Dear Dr. Ting,

Thank you very much for your quick and positive assessment of our revised manuscript and for your careful evaluation throughout the review process. We greatly appreciate your comments and are pleased that the revisions have satisfactorily addressed the referees' concerns.

Following your suggestion, we have revised the title to better reflect the study's scope and clarify that the evaluation concerns the WRF-driven radiative transfer component of PALM. The revised title now reads:

“Evaluating the radiative fidelity of WRF-driven PALM (v25.04) in high-resolution using RTM: impact of diverse urban morphology and vegetation on short-wave radiation”

We intentionally omitted the RTM version number from the title because RTM is an integral component of PALM. The RTM version is tied to the corresponding PALM release; therefore, PALM v25.04 inherently includes RTM v4.1. We felt that including both version numbers in the title would unnecessarily increase its complexity. The specific RTM version is, of course, stated in the manuscript.

For consistency and to ensure that all submission materials match, we have also updated the manuscript title in the **"Manuscript title"** field of the submission system.

If you feel that explicitly including the RTM version in the title would improve clarity, we would be happy to revise the title accordingly.

Thank you again for your constructive guidance and for handling our manuscript. We appreciate your time and consideration.

Kind regards,

Jelena Radović, on behalf of all co-authors

Reviewer 1

General

This is a comprehensive and timely GMD style evaluation paper. It benchmarks short wave radiation in the microscale model PALM version 25.04 across different urban and vegetated settings in Prague Dejvice, and it does so over a meaningful ensemble of episodes from clear sky to cloudy conditions. The takeaway lands well that PALM can reproduce street scale shading and reflection patterns very convincingly when the incoming radiation is right, but it cannot fix errors coming from the mesoscale cloud and radiation forcing. The results and figures support that message clearly. I generally find the paper well prepared and would recommend a minor to moderate revision focused on reproducibility details and a slightly tighter interpretation at the most problematic site.

Specific comments

1. **The paper would benefit from a compact run description inside the manuscript itself, not only in the archive. Right now a reader has to piece together key setup details. *I suggest to add a short block that states PALM horizontal and vertical grid, time step, episode length, spin up length, and output frequency, plus which modules were active and which radiation options were used.***

   It would also help to explicitly describe how station values were sampled from the model, including the sampling height, whether it is nearest grid cell, and how reflected short wave from the model is matched to the up looking and down looking pyranometers.

Answer: *In this study, the term 'spin-up' refers to the specific operational mode of the PALM model system used to perform the primary micro-scale simulations, rather than a separate pre-processing phase; for the purposes of this study, these 'spin-up' runs constitute the primary simulations. Hence, PALM simulations were performed with horizontal and vertical grid resolutions of 1 m, utilising a 5 s timestep during spin-up. Regarding the simulation duration (i.e., the spinup length), while the selected physical episodes covered 24 to 48 hours, the simulations for each day were initiated at 23:00 UTC (one hour prior to the target date) and lasted 23 or 46 hours, depending on the selected physical episode. A one-hour initialisation offset (starting at 23:00 UTC) was implemented to address technical constraints within the PALM/palm_meteo framework, with no loss of critical data, since short-wave radiation is zero at midnight. PALM output data were recorded at a frequency of 600 s (10 minutes).*
The modelled irradiance values for the pyranometers were taken at the grid cell nearest to the sensors' actual locations. In the vertical direction, this translates to 1 m above ground for both the upward- and downward-facing sensors. The short-wave irradiance of the sensors was modelled identically to that of actual surfaces in the RTM module of the PALM model, utilising fully 3-D radiative interactions with multiple reflections and shading by the modelled terrain and buildings, and partial attenuation by the fully resolved vegetation. More details about RTM are

available in Krč et al. (2021). The identified limitation was the agreement between the modelled representation and reality, particularly regarding tree crown shapes in the input data (see Section 4.1 and Figure 8, which compares the photographed reality and the modelled representation).
As noted by the reviewer, the active modules (LSM, BSM, RTM, and MESO) are detailed in Section 2.1 Line 138). The complete section 2.1 has been rewritten, and an additional paragraph, 2.1.1 "Spin-up simulations," has been added, in which the specific issues are elaborated in detail.

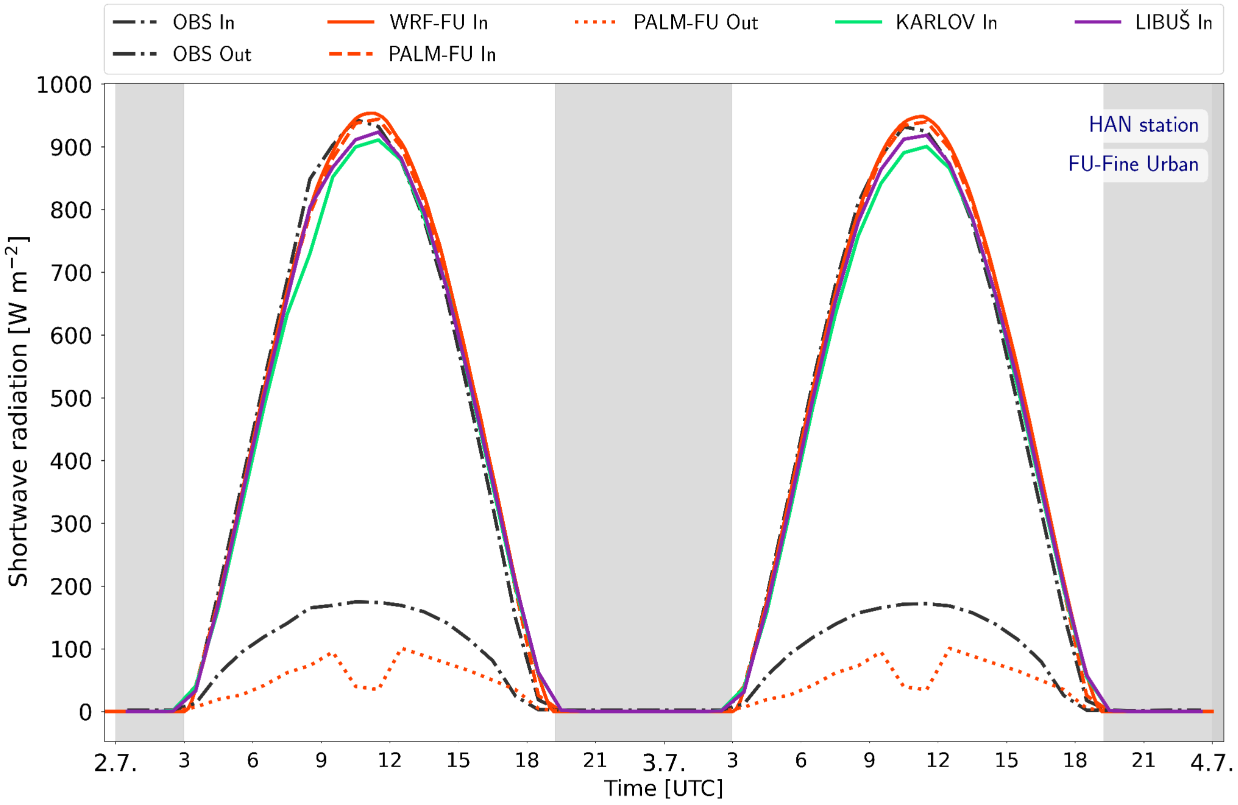
2. **The spin up approach is described as not affecting short wave radiation, but the interpretation needs one more sentence of care. Reflected short wave depends on surface albedo and on surface and canopy state. Please clarify what can change during spin up for the land and vegetation state, especially soil moisture and grass or vegetation parameters, and what is fixed. This matters directly for your explanation of the reflected radiation behavior at the vegetated site.**

Answer: *The reviewer makes a valid point regarding the potential influence of surface and canopy states on reflected short-wave radiation. In this study, however, the land and vegetation parameters—specifically albedo, leaf area density (LAD), and canopy height—are prescribed as static inputs via the static driver file and remain constant throughout the simulation. For a detailed specification of the static driver data included in this study (i.e., individual land-surface and plant-canopy input parameters), we refer to the validation study by Resler et al. (2020).*

*Parameters used in the Land Surface Model (LSM), i.e., **land_surface_parameters** such as **"root_fraction"**, "**soil_moisture"**, "**soil_temperature"**, and "**deep_soil_temperature"**, are fixed and prescribed in advance in the LSM model parameter list of the configuration p3d file. The same applies to the Plant Canopy Model (PCM) parameter list and "**plant_canopy_parameters"** in the p3d configuration file.*

During the spinup phase for the original set of simulations, the option "calc_soil_moisture_during_spinup " was switched off, which means that the additional prognostic equation for the volumetric moisture content of the soil layer is not solved.
To assess the effect of this configuration parameter, we performed an additional simulation with this option set to True, which enabled the soil moisture prognostic equation to be solved. However, its activation did not affect the simulated short-wave radiation at the HAN station. The results indicate that for episode e5, radiative conditions were largely insensitive to the soil moisture settings. Furthermore, the absence of differences suggests that the modified soil moisture treatment did not substantially alter the surface thermal regime, with the soil moisture remaining similar between the simulations, or the atmospheric forcing dominated the surface energy balance during the episode.
In addition, for the revised manuscript and changes included we refer to Section 2.1.1 "Spin-up simulations".

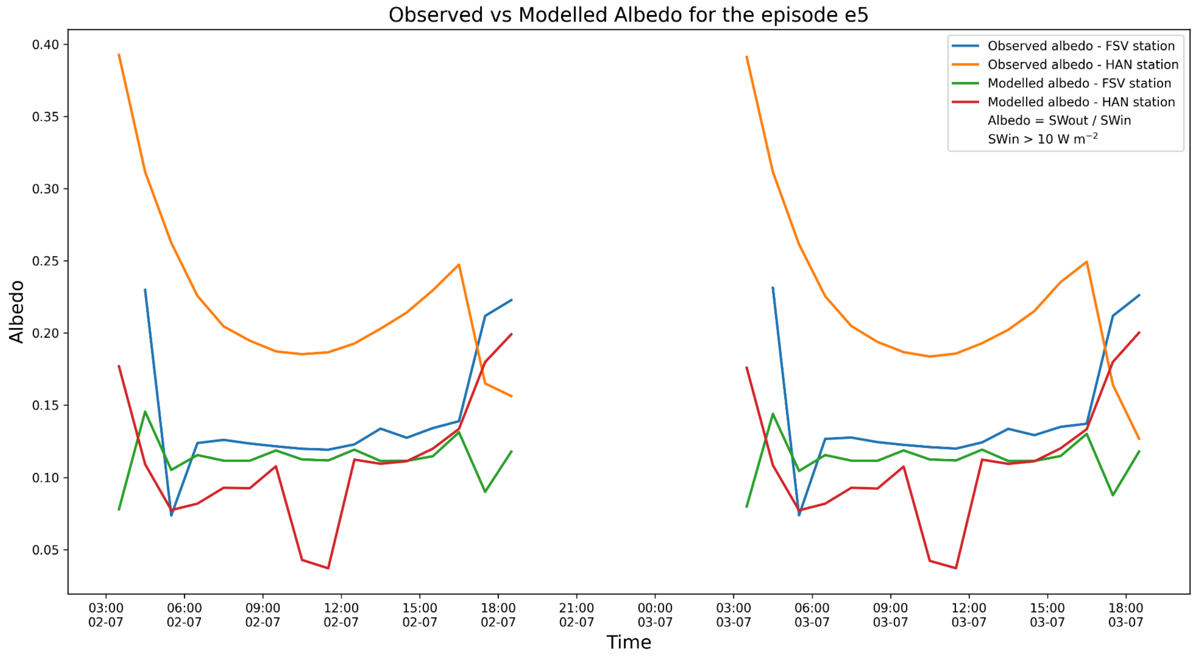
Figure: *The comparison of observed hourly averages of incoming (In) and outgoing (Out) short-wave radiation for the e5 episode at the HAN station, with PALM model output for WRF Fine Urban (WRF-FU) configuration with the soil moisture parameter switched ON. Additional lines represent the raw WRF output, Fine Urban (WRF-FU), and measurements from professional meteorological stations in Karlov and Libuš.*



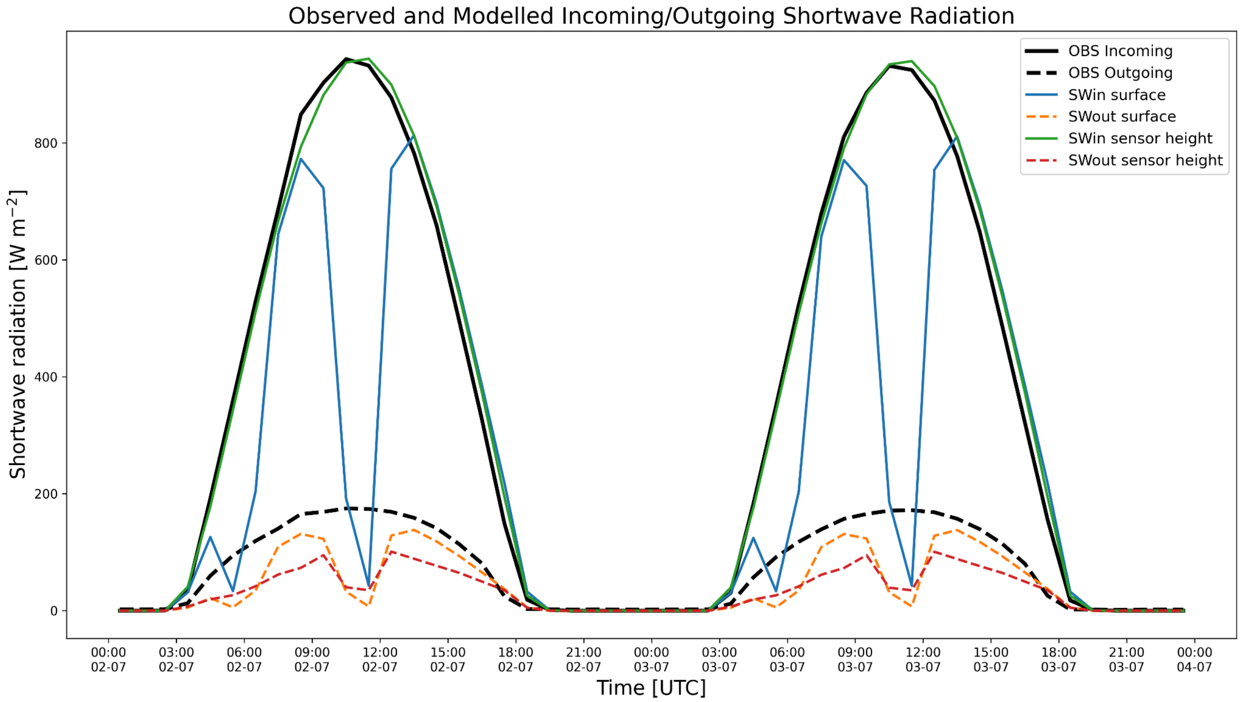
3. **The persistent reflected short wave bias at the HAN location is an important result and it is already well highlighted in the tables and discussion, but it still reads a bit like a likely story rather than a demonstrated diagnosis. Since this site drives several conclusions about static driver limitations, I suggest strengthening it with one concrete extra analysis. I guess the simplest option is to show a time series of observed albedo, reflected divided by incoming, and the modeled equivalent for HAN and one well behaving site. Another option is a small sensitivity test with plausible albedo changes, for example plus or minus 0.05 to 0.10, or an alternative grass parameter set, to show how much of the bias is realistically explainable. One of these would turn the HAN section into a clean quantitative lesson for the community. However, I raise it here as an optional extra analysis which, if doable, will enhance the analysis.**

Answer: *Thank you for highlighting this important issue and for your valuable suggestion. We have conducted the analysis and included the relevant information in the revised manuscript on lines 491-499. Additionally, we have provided an extended analysis in the figures presented below.*

1. EVALUATION #1 TIME SERIES OF OBSERVED ALBEDO: *The figure below compares observed and PALM modelled albedo at the FSV (well-behaving) station and the problematic HAN station during episode e5. Albedo was calculated as the ratio between reflected and incoming short-wave radiation, and only periods with SWin > 10 W* m^{-2} *were considered to avoid unrealistically large values associated with low incoming short-wave radiation. The observations indicate differences between the two observation stations. The HAN station exhibits substantially higher albedo than FSV throughout most of the daytime period, with albedo ranging from approximately 0.16 to 0.39, whereas at FSV it remains lower, generally between 0.11 and 0.23. Both stations show a pronounced diurnal cycle, with larger albedo values during the early morning and late afternoon and lower values around midday, consistent with the influence of solar zenith angle on surface reflectivity. The PALM model reproduces the magnitude and temporal evolution of albedo at FSV reasonably well, although it systematically underestimates the albedo calculated from observations by approximately 0.01–0.05. On the other hand, PALM underestimates albedo at the HAN station, with simulated values remaining near 0.08–0.20, whereas albedo calculated from observed values frequently exceeds 0.20 and reaches nearly 0.40 during the morning hours. Overall, the results suggest that PALM captures the albedo characteristics at the FSV station more successfully than at the HAN station. The systematic low bias at HAN may indicate deficiencies in the representation of surface properties and land-cover characteristics that control surface albedo during the episode.*



2. EVALUATION #2 VALUES AT THE SURFACE VERSUS SENSOR HEIGHT VALUES: *An additional analysis was performed to show the differences between outputs at the surface and the sensor sampling height. The figure below compares observed and modelled incoming and outgoing short-wave radiation at the HAN station during episode e5. In addition to the observational data, the modelled radiation fields are shown at the surface and at the sensor height corresponding to the measurement location. The incoming short-wave radiation simulated at sensor height reproduces the observations very well. The model captures both the timing and magnitude of the diurnal cycle, including sunrise, sunset, and the daytime maximum. In contrast, the incoming short-wave radiation at the surface shows pronounced reductions. These decreases are attributed to potential shadowing effects caused by varying plant canopy densities, which affect the radiation reaching individual surface elements. The lower plant canopy does not affect the radiation field at the measurement height above the surface; grass is typically lower, at a maximum of 1 m. The outgoing short-wave radiation also differs substantially between the two model levels. The surface level one produces larger outgoing short-wave fluxes than at the sensor height, but both modelled fluxes remain below the observed values throughout most of the daytime period. This comparison suggests the importance of diagnostic height within the urban canopy, as surface radiation fields are strongly influenced by local shading effects that are not apparent at sensor height.*



4. **The clear sky versus cloudy episode grouping is a good idea, but it needs an objective definition so others can reuse the protocol. A short criterion based on observed incoming short wave smoothness, or a threshold in modeled cloud fraction, would be enough.**

Answer: *Thank you for this suggestion. We agree that a more explicit description improves clarity. We have now added a detailed explanation of the protocol and criteria used to select and classify clear-sky versus non-clear-sky episodes. Please see the updated text in the revised manuscript in the 2.5 "Simulation episodes and meteorological conditions" section and Lines 273-177.*

5. **The paper is right to emphasize PALM superiority in resolving geometry driven shading and reflections, but please keep the wording careful so it does not sound like PALM is resolving cloud processes. The results show excellent redistribution and timing under clear sky and limited ability to correct wrong downwelling radiation when cloud timing is off in the driving model. That nuance is important for a GMD audience.**

Answer: *We thank the reviewer for this observation. We agree that the manuscript should more clearly distinguish between the simulation of urban radiative processes and the representation of cloud effects. In the PALM configuration used in this study, clouds are not explicitly simulated. Instead, the incoming radiative forcing, including the effects of cloudiness, is prescribed through the external dynamic driver. To avoid this potential misunderstanding, we have revised the manuscript to explicitly state this limitation. (Please see L139 in the 2.1 PALM configuration section).*

6. **There seems to be a small inconsistency between the station year labeling in the table and the narrative description of which sites were active in which year. Please double check that so readers do not get confused about the measurement periods.**

Answer: *The measurement periods and station nomenclature have been harmonised for consistency. Please refer to the revised text in section 2.4 and to Table 2 for these changes.*

7. **Please state clearly the model sampling height and how it relates to the pyranometer installation height, and confirm whether reflected radiation is taken directly from the surface flux or from a level that matches the sensor exposure.**

Answer: *Thank you for this comment. We agree that the relationship between the model sampling height and the pyranometer installation height was not described clearly in the original manuscript. We have revised the text to explicitly state the requested. This clarification has been added to the description of the PALM simulation setup, and we kindly refer the reviewer to our detailed response to Comment #1 and in the manuscript, Lines 170-174.*

8. **The pyranometer specification values in the sensor table look inconsistent in places, especially typical accuracy versus resolution. Please double check the units and the transcription from the original sensor documentation.**

Answer: *We have checked the pyranometer specifications with the original technical documentation and calibration lists. We have corrected these values and verified all units in the sensor table to ensure complete consistency with the official sensor data sheets. Please see the updated Table 3.*

9. **(Optional) A simple schematic showing the chain from WRF to PALM meteo to dynamic driver to PALM RTM, including how direct and diffuse short wave are handled, would improve readability. The text explains this, but a diagram would make it much easier to follow at a glance.**

Answer: *We thank the reviewer for this nice suggestion. We agree that a visual representation of the modelling chain enhances clarity. Following the reviewer's recommendation, we have added a new schematic diagram as Appendix B at Line 735. Furthermore, we offer a brief explanation to enhance the answer provided:*

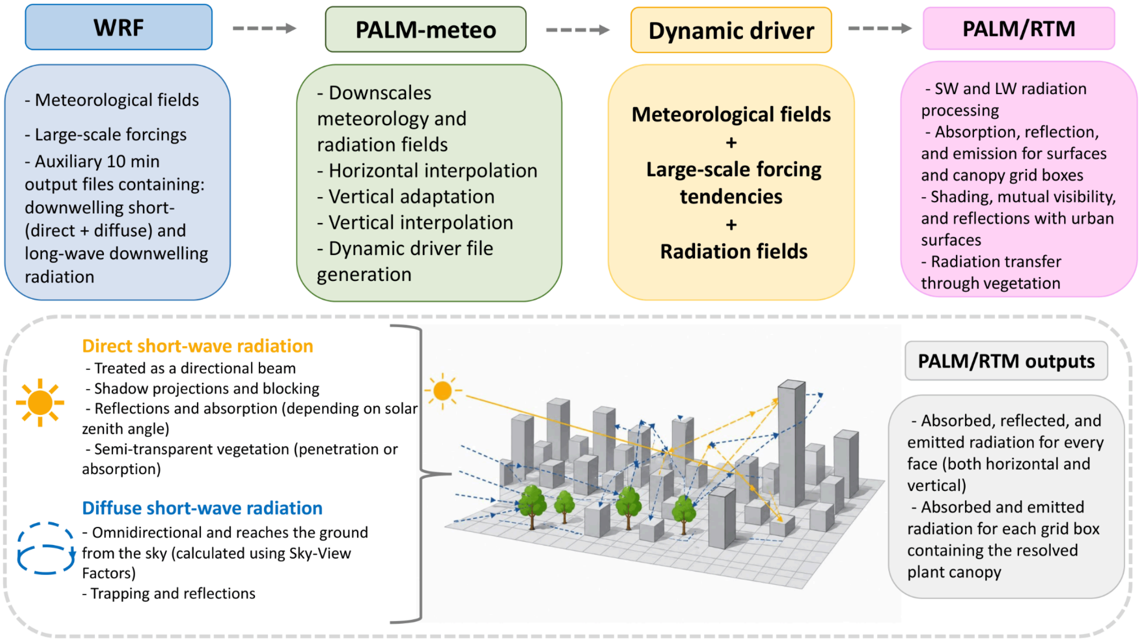
Essentially, this diagram illustrates the sequential data flow from the meso-scale WRF forcing through the PALM-meteo pre-processor, the dynamic driver, and PALM/RTM. In essence, direct and diffuse short-wave radiation components are taken from the WRF model, processed by the PALM-meteo into the dynamic driver file, transferred through the PALM dynamic driver into the PALM simulation, and subsequently modified by the PALM radiative transfer model (RTM) through interactions with urban geometry and vegetation.

Direct Radiation: *Handled using a ray-tracing algorithm. The solar position from ray-tracing, which determines shading by terrain, buildings or trees, is discretised to the nearest pre-calculated position (resolution 4.5°), but the angle of incidence to the target surface is exact at each time-step.*

Diffuse Radiation: *Handled using Sky View Factors (SVFs). Diffuse radiation scatters across the entire sky hemisphere. Instead of tracking individual rays, PALM uses the SVF to calculate the percentage of the open sky visible to any given surface (walls, roofs, or ground), adjusting the overall diffuse downwelling flux.*

Reflected Radiation: *Handled using a radiosity algorithm using view factors towards other surface elements (pre-calculated by ray-tracing) and their short-wave radiosity.*

Diagram: *Schematic representation of the WRF-to-PALM modelling workflow applied in this study. The figure illustrates how meteorological and radiation fields are passed from the WRF model through the PALM-meteo processor and the dynamic driver file into PALM and its radiation transfer module (RTM).*



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/m²]**	**MBE [MJ/m²]**	**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/m^2]	MBE [MJ/m^2]	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/m^2]	MBE [MJ/m^2]
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/m^2]**	**MBE [MJ/m^2]**
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).*

Reviewer #3

The paper by Radovic et al. (2026) presents an evaluation of the PALM model's radiation module RTM against measurements from four sites in terms of shortwave radiation. PALM is forced with radiation from two different WRF set-ups for 16 days and run in spin-up mode, i.e. without the resource-consuming calculation of air quantities. PALM is a complex model and has mostly been evaluated with full, realistic set-ups and quantities that require different components of PALM. This paper focusses on a systematic evaluation of one component and is thus highly welcome. In most parts of the paper, in particular in the title and the abstract, this, however, does not become clear. In addition, since the model WRF is used as input, it is actually an evaluation of the coupled WRF/PALM-RTM set-up, which introduces additional uncertainties. I would like to ask the authors to make this clearer in the paper and to clarify the relationship with the stations Karlov and Libus (details below). I thus recommend consideration for publication after major revision.
Major issues:

1. **This paper mostly evaluates the RTM module of PALM. While the land-surface and the building surface module as well as the mesoscale nesting module are also used, the former only supplies surface albedo and the latter facilitates only the radiation input to my understanding. In particular, I assume that only WRF *radiation* data is used and not other fields; the latter is implied by "dynamic meteorological forcing" (L159). Please clarify this in the paper, in particular in the abstract. Are the surface albedos constant in time or (partly) sun-angle-dependent?**

Answer: *During the spinup mode, PALM uses only the incoming shortwave and longwave radiation from the WRF mesoscale forcing to drive the RTM module and performs surface energy balance calculations. Other meteorological variables are not dynamically coupled to the PALM simulation per se; they are used from WRF to create the “complete” dynamic driver set. The wind speed, humidity, and temperature are prescribed through the configuration p3d file. In the present setup, temperature follows an idealised daily sinusoidal cycle, while the remaining atmospheric variables are kept constant in time, and present only in the dynamic driver.*
We will clarify in the manuscript that mesoscale nesting is used solely to provide radiative forcing fields rather than full dynamic meteorological forcing, and that the remaining fields are downscaled into the dynamic driver file but not used. Please see Lines 204-206 in the revised manuscript.
Surface albedo (reflectance) values are prescribed for each individual surface grid element and remain constant in time; no explicit incidence-angle-dependent albedo parameterisation is applied. On the other hand, the total effective albedo (ratio of reflected to incoming radiation) for larger areas does change in time as a result of changing solar geometry when the 3-D radiation with multiple reflections is simulated by RTM. We have revised the manuscript to address the comments regarding albedo. Please refer to the newly included section 2.1.1, “Spin-up

simulations," in which all albedo-related concerns are addressed. Additionally, we would also like to point to our reply to Reviewer #1's comment #3 regarding albedo, in which we evaluated the time series of observed albedo and discussed the additional analysis presented.

2. **The authors also imply that PALM-RTM could partly correct errors in the radiation input (L12, L290, L357, L376). I think that this is not the case. I consider the radiation fluxes of WRF to be above its (unresolved) canopy and exactly like this, it is considered as input in RTM. RTM mostly distributes it geometrically within the canopy without doing any atmospheric adjustments. This is why using WRF as input actually results in an evaluation of the coupled WRF/PALM-RTM system. Thus, forcing with radiation measurements above the canopy, for example from rooftops, would have removed the uncertainty from the evaluation. Please discuss this. What about the stations Karlov and Libus? Their data is used in the paper but both stations are not introduced at all. Please describe these stations as well. Could their data be used as forcing?**

Answer: *We agree that PALM-RTM does not explicitly correct biases in the incoming radiation forcing from WRF. The RTM primarily redistributes incoming shortwave and longwave radiation within the urban canopy through geometric shading, absorption, reflection, and multiple scattering. Therefore, uncertainties in the WRF radiation forcing may propagate into the PALM simulations.*

Using radiation measurements above the canopy as forcing would indeed reduce uncertainties associated with mesoscale radiation forcing and would allow a more isolated evaluation of RTM performance. However, the objective of the present study was to assess the applicability of PALM-RTM under realistic mesoscale-forced conditions representative of operational urban climate simulations.

We will additionally expand the description of the Karlov and Libuš stations in the manuscript, including their locations, measurement characteristics, and observational datasets used in the evaluation. While these stations could potentially provide observational radiation forcing data, in the present study, they were used exclusively as out-of-domain reference evaluation stations and not as direct model forcing. Please see Lines 254-262 in the revised manuscript.

Minor issues:

3. **L12: Does PALM, in particular RTM, compensate for any errors in the radiation input? My understanding is that RTM distributes the radiation within the canopy received as input at the top of the canopy. This input is expected to be correct.**

Answer: *The incoming radiation input coming from WRF (and further dynamic driver) can contain errors (e.g., cloud representation and timing of cloud cover, aerosols). So, the radiation arriving at the canopy top of the PALM domain is not guaranteed to be necessarily "correct" or close to reality. The RTM uses whatever radiation it receives as boundary/input forcing and, per se, is not designed to diagnose or correct biases in the forcing; it just redistributes and* *physically processes the given data within the canopy. Hence, the quality of the RTM results still depends on the accuracy of the incoming radiation forcing.*

4. **L31: "recognised by THE World Meteorological Organization"**

Answer: *We thank the reviewer for the correction. The text has been updated to include the missing article (please see Line 32).*

5. **L46: As the authors write in L45, MRT cannot be derived from shortwave radiation alone, but longwave radiation needs to be considered as well.**

Answer: *We thank the reviewer for pointing out this inconsistency. The sentence is revised to clarify the issue. Now it states: "Outdoor MRT combines the impacts of short-wave and long-wave radiation fluxes in outdoor environments. While long-wave radiation is an important component of the overall radiative balance, especially in indoor or shaded environments, short-wave radiation is often the dominant driver of elevated MRT values outdoors, particularly at high solar altitudes and under clear-sky conditions. Although MRT derivation strictly requires integrating both short-wave and long-wave radiation flux densities, the variability of outdoor MRT during the daytime is heavily influenced by short-wave radiation flux \citep{lee2014}." For the corrections, please see Lines 45-50.*

6. **L106: Without parentheses around the citation.**

Answer: *Done. The parentheses around the citation have been removed. For the corrected manuscript, please see Line 106.*

7. **L144: Are there any differences in the results of RTM 4.1 compared to RTM 4.0 or RTM 3.0 described in Krc et al. (2021) when only a 2.5D geometry is used? The description mentions only numerical advancements.**

Answer: *The differences are indeed mostly computational enhancements (such a, localized raytracing, improved MPI exchanges, etc.), with the single exception of the full 3D geometry as opposed to the 2.5D geometry in RTM 3.0. However, the actual number of overhanging structures in the simulated scenario is very small, and none of them are close to the measurement locations, therefore, we do not expect this to have an impact on the results.*

8. **Section 2.2: Please include more details of the WRF simulations:**

   *** How is WRF forced? Only the discussion section mentions ERA5.**

Answer: *We apologise for the omission in the Methods section. The WRF model was indeed driven by ERA5 reanalysis data. We have updated section 2.2 "Initial and boundary conditions" and Lines 206-207 to explicitly state this and provide the necessary technical details regarding the WRF's forcing data.*

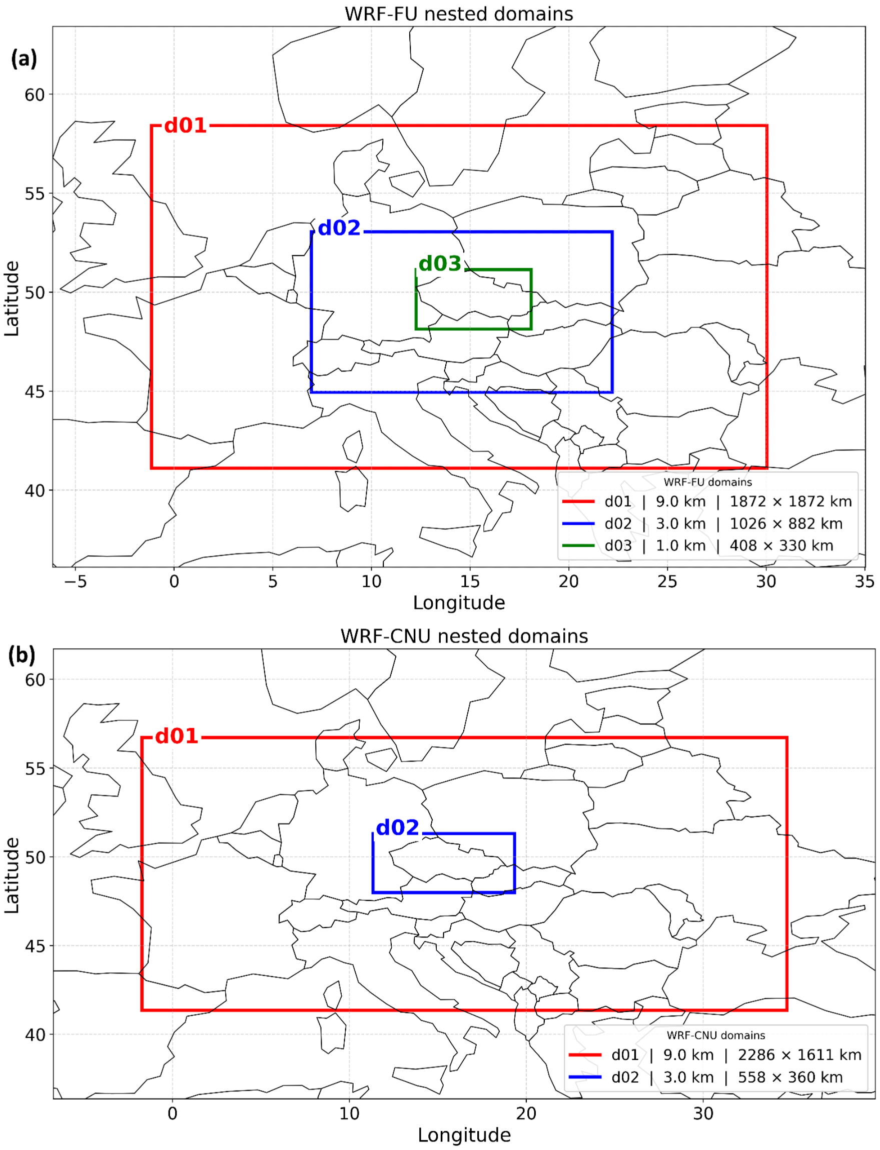
   *** What are the domain sizes, and are the FU simulations nested into CNU? Or are there any other nesting steps in between? If not, is the difference in**

resolution between the forcing of WRF and WRF itself (in particular for the FU simulations) not a problem?

Answer: *We thank the referee for pointing out the omission of the description of WRF domains and nesting. We added the information to section 2.2, “Initial and boundary conditions,” and to Lines 200-202. In addition, we provided a detailed Figure S14 in the supplementary materials to support our answer, as well as here as the continuation of our answer.*

To clarify, the CNU and FU WRF simulations were performed in two independent domain configurations. The CNU simulations were performed as a two-domain setup, with an inner domain at 3 km resolution and an outer domain at 9 km resolution. The FU simulations have a triple-nesting chain with resolutions of 9 km, 3 km and 1km. We tried to make the WRF setup standard. We think that ERA5 (with 0.25° resolution) forcing WRF with 9km horizontal resolution is a frequent choice, as well as further nesting of WRF domains to 3 km and 1 km resolution. From today’s operational forecasting perspective, the CNU setup is rather minimalistic. The triple nesting in FU with domain sizes as depicted in the picture is (in our opinion) comparable to similar cases reported in the literature.

Figure: WRF model simulation domain configurations and nested grid hierarchies for the two experimental setups: (a) the three-domain WRF-FU setup (red: d01 at 9.0 km, blue: d02 at 3.0 km, and green: d03 at 1.0 km horizontal resolution) and (b) the two-domain WRF-CNU setup (red: d01 at 9.0 km and blue: d02 at 3.0 km horizontal resolution). The legends denote the horizontal grid spacing and total geographic coverage area for each parent and nested domain.



9. L170: Is the diffuse shortwave radiation also stored?

Answer: *Yes, the diffuse short-wave radiation is stored in the WRF model's 10 min-resolution output files and passed to PALM-meteo for further processing into the dynamic driver. In our setup, the* ***SWDDIF*** *variable ("Shortwave surface downward diffuse irradiance") from the WRF output is mapped to the* ***rad_sw_in_dif*** *variable within the PALM dynamic driver. This allows PALM to use the meso-scale model's specific partitioning of direct and diffuse radiation as a* *boundary condition, rather than relying on an internal parameterisation for the split into direct and diffuse components, which would be applied if the external diffuse component were unavailable.*

10. **Table 1: Which urban canyon parametrization is used? Probably BEP (Martilli et al. 2002), however, SLUCM+BEM (Takane et al. 2024) is also available.**

Answer: *The WRF simulations utilised the building environment parameterisation (Martilli et al., 2002), in combination with a building energy model, so BEP+BEM by Salamanca and Martilli (2009), corresponding to* `sf_urban_physics = 3` *in WRF. We have clarified this in the revised manuscript in Table 1. We also thank the reviewer for pointing out the recent work by Takane et al. (2024) on the SLUCM+BEM implementation, and we believe it could be explored and tested in future studies.*

11. **Table 4: Columns Category and CNU/FU are redundant.**

Answer: *We agree with the reviewer's observation regarding the redundancy in Table 4. We have removed the redundant columns in Table 4.*

12. **L220: Please explain exactly the output quantities: Are the values taken from the bottom surface (height 0m) at the locations of the measurements? For completeness: what does SWin include, only direct and diffuse radiation from the sky or also reflections from the surroundings?**

Answer: *We clarify that the SWin variable is not limited to the direct and diffuse solar radiation coming down from the sky, but it accounts for secondary reflections from the surrounding environment (such as neighbouring walls, ground, and vegetation). The model simulates reflection steps among surfaces that share a mutual view factor. At each step, surfaces and canopy elements absorb a portion of the radiation, while the remainder is re-reflected according to surface albedo, whereas the facets with no mutual view to other structures (e.g., unobstructed roof surfaces) receive only the direct and diffuse sky radiation, meaning their incident shortwave irradiance does not receive reflected environmental components. The values are taken from the grid cell nearest to the vertical position of the pyranometers, that is, 1 m above ground.*

For the revised manuscript, please look at section 2.1.1 "Spin-up simulations" and Lines 170-274 for the first part of your comment.

For the second part, please see our response to Reviewer 1, which addresses the optional suggestion to include a diagram. We have implemented this suggestion in Appendix B of the manuscript.

13. **Figures 3 and 4: According to Table 4, the common episodes are e1 to e6. Why do the captions say it is (e3, e5, e6, e8, e9, e16)?**

Answer: *We apologise for this inconsistency. The discrepancies in the episode numbering in the captions of Figures 3 and 4 were a clerical error resulting from a later episode's re-indexing* *during the final manuscript preparation. We have updated the captions to correctly match the numbering (e1–e6) used in Table 4. Please see Figures 3 and 4.*

14. **Figure 5, in particular (a): Please discuss why the ratio of PALM-CNU In to WRF-CNU In is so different from the ratio of PALM-FU In to WRF-FU In. Is this related to the the relationship of diffuse and direct radiation? This would highlight that not only the total shortwave input needs to be correct but also the distinction between diffuse and direct.**

Answer: *We thank the reviewer for highlighting this important topic. To investigate this observation, we examined the diffuse shortwave radiation supplied to PALM via the dynamic driver, as well as the WRF input data for both simulation setups. The qualitative analysis reveals substantial differences between the PALM-CNU and PALM-FU simulations in the diffuse shortwave radiation forcing itself, with PALM-CNU exhibiting considerably larger diffuse radiation values throughout the daytime period (see Figure 1). This finding supports the reviewer's interpretation that the differing relationships between WRF and PALM incoming shortwave radiation are likely related to differences in the partitioning between direct and diffuse radiation components. Since PALM further processes the incoming radiation through urban radiative transfer mechanisms such as shading, obstruction, and reflections, the diffuse-to-direct ratio becomes highly important for the resulting near-surface radiation fields. Consequently, even when total incoming shortwave radiation is comparable (as shown in Figure 2), differences in diffuse radiation forcing can produce substantially different urban radiative responses within PALM. These results, therefore, highlight that accurate representation of the diffuse and direct shortwave radiation components is crucial for reproducing total radiation fluxes, and that PALM simulations are sensitive to the diffuse/direct partitioning inherited from the mesoscale forcing model.*

An additional discussion addressing this valuable point has been added to Section 4.2, "Driving data impact: WRF-CNU vs. WRF-FU configuration performance" (Lines 249–255).

Figure 1. *Comparison of diffuse short-wave radiation supplied to PALM-CNU (solid blue) and PALM-FU (dashed orange) setups through the dynamic driver (panel (a); variable rad_sw_in_dif) with the corresponding forcing data from WRF-CNU (solid blue) and WRF-FU (dashed orange) (panel (b); variable SWDDIF) for the clear-sky episode e6 at the FLE station.*

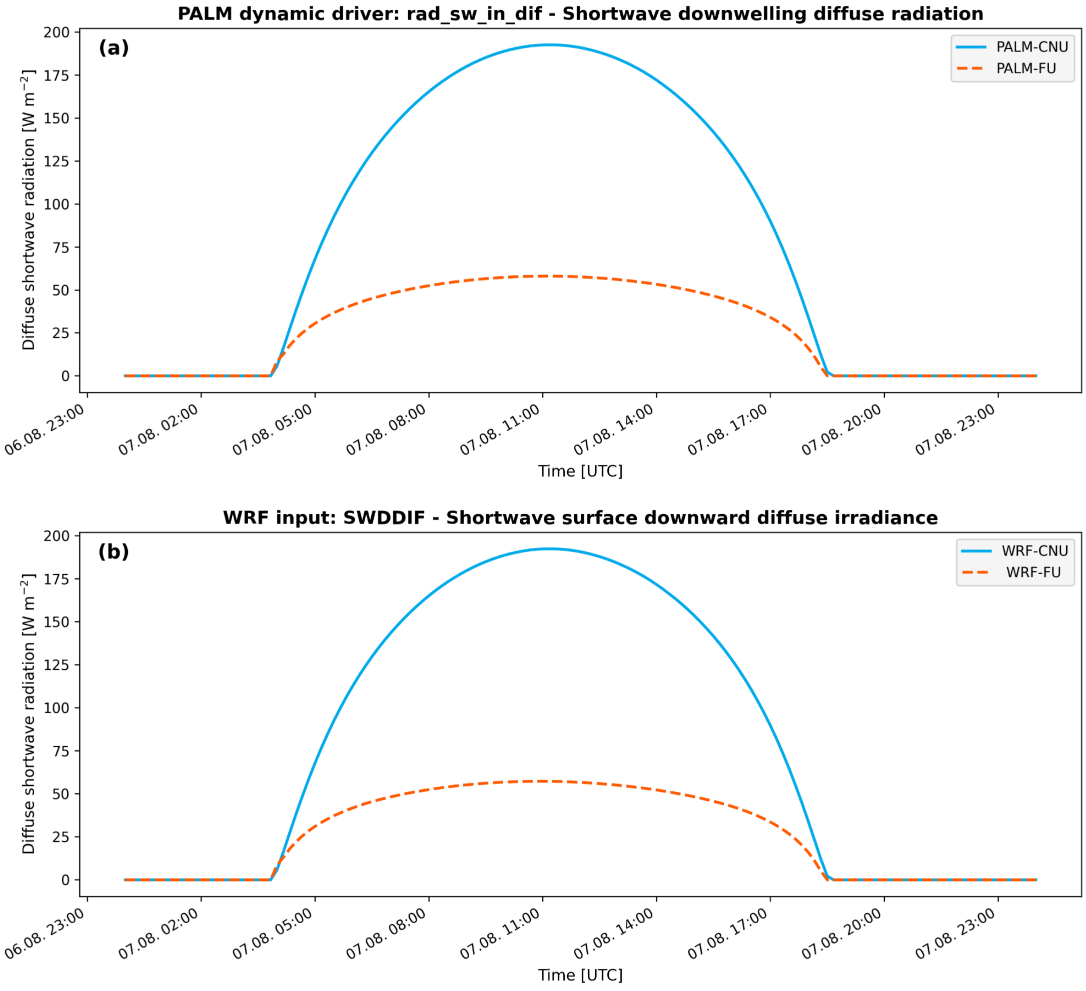
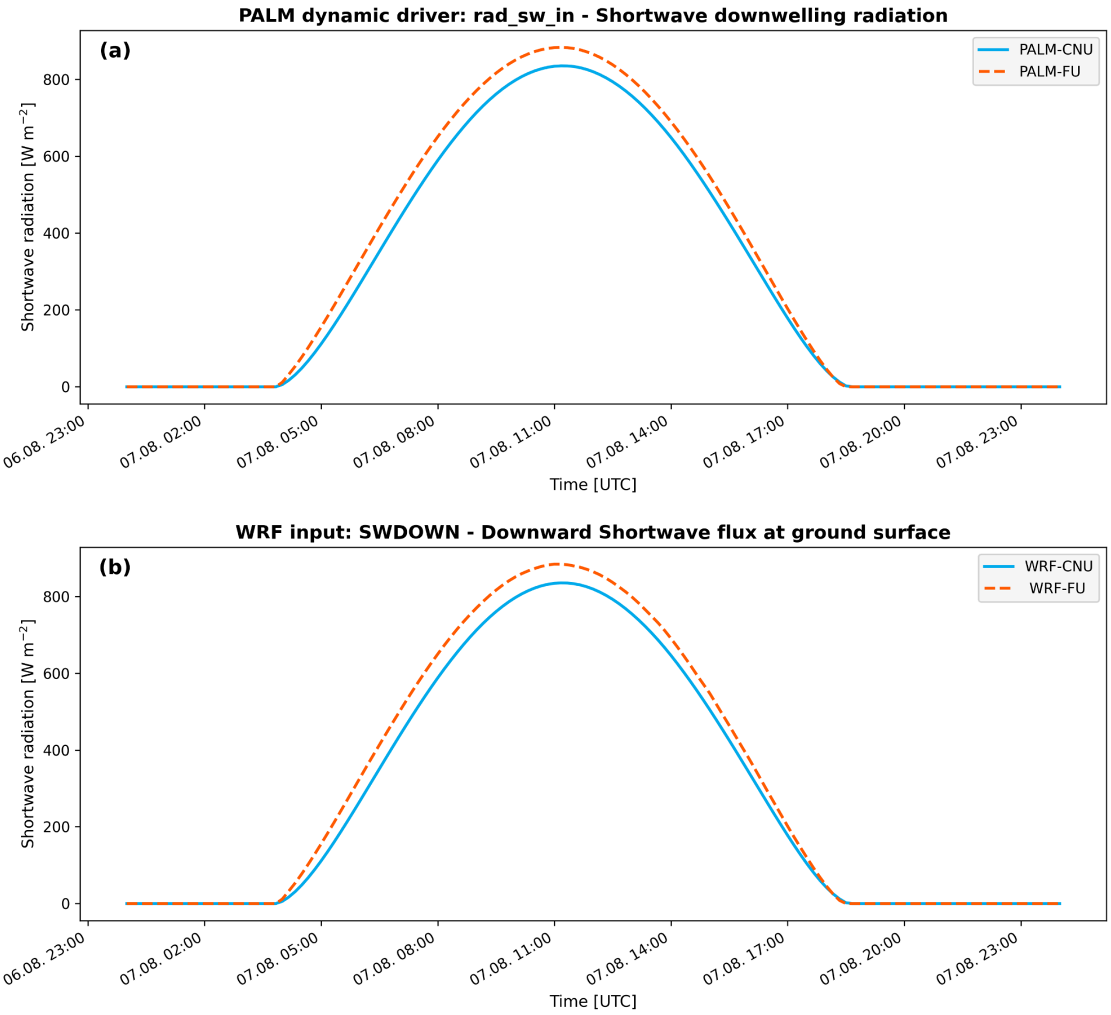


Figure 2. *Comparison of the direct short-wave radiation supplied to PALM-CNU (solid blue) and PALM-FU (dashed orange) setups through the dynamic driver (panel (a); variable rad_sw_in) with the corresponding forcing data from WRF-CNU (solid blue) and WRF-FU (dashed orange) (panel (b); variable SWDOWN) for the clear-sky episode e6 at the FLE station.*



15. L331: episode e5 while Figure 6 says e9.

Answer: *Once again, we apologise for this inconsistency. The discrepancies in the episode numbering are due to the reasons explained in our answer to your previous comment. We have updated the caption to correctly match the numbering used in Table 4. Please see Figure 6.*