293
e16 (20/08/2018)	SWin	1.240	0.538
e16 (20/08/2018)	SWout	1.741	-1.373

c) ***Statistical evaluation metrics: Integrals of daily SWin and SWout across stations for FU-clear-sky:***

Station	**Variable**	**RMSE [MJ/m^2]**	**MBE [MJ/m^2]**	**Correlation**
FLE	SWin	0.281	0.401	0.948
FLE	SWout	0.393	-0.199	0.816
FSV	SWin	1.170	-0.013	0.995
FSV	SWout	0.465	-0.433	0.989
HAN	SWin	1.030	0.796	0.992
HAN	SWout	2.923	-2.835	0.979
NTK	SWin	0.92	-0.213	0.906

NTK	SWout	0.580	-0.471	0.727

2. NON-CLEAR SKY

a) Statistical evaluation metrics: Integrals of daily SWin and SWout across FU-non-clear-sky episodes:

Variable	RMSE [MJ/m^2]	MBE[MJ/m^2]	Correlation
SWin	3.384	2.891	0.862
SWout	1.535	-0.744	0.430

b) Statistical evaluation metrics: Integrals of daily SWin and SWout for each episode for FU-non-clear-sky:

Episode (date)	Variable	RMSE [MJ/m^2]	MBE [MJ/m^2]
e9 (7/7/2017)	SWin	4.712	3.988
e9 (7/7/2017)	SWout	0.672	0.379
e10 (16/06/2018)	SWin	1.915	1.803
e10 (16/06/2018)	SWout	1.964	-1.417
e12 (21/7/2018)	SWin	2.913	2.881
e12 (21/7/2018)	SWout	1.662	-1.197

c) Statistical evaluation metrics: Integrals of daily SWin and SWout across stations for FU-non-clear-sky:

Station	Variable	RMSE [MJ/m^2]	MBE [MJ/m^2]
FLE	SWin	1.478	1.478
FLE	SWout	0.174	-0.174
FSV	SWin	1.919	1.806
FSV	SWout	0.050	-0.061
HAN	SWin	2.910	2.878
HAN	SWout	2.572	-2.564
NTK	SWin	6.498	6.498

NTK	SWout	0.934	0.934

3. COMMON EPISODES

a) Statistical evaluation metrics: Integrals of daily SWin and SWout across common episodes

PALM config.	Variable	RMSE [MJ/m²]	MBE [MJ/m²]	Correlation
PALM - CNU	SWin	3.077	-2.768	0.985
PALM - CNU	SWout	1.809	-1.395	0.806
PALM - FU	SWin	0.987	-0.340	0.988
PALM - FU	SWout	1.627	-1.089	0.800

b) Statistical evaluation metrics: Integrals of daily SWin and SWout for each episode for common episodes

PALM config.	Episode (date)	Variable	RMSE [MJ/m²]	MBE [MJ/m²]
PALM - CNU	e1 (19-20/06/2017)	SWin	3.761	-3.514
PALM - CNU	e1 (19-20/06/2017)	SWout	1.157	-1.127
PALM - FU	e1 (19-20/06/2017)	SWin	0.630	-0.500
PALM - FU	e1 (19-20/06/2017)	SWout	0.521	-0.509
PALM - CNU	e2 (19/07/2017)	SWin	2.964	-2.603
PALM - CNU	e2 (19/07/2017)	SWout	1.078	-0.716
PALM - FU	e2 (19/07/2017)	SWin	0.761	-0.056
PALM - FU	e2 (19/07/2017)	SWout	0.630	-0.215
PALM - CNU	e3 (7/8/2017)	SWin	2.216	-1.595
PALM - CNU	e3 (7/8/2017)	SWout	1.051	-0.843
PALM - FU	e3 (7/8/2017)	SWin	1.146	0.618
PALM - FU	e3 (7/8/2017)	SWout	0.623	-0.433
PALM - CNU	e4 (30/06/2018)	SWin	4.006	-3.957

PALM - CNU	e4 (30/06/2018)	SWout	2.853	-2.412
PALM - FU	e4 (30/06/2018)	SWin	1.536	-1.090
PALM - FU	e4 (30/06/2018)	SWout	2.654	-2.168
PALM - CNU	e5 (02-03/07/2018)	SWin	3.063	-2.999
PALM - CNU	e5 (02-03/07/2018)	SWout	2.641	-2.217
PALM - FU	e5 (02-03/07/2018)	SWin	1.148	-0.489
PALM - FU	e5 (02-03/07/2018)	SWout	2.497	-2.021
PALM - CNU	e6 (11-12/09/2018)	SWin	2.054	-1.874
PALM - CNU	e6 (11-12/09/2018)	SWout	1.099	-0.946
PALM - FU	e6 (11-12/09/2018)	SWin	0.737	-0.364
PALM - FU	e6 (11-12/09/2018)	SWout	1.222	-0.962

c) ***Statistical evaluation metrics: Integrals of daily SWin and SWout across stations for common episodes:***

Station	**PALM config.**	**Variable**	**RMSE [MJ/m²]**	**MBE [MJ/m²]**
FLE	PALM - CNU	SWin	1.694	-1.432
FLE	PALM - CNU	SWout	0.643	-0.479
FLE	PALM - FU	SWin	0.946	0.406
FLE	PALM - FU	SWout	0.454	-0.190
FSV	PALM - CNU	SWin	3.464	-3.341
FSV	PALM - CNU	SWout	0.689	-0.657
FSV	PALM - FU	SWin	1.498	-1.386
FSV	PALM - FU	SWout	0.477	-0.444
HAN	PALM - CNU	SWin	2.303	-2.135
HAN	PALM - CNU	SWout	3.054	-2.383
HAN	PALM - FU	SWin	0.435	0.268

HAN	PALM - FU	SWout	2.962	-2.810
NTK	PALM - CNU	SWin	4.257	-4.181
NTK	PALM - CNU	SWout	1.434	-1.428
NTK	PALM - FU	SWin	0.669	-0.625
NTK	PALM - FU	SWout	0.676	-0.643

Specific comments & suggestions

1. **The wording in the abstract could be a bit more careful:**
    1. **L11-12: “a capability that mesoscale models cannot match”**
    I think the statement is too general. Mesoscale models can implement coupling to a 3D urban surface model which could match PALM’s capabilities in this regard, one example would be 3DUCM and CSUMM (Conigliaro et al., 2021). This could be possible with WRF-SUEWS as well, using the SPARTACUS-Surface for 3D radiative interactions, however I’m not sure if this is tested in practice. There could be some other examples as well. Nevertheless, my point here is that this statement is not necessarily valid in general.

Answer: *We thank the reviewer for this comment. After considering reframing the sentence or simply replacing the phrase with alternative text to clarify the distinction, we decided that any replacement would unnecessarily overcomplicate the text. We think the text is more precise if we remove this clause entirely, as it avoids an inaccurate generalisation about mesoscale capabilities. Please take a look at the updated text at Lines 10-12, which now refers to: “Results demonstrate that PALM resolves urban- and vegetation-induced short-wave radiative exchange (i.e., canyon trapping, vegetation shading, building reflections, interaction with urban surfaces and dynamic timing) with high fidelity regardless of the urban setting.”*

2. **L12: “PALM’s superiority”**
PALM’s superiority is context-dependent, and while in the present study the model performs very well on capturing the SW exchanges in the urban canopy, this statement seems too general.

Answer: *We have revised the wording in the abstract at L12 to rephrase subjective terms such as 'superiority.' Please review the updated text on Line 14, which now refers to the 'explicit physical detail' of the micro-scale approach.*

2. **L1: “Validating short-wave …” → “Validating urban short-wave …”**

Answer: *We agree with the reviewer's suggestion. The change is implemented in the revised manuscript (please see Line 1).*

3. **L126: "extensible" → "modular" would perhaps be more fitting here.**

Answer: *Accepted. The change is implemented in the revised manuscript (please see Line 127).*

4. **L127-129: The spin-up mode should be explained in detail, as this is a key feature of the modelling setup. E.g. what processes are included and what excluded from the computation, how does the solved model system look like, what are remaining factors affecting the SW radiation and what are fixed constants. Given the context of the study, it would be especially important to know whether the albedo of the surfaces can change throughout the simulation (and how so).**

Answer: *We agree with the reviewer's suggestion. We have included detailed information that addresses the requested questions in the separate subsection 2.1.1 titled "Spin-up simulations". Additionally, to complete the missing descriptions, the changes were implemented in 2.1 "PALM model configuration " and 2.1.2 "RTM configuration" sections.*

5. **L134-135: "by external forcing" → "by prescribed external forcing" to be more specific.**

Answer: *Done. The change is implemented in the revised manuscript (please see Line 134).*

6. **L149-150: Specify how the data was interpolated to radiation model time steps and what time step was used to compute the radiation interactions.**

Answer: *Thank you for this comment. We agree that the original description did not provide sufficient detail on this aspect of the methodology. To address your concerns, we have expanded Section 2.1.2, " RTM configuration," in the revised manuscript to provide a more complete description of the simulation process.*

7. **The angular resolution used with RTM ray-tracing is reported, but not the spatial resolution of the PALM surface representation. I think the authors should add a summary of the PALM model setups (e.g. resolution, number of grid points, time step, integration scheme, and any other information that may be important for reproducibility).**

Answer: *We thank the reviewer for this suggestion. A similar point regarding the model setup details was also raised by Reviewer #1. In response to both comments, we have added a dedicated paragraph clarifying the technical specifications and configuration details. Please see the section 2.1.1 titled "Spin-up simulations" in the revised manuscript.*

8. **L174-178: Perhaps some information on solar elevation could be added, e.g., the range of maximum solar elevation over the episodes.**

Answer: *We agree that adding information on solar elevation provides valuable context for the radiative exchange within the urban canyons. Based on Prague's coordinates and the specific simulation dates, we have calculated the maximum solar elevations for all episodes. The values range from 44.41° (mid-September) to 63.37° (summer). We have updated Section 2.5 "Simulation episodes and meteorological conditions," and Table 4 to provide additional description and to include these values for each selected episode.*

9. **Table 2: Measurement heights would be needed here. The authors could also report the view factors (VFs) for surface types for both incoming and outgoing SW radiation (e.g. FLE incoming: 0.xx sky, 0.xx building walls 0.xx tree canopy, …; outgoing: 0.xx road, 0.xx low vegetation, 0.xx …), as computed by PALM. This would help comparing the results across sites.**

Answer: *The measurement heights are provided in section 2.1.1 (see lines 171-172). In addition, we have revised Table 2 to include the sky, obstacle, and tree view factors for all measurement locations.*

10. **The manufacturer as well as the manufacturer's country of origin should be given for the instruments.**

Answer: *We have further updated the sensor/instrument information in Table 3 to explicitly include both the manufacturer's name and the manufacturer's country of origin for all listed instruments. Please, see Table 3.*

11. **Table 3: The given CMP3 accuracy for incoming SW seems to be unrealistically high for its resolution. Please recheck the numbers for all instruments from the official data sheets.**

Answer*: We have rechecked the official manufacturer data sheets for all instruments listed in Table 3. The accuracy and resolution values for CMP3, as well as those of the other sensors, have been corrected, and the updated values have been integrated into Table 3 of the revised manuscript.*

12. **Table 5: Instead of absolute and relative differences, perhaps report bias (with the sign) and the relative bias, as the sign is important here.**

Answer: *We appreciate the reviewer's suggestion to report bias. Regarding Table 5, we prefer to maintain the 'Absolute and Relative Differences' format as it provides a clear, station-by-station comparison of the magnitude of the deviations.*

To address the reviewer's valid point about the importance of the sign of the error, Figure 2 provides a comprehensive summary of the mean bias across all categories. We believe that presenting the data in this way—the detailed differences in the table and the bias in the figure—offers the most complete picture of model performance without introducing redundant metrics into the table.

13. **L313: “bottlenecks” → “degradation” or similar, I think the audience of GMD would associate “performance bottlenecks” solely with computational bottlenecks.**

Answer: *Accepted. The change is implemented in the revised manuscript (please see Line 402).*

14. **L316-319: As discussed earlier, this is not always true for all mesoscale models. But definitely an argument for resolving 3D radiation. This is also an example of text in the Results section that would be better suited for Methods.**

Answer: *Thank you for this comment. We agree that the original wording was overly general and could be interpreted as applying to all meso-scale models. We have therefore revised the text to explicitly refer to the ability to resolve three-dimensional radiative processes and shading patterns associated with buildings and vegetation, which is a key advantage of high-resolution urban models such as PALM (Please see Lines 405-407).*
In addition, the Methods section has been substantially expanded in the revised manuscript as suggested by your second comment and this comment as well, and now contains a more detailed description of the modelling framework and the treatment of three-dimensional radiative processes. We have nevertheless retained brief explanatory statements in the Results section because we believe it helps introduce the physical mechanisms responsible for the observed radiation patterns before presenting the corresponding results. The retained text is intended only as a short reminder of the underlying process and to improve the readability and flow of the discussion.

15. **The font size in Figures 5-7 is really small, especially in Figure 7. Check that the texts are readable at true paper size.**

Answer: *As suggested, we have increased the font size in Figures 5-7.*

16. **The definitions of the evaluation metrics could be moved from supplementary material to the appendix section of the paper so that they would be more accessible for the reader.**

Answer: *We agree with the reviewer’s suggestion. The formulas used for statistical evaluation are moved from the supplementary material to the manuscript as Appendix A: Statistical evaluation (Lines 705-734).*

17. **L587-589: I would perhaps rephrase this a bit as the robustness of radiative transfer simulation is subjective. I would state that the quality and accuracy of the prescribed datasets and mesoscale input forcing data are clearly the dominant sources of errors, not the internal radiative transfer simulation.**

Answer: *We thank the reviewer for pointing this out and agree with the reviewer’s view that “robustness” is subjective in this context. The sentence is rephrased accordingly (please see Lines 691-693).*