

# Response to Reviewer #1 Comments

*WRF-MLGR v2.0 / EGUsphere Green-Roof Manuscript*

## Reviewer Comment 1

---

**Reviewer comment:** The green-roof soil depth is only 0.03 m per layer. Please discuss potential numerical stability issues under extreme conditions (e.g., prolonged heat waves) when soil moisture may dry out and temperatures may rise rapidly. Have you tested model behavior under such extremes, and are any timestep or scheme adjustments needed?

### Response:

Thank you for this important comment. We agree that numerical stability under dry and hot conditions can be an important consideration for multilayer diffusion schemes, particularly when substrate moisture becomes low and surface temperatures can change rapidly.

First, we would like to clarify that the green-roof substrate layers in our WRF-MLGR configuration are not all fixed at 0.03 m. The multilayer green-roof scheme uses non-uniform layer thicknesses, and the layer depths are specified according to the model configuration rather than being constrained to a constant 0.03 m for each layer.

Second, the model has been evaluated under contrasting environmental conditions, including dry, wet/moist, high-radiation, and warmer periods. In a companion study, [Martinez Mendoza et al. \(2026\)](#) evaluated green-roof model performance, including WRF-MLGR, across different moisture and radiative-forcing regimes using observations from the same green-roof site. That analysis included dry and rainy periods, as well as high- and low-net-radiation conditions. These conditions represent the range available in the observational dataset, although they may not fully encompass more extreme or prolonged multi-day heatwave scenarios. The model remained numerically stable across the tested conditions, and no numerical instabilities associated with rapid substrate drying or rapid increases in surface temperature were observed. Although the version evaluated in the companion study did not include the updates introduced in the present article, the fundamental numerical solution of soil heat and moisture diffusion remains unchanged.

Third, heat conduction and moisture movement within the green-roof substrate are solved using an implicit matrix-based approach. In the heat routine, the temperatures of the soil layers are solved together using a tridiagonal matrix, and in the moisture routine, the soil-water contents are updated using the same type of matrix solution. This makes the model more numerically robust than a purely explicit layer-by-layer update. However, an implicit method does not automatically guarantee accurate results. If the time step is too large compared with the layer thickness and the speed of heat or moisture movement, the solution can become overly smoothed or physically unrealistic. Therefore, reasonable layer thicknesses and time steps are still needed for reliable model performance.

## Reviewer Comment 2

---

**Reviewer comment:** Section 2.2 describes the original model in text. Consider adding a schematic or flow diagram that contrasts the original scheme and the modified scheme to make the differences clearer (e.g., layer structure, heat flux pathways, and where parameter changes are applied).

### Response:

Thank you for this helpful suggestion. We agree that a schematic comparison of the original and modified schemes improves the clarity of the model description. In response, we have added a new schematic diagram as Figure 1 in

the revised manuscript. This figure contrasts the structure of the original model with that of the modified model and highlights the main process-based changes introduced in this study. In particular, it illustrates the layer structure, the main heat and moisture flux pathways, and the locations where the parameter and process modifications were applied. We believe this addition makes the differences between the original and modified schemes much clearer for the reader.

### Changes made in the manuscript:

We revised Sect. 2.2 of the manuscript, (page 6, lines 160–163), to clarify the structural and process-level differences between the original and modified WRF multi-layer green-roof schemes. Specifically, we added a new schematic as Fig. 1 and added explanatory text introducing the figure. The revised text now reads:

*“To clarify the structural and process-level differences between the original and modified WRF multi-layer green-roof schemes, Fig. 1 provides a schematic comparison of the two model formulations. The figure highlights the main layer structure, the key heat and water flux pathways, and the locations of the five process-based modifications introduced in Sect. 2.3.”*

The new figure added to the revised manuscript is reproduced below for clarity:

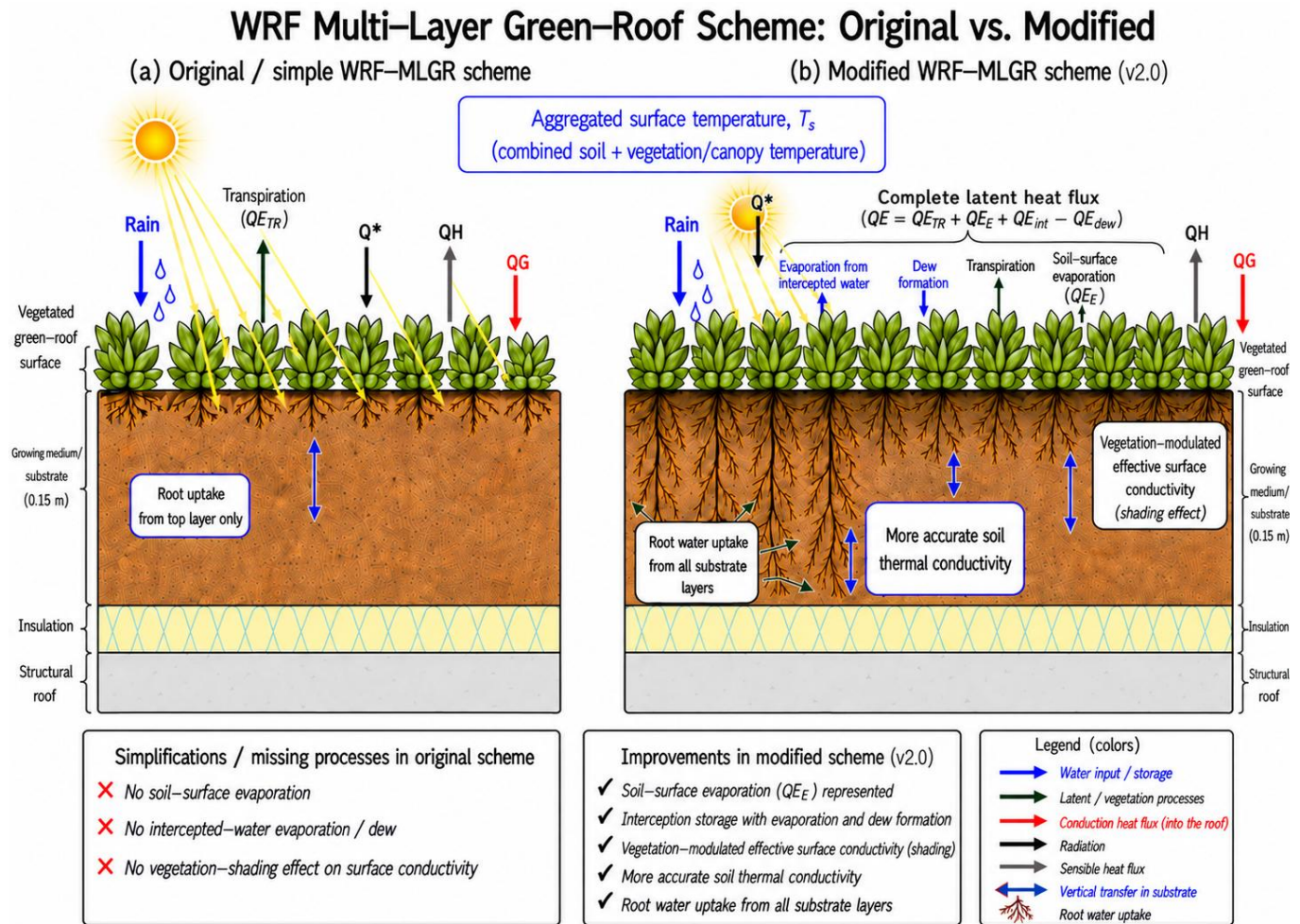


Figure 1. Schematic comparison of the original and modified WRF multi-layer green-roof schemes.

## Reviewer Comment 3

**Reviewer comment:** In Sections 2.3.1 and 2.3.2, please quantify how much the soil and vegetation thermal conductivities were changed relative to the original model. Are the modified values still within physically reasonable/observed ranges? If possible, include a table or figure showing the original vs. modified conductivity values and their sources or justification.

### Response:

Thank you for this helpful suggestion. We agree that quantifying the changes in soil and vegetation thermal conductivity relative to the original model improves the clarity of Sections 2.3.1 and 2.3.2.

In response, we have added a new comparison figure as Fig. S1 in the Supplement. This figure compares the depth-averaged soil thermal conductivity simulated by the original and modified model configurations during summer 2014. Because thermal conductivity varies among substrate layers in the multilayer scheme, we calculated and presented the average thermal conductivity across the green-roof substrate layers for both the original and modified models. This allows a clearer comparison of the overall change in effective substrate thermal conductivity between the two model versions.

### Changes made in the manuscript and Supplement:

We revised Sect. 2.3.2 of the manuscript, (page 8, lines 215–221), to quantify the effect of the revised thermal treatment and to explain why the modified conductivity values are physically reasonable. We also added a new supplementary figure, Fig. S1, to show the depth-averaged soil thermal conductivity simulated by the original and modified schemes during summer 2014. The revised manuscript text now reads:

*“To quantify the effect of the revised thermal treatment, we compared the time-varying average soil thermal conductivity produced by the original and modified green-roof schemes during summer 2014 (Fig. S1 in the Supplement). The original formulation produced unrealistically large and highly variable conductivity values, with peaks exceeding  $49 \text{ W m}^{-1} \text{ K}^{-1}$ , indicating excessive conductive heat transfer under wet conditions. In contrast, the modified formulation produced conductivity values mainly within the range  $0.32$  to  $0.98 \text{ W m}^{-1} \text{ K}^{-1}$ , which is physically more reasonable for engineered green-roof growing media, for which thermal conductivity has been shown to vary strongly with substrate moisture content and composition (Barozzi et al., 2017; Shao et al., 2021).”*

*The new supplementary figure is reproduced below for clarity:*

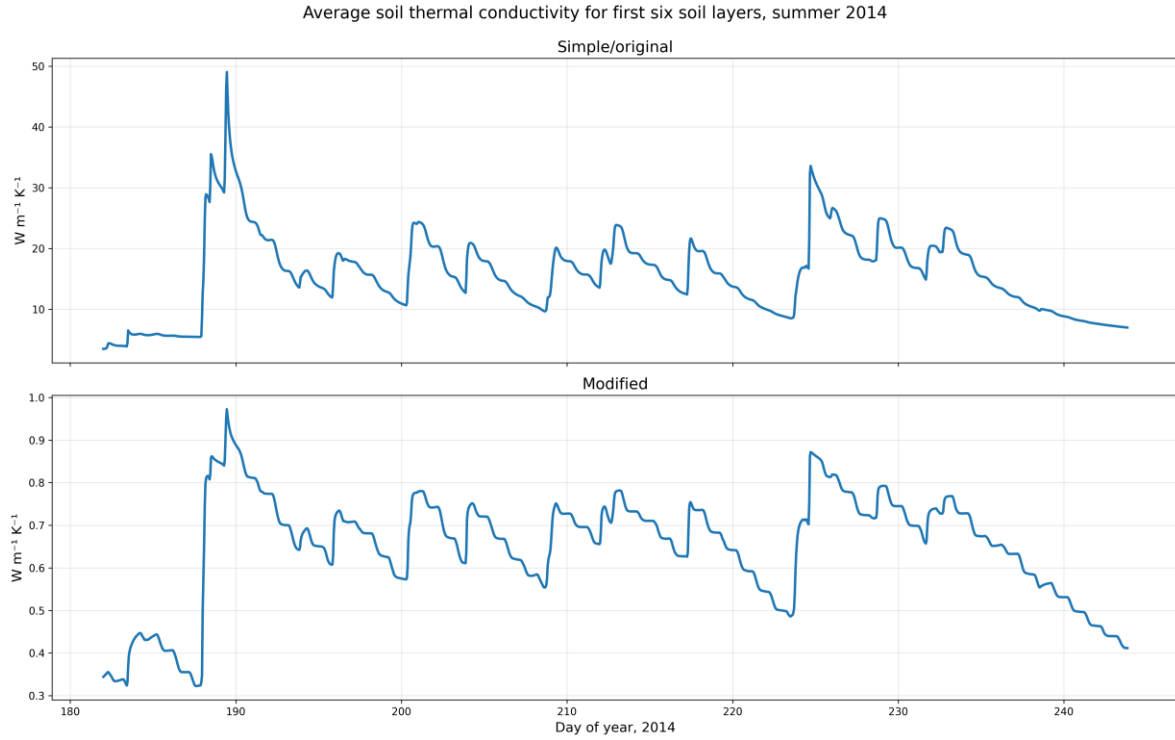


**Figure S1.** Comparison of depth-averaged soil thermal conductivity between the original and modified WRF multi-layer green-roof schemes during summer 2014.

The modified values remain within physically reasonable ranges reported for green-roof substrates and vegetation materials. We have also revised the text in Sections 2.3.1 and 2.3.2 to clarify the magnitude of the changes and to explain the physical basis for the modified parameter values. Because several additional diagnostic figures are already included in the Supplementary Material, we placed this comparison there rather than in the main manuscript.

## Reviewer Comment 4

**Reviewer comment:** Table 1 and Fig. 1 show degraded performance in  $T_s$  prediction across all cases. Please discuss possible causes. Clarify which surface temperature is reported (soil surface, vegetation canopy surface, or some aggregated surface skin temperature). You mention two thermal conductivity calculations-how do those relate to the reported  $T_s$ , and might they explain the degradation? Consider adding separate diagnostics for soil-surface and vegetation-surface temperatures if available.

### Response:

Thank you for this important comment. We agree that the small drop in model performance for surface temperature prediction requires further clarification and discussion.

In the current WRF-MLGR formulation, the reported surface temperature,  $T_s$ , represents a simplified bulk/skin surface temperature of the green-roof system. The model does not explicitly solve separate temperatures for the vegetation canopy surface, the air space within the canopy, and the soil/substrate surface. Instead, these components are represented in a simplified manner through a single surface temperature. Therefore,  $T_s$  should be interpreted as an aggregated surface skin temperature rather than a separately resolved soil-surface or vegetation-canopy temperature.

We believe this simplified representation is one of the main reasons for the weaker  $T_s$  performance. Because the current model does not explicitly resolve these separate thermal components, improvements in substrate heat transfer and vegetation-related parameterizations do not necessarily translate directly into improved  $T_s$  prediction.

The two thermal conductivity calculations discussed in Sections 2.3.1 and 2.3.2 affect the heat transfer within the substrate. These changes improve the representation of heat storage and ground/substrate heat fluxes, but they do not introduce a fully separate canopy energy-balance model. Therefore, they may influence  $T_s$  indirectly, but they do not fully address the structural limitation associated with representing the green roof using a single bulk surface temperature.

Although  $T_s$  remains an important diagnostic variable, improving the surface energy-flux partitioning has higher priority at this stage of model development, because  $Q_G$ ,  $Q_E$ , and  $Q_H$  are more directly influential in the coupled WRF climate-model system. These fluxes control the exchange of heat and moisture between the urban surface and the lower atmosphere and can therefore affect the broader atmospheric response simulated by the model. While  $T_s$  affects upward longwave flux, the latter is less influential in terms of surface climate than  $Q_H$  (and  $Q_E$  and  $Q_G$ ).

We have revised the manuscript to clarify the definition of  $T_s$  and to discuss this limitation more explicitly. Separate diagnostics for soil-surface temperature, vegetation-surface temperature, and canopy-air temperature are not available in the current WRF-MLGR formulation. Addressing this limitation will require future model development in which canopy temperature, substrate/soil-surface temperature, canopy-air exchange, and their separate energy interactions are represented explicitly rather than being aggregated into a single bulk skin temperature. Such a formulation would allow the respective contributions of vegetation and substrate processes to be computed separately and would provide a more physically consistent basis for evaluating  $T_s$ . It is largely for this reason that we identified the separation of canopy and substrate thermal processes as an important direction for future development of the WRF-MLGR scheme in the final section of the manuscript.

### Changes made in the manuscript:

We revised Sect. 2.2, (page 5, lines 135–137), to clarify that the reported surface temperature,  $T_s$ , should be interpreted as a bulk or aggregated green-roof skin temperature rather than as a separately resolved soil-surface temperature, vegetation-canopy temperature, or canopy-air temperature. The revised text now reads:

*“Therefore, the reported surface temperature  $T_s$  in the current WRF-MLGR formulation should be interpreted as a bulk or aggregated green-roof skin temperature, rather than as a separately resolved soil-surface temperature, vegetation-canopy temperature, or canopy-air temperature.”*

We also revised Sect. 4.1, (page 24, lines 521–527), to explain why biases in  $T_s$  remain after the model updates. The revised text now reads:

*“A key reason for this remaining  $T_s$  bias is that the current WRF-MLGR formulation calculates only a single bulk surface skin temperature for the green-roof system. Separate computation of the soil/substrate surface temperature, vegetation/canopy surface temperature, and canopy-air temperature is not available in the present formulation. As a result,  $T_s$  cannot be directly interpreted as either a soil-surface temperature or a vegetation-canopy temperature, but instead represents an aggregated surface temperature produced by the simplified bulk representation of the green roof. This limits the ability to diagnose whether the remaining  $T_s$  bias arises primarily from the substrate thermal state, vegetation thermal response, or both.”*

Finally, we revised Sect. 4.1, (page 25, lines 535–540), to identify this limitation as an important direction for future model development. The revised text now reads:

*“Future development should therefore separate the vegetation canopy, canopy air, and soil/substrate surface in the model, instead of representing them with one bulk surface temperature. This would allow the model to calculate their heat exchange processes separately, including radiation, turbulent heat exchange, and heat conduction into the substrate. It would also make it possible to identify whether any remaining errors in  $T_s$  come mainly from the vegetation or the soil surface, and to improve the inaccurate model component.”*

## Reviewer Comment 5

---

**Reviewer comment:** The resolution of all figures is quite low. Please replace figures with higher-resolution versions and ensure axes, legends, and labels are clearly legible.

### Response:

Thank you for this helpful comment. In response, we have revised all figures and replaced them with higher-resolution versions. We also improved the readability of the axes, legends, labels, and text to ensure that all figures are clear and legible in the revised manuscript.

## References

---

Martinez, M., Saeedi, A., Voogt, J., and Krayenhoff, S.: Process-based evaluation of green roof models for assessment of heat mitigation efficacy in WRF (v4.3.1) and EnergyPlus (v8.6.0), *Geosci. Model Dev.*, under review, 2026.

We thank the reviewer for these comments and hope the revisions are satisfactory.

Sincerely,

Alireza Saeedi

# Response to Reviewer #2 Comments

## WRF-MLGR v2.0 / EGUsphere Green-Roof Manuscript

*All page and line numbers refer to the revised manuscript submitted with these responses.*

### Reviewer Comment 1

**Reviewer comment:** As mentioned by the authors, there are five parts being improved/revised. However, based on the current methodology, it is not clear how these five parts affect the physical processes included in the module. In addition to a clearer description, a schematic figure showing the structure and improved processes could help. It would also be better to show a comparison of the module structures between the old and modified versions.

#### Response:

Thank you for this helpful comment. We agree that the physical role of the five updates should be stated more directly. We therefore added a short process-level overview at the beginning of Sect. 2.3 that separates the thermal and hydrological changes and explains which fluxes they mainly affect. The revised manuscript also includes Fig. 1, which directly compares the original and modified model structures and shows the locations of the five updates. We also clarified the purpose of the vegetation-dependent effective surface conductivity.

#### Changes made in the manuscript:

We added the following overview in Sect. 2.3 (page 7, lines 169–178):

*“To improve the representation of heat and water exchange in the WRF multi-layer green-roof scheme, we implemented five physically based modifications. The two thermal modifications mainly change how heat moves through the green roof soil and how much energy is stored as  $Q_G$ . The revised soil thermal conductivity provides a more realistic representation of heat transfer as soil moisture changes, while the vegetation-dependent surface conductivity reduces excessive heat storage at the soil surface to represent the effect of vegetation shading on heat storage.”*

*“The three hydrological modifications mainly affect evapotranspiration and soil water. Soil evaporation allows water to evaporate directly from the substrate surface. Multilayer root uptake allows plants to take up water from several soil layers rather than only from the top layer. Canopy interception allows rainfall to be temporarily stored on the vegetation, evaporated back to the atmosphere, and also allows dew formation. Together, these modifications change  $Q_G$  and  $Q_E$  and therefore also affect how the remaining available energy is partitioned into  $Q_H$ .”*

We also clarified the vegetation-dependent conductivity description in Sect. 2.3.2 (page 8, lines 201–205):

*“In the original green-roof scheme, the model does not include an explicit vegetation canopy, so it does not represent vegetation cover or the shading effect of plants. This means the soil surface receives the full incoming radiation, which can lead to unrealistically large surface heat storage, particularly in the morning, with associated impacts on morning sensible and latent heat fluxes. To make the simulated surface heat storage more realistic, we added a vegetation dependent effective surface conductivity that acts as a simplified representation of the vegetation effect on surface heat storage.”*

We also added Fig. 1 in Sect. 2.2 (page 6, around lines 160–165). This figure provides a side-by-side schematic of the original and modified WRF-MLGR schemes. The left panel shows the original model structure and the main simplified or missing processes, while the right panel shows the modified scheme and where the five updates act on the heat and water pathways. We added this schematic because the text description alone made it difficult to see how the individual modifications affect the physical processes. The figure allows the reader to identify the model structure and the role of each update directly.

The schematic added to the revised manuscript is reproduced below for clarity:

# WRF Multi-Layer Green-Roof Scheme: Original vs. Modified

(a) Original / simple WRF-MLGR scheme

(b) Modified WRF-MLGR scheme (v2.0)

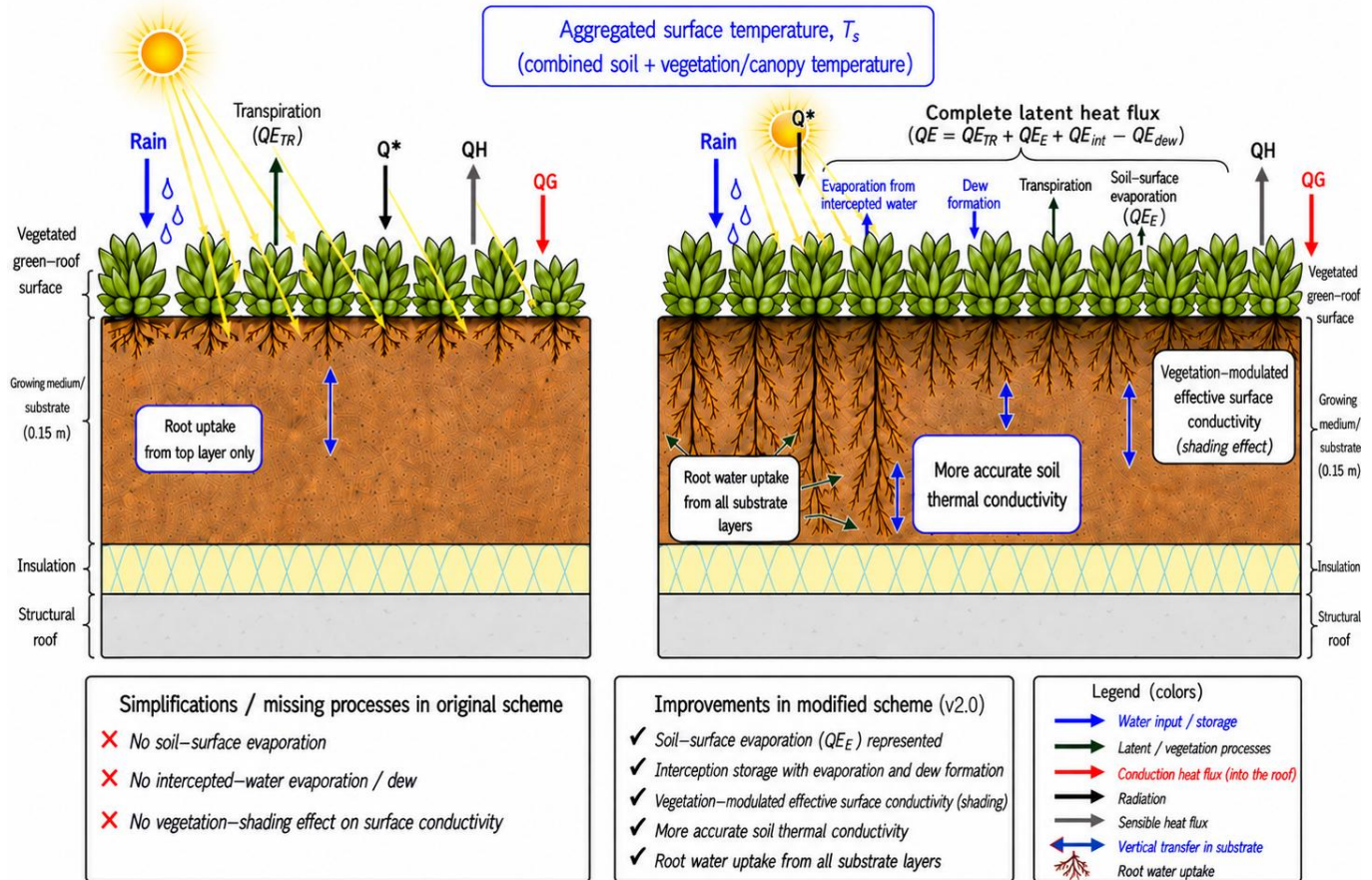


Figure 1. Schematic comparison of the original and modified WRF multi-layer green-roof schemes.

The schematic was included specifically to make the five process changes easier to follow: revised soil thermal conductivity and vegetation-dependent surface conductivity mainly affect  $Q_G$  and heat storage, while soil evaporation, multi-layer root uptake, and canopy interception mainly affect soil water and  $Q_E$ .

## Reviewer Comment 2

**Reviewer comment:** Some technical details regarding the module improvement: 1) How many layers of soil are below the green vegetation for the roof? 2) Is it possible to show the soil layer temperature and moisture? 3) Does the indoor temperature influence the simulation, especially the storage heat? How does the current model deal with this?

### Response:

Thank you for these technical questions. We agree that the model layers and the lower boundary needed to be explained more clearly. The offline model has 10 thermal layers in total. The lower four layers represent the structural roof and have a total thickness of 0.15 m. Above them, six soil layers represent the green roof and also have a total thickness of 0.15 m. Temperature is calculated in all 10 layers, while soil water content is calculated only in the six green-roof soil layers.

In the offline simulation, indoor temperature is not calculated. Instead, the model uses a fixed temperature of 23 °C at the bottom of the roof. The fixed indoor temperature of 23 °C was selected as a representative thermal comfort setpoint for occupied buildings. Indoor temperatures in conditioned buildings are typically controlled near this value to maintain

human thermal comfort. Therefore, 23°C was used as a constant lower-boundary temperature in the offline simulations. We also added the temperature and soil water content of the individual green roof soil layers as Figs. S2 and S3 in the Supplement.

**Changes made in the manuscript and Supplement:**

We revised Sect. 2.2 (pages 4–5, lines 120–129) as follows:

*“The WRF multi-layer green-roof scheme represents a vegetated roof as a one-dimensional (vertical) column of soil (growing medium) placed on top of the structural roof. This column simulates the coupled heat and water balances within the green-roof system. Following the formulations of de Munck et al. (2013) and Gutierrez (2015), as summarized by Zonato et al. (2021), the offline WRF-MLGR configuration contains 10 thermal layers in total. Four layers represent the underlying roof construction, while six layers represent the 0.15 m green-roof soil. Temperature is calculated in all thermal layers, while volumetric water content is calculated for the six soil layers. Layer-resolved soil temperature and moisture diagnostics are provided in Figs. S2 and S3 of the Supplement.”*

*“During the offline evaluation, the indoor temperature is not simulated. Instead, the model uses a fixed temperature of 23 °C at the bottom of the green-roof system, representing a typical indoor/building temperature and providing the lower thermal boundary condition for the offline simulations. This temperature affects the amount of heat transferred between the green roof and the underlying roof.”*

We added a new Sect. S2 to the Supplement (Supplement, page 2) and two new figures. For clarity, the six soil layers in the plots are numbered from the soil surface downward, with Soil layer 1 as the uppermost layer and Soil layer 6 as the deepest layer.

*“The offline WRF-MLGR setup used in this study has 10 thermal layers in total. The lower four layers represent the underlying roof construction, while the upper six layers represent the 0.15 m green-roof soil. Temperature is calculated for all thermal layers, while volumetric water content is calculated for the six soil layers. To illustrate the vertical variation within the soil profile, Fig. S2 shows the mean diurnal temperature cycle of the six soil layers during summer 2014, while Fig. S3 shows the daily mean volumetric water content of the six soil layers over the same period. For clarity, the six soil layers in Figs. S2 and S3 are numbered from the soil surface downward, with Soil layer 1 representing the uppermost soil layer and Soil layer 6 the deepest soil layer.”*

The two new supplementary figures are reproduced below for clarity:

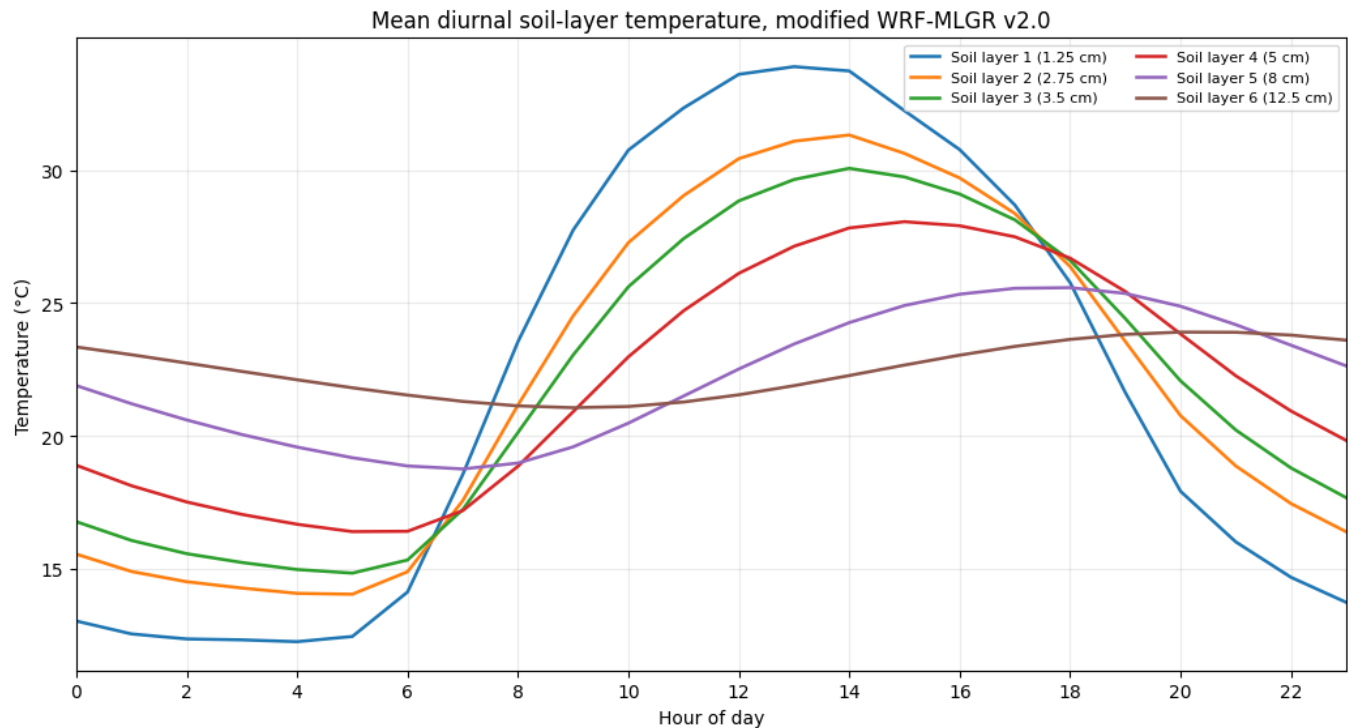


**Figure S2.** Mean diurnal cycle of temperature for the six green-roof soil layers in the modified WRF-MLGR v2.0 simulation during summer 2014. Soil layers are numbered from the soil surface downward (Soil layer 1 = uppermost soil layer), and labels also give the center depth below the soil surface.

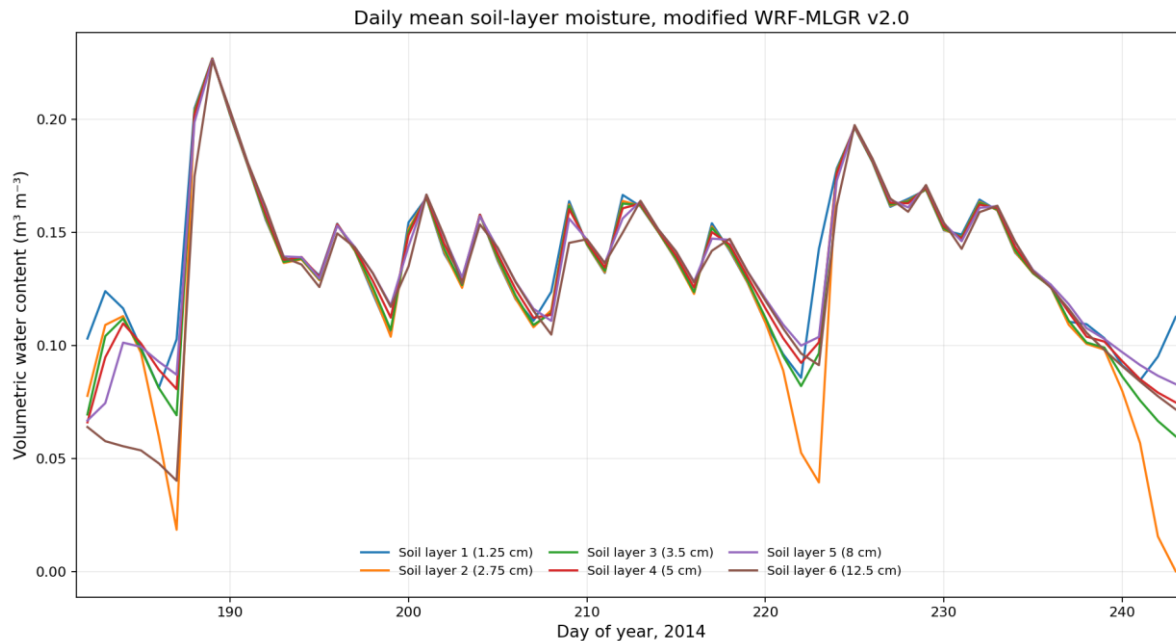


**Figure S3.** Daily mean volumetric water content of the six green-roof soil layers in the modified WRF-MLGR v2.0 simulation during summer 2014. Soil layers are numbered from the soil surface downward (Soil layer 1 = uppermost soil layer), and labels also give the approximate center depth below the soil surface.

### Reviewer Comment 3

**Reviewer comment:** A clarification is needed for the reason for the included validation variables. Why are the sensible heat flux and near-surface 2-m air temperature not included? Is it because of the limits of the field measurement dataset? Since the field measurement description is very limited in the current manuscript, I suggest adding a bit more detail, along with the previous reference; this would be more rigorous and comprehensive for the current focus.

#### Response:

Thank you for this comment. We agree that the choice of validation variables and the limitations of the field measurements should be explained more clearly.

$Q_E$ ,  $Q_G$ ,  $T_s$ , and soil water content were selected as the main validation variables because they were directly measured and are closely related to the processes modified in this study. The 2-m air temperature was also measured, but in the offline model it is used as forcing data. Because the green-roof model was separated from the full WRF atmospheric model for this evaluation, it does not calculate its own 2-m air temperature. Therefore, the observed 2-m temperature is an input to the model and cannot also be used as an independent validation variable.

Sensible heat flux ( $Q_H$ ) was not measured directly. Instead, it was calculated as a residual of the surface energy balance, using net radiation,  $Q_E$ , and  $Q_G$ . Therefore, any uncertainty or error in these measured fluxes, especially  $Q_G$ , also affects the calculated  $Q_H$ . As discussed in the manuscript, the 2014 green-roof modules were located on an elevated platform with an air space below them, and the individual modules could also lose heat through their sides. This makes the  $Q_G$  measurements, especially at night, more difficult to interpret. For this reason,  $Q_H$  is used mainly to discuss changes in energy partitioning rather than as an independent validation variable.

We have clarified these points in the manuscript. Additional details about the field instruments, measurement setup, data processing, and quality-control procedures are provided in Martinez et al. (2026), which we cite for the complete description of the observational dataset.

#### Clarifications in the manuscript:

The field setup and data processing are described in Sect. 2.1.1 (pages 3–4, lines 81–96), including the instrumented green-roof array, Sedum vegetation, 0.15 m soil depth, quality control, hourly processing, and the reference to Martinez et al. (2026). The revised text includes:

*“Raw measurements were quality controlled, and time periods affected by instrument malfunction, maintenance, calibration events, or missing data were excluded. For this paper, both the observational time series and the corresponding model output were processed to hourly resolution (hourly means for state variables and fluxes, and hourly totals for precipitation). Further details on the site instrumentation, observational processing, and quality-control procedures are provided in Martinez et al. (2026); readers are referred there for a complete description of the observational dataset.”*

Sect. 2.1.2 (page 4, lines 101–104) makes clear that near-surface air temperature is prescribed forcing:

*“Atmospheric forcing data were prescribed from observations collected on the roof at, or near, the green-roof array. The forcing time series includes incoming shortwave radiation ( $W\ m^{-2}$ ), incoming longwave radiation ( $W\ m^{-2}$ ), near-surface air temperature ( $^{\circ}C$ ), relative humidity (%), wind speed ( $m\ s^{-1}$ ), precipitation (mm; hourly accumulation), and surface air pressure (Pa).”*

The limitation of  $Q_H$  as an independent validation variable is stated in Sect. 4.2 (page 26, lines 564–569):

*“A related limitation is that the “observed” sensible heat flux ( $Q_H$ ) is not measured directly. Instead, it is calculated as a residual: net radiation minus the sum of measured  $Q_E$  and measured  $Q_G$ . This means that any bias in  $Q_G$ , during the day and especially at night—will automatically appear as a bias in the residual  $Q_H$ , which makes nighttime  $Q_H$  comparisons harder to interpret. Importantly, this does not change one of the main conclusion of this study: the modified model dramatically reduces the large  $Q_G$  errors seen in the original scheme. Rather, the residual nature of  $Q_H$  mainly limits how confidently we can interpret the remaining nighttime  $Q_H$  differences as resulting from model physics.”*

#### Reviewer Comment 4

**Reviewer comment:** As mentioned by the authors, the green-roof scheme in the single-layer urban canopy model, I would like to ask how its performance compares to the previous scheme and the current modified scheme. Because the green-roof scheme itself has been quite intensively studied, it is essential to show why and how the current improvements are significant.

#### Response:

Thank you for this important comment. We agree that the difference between the single-layer and multi-layer green-roof schemes needed to be explained more clearly. In response, we added further explanation to the revised manuscript to clarify that these are two different model implementations.

The single-layer and multi-layer green-roof schemes are two different model implementations with different structures and physical processes. The single-layer green-roof scheme was developed for the WRF single-layer urban canopy model and has been evaluated in previous work. In contrast, our study uses the multi-layer green-roof scheme developed by Zonato et al. (2021) for WRF-BEP+BEM. This model represents the green roof and underlying roof using multiple vertical layers and is therefore different from the single-layer scheme.

A direct numerical comparison between the two models would not be fully appropriate because the previous single-layer study used a different site, climate, forcing data, simulation period, and model configuration. Therefore, differences in their reported performance cannot be attributed only to the green-roof model itself. For this reason, we compare their structures and previously reported performance, rather than directly comparing their numerical statistics.

The focus of the present study is the Zonato et al. (2021) multi-layer scheme. This scheme had not previously received comprehensive process-level evaluation and development against long-term field observations. In this study, we

evaluate the original multi-layer scheme against observations, identify its main problems, introduce five process-based modifications, and then evaluate the modified WRF-MLGR v2.0 using exactly the same forcing and observations. We also use the one-at-a-time sensitivity tests in Figs. 8 and 9 to show which of the new processes are responsible for the changes in  $Q_E$  and  $Q_G$ . This direct original-versus-modified comparison provides the main evidence for the importance of the improvements introduced in this study.

### Changes made in the manuscript:

We revised Sect. 4.1 (page 22, lines 447–456) to make this distinction explicit:

*“It is relevant to additionally compare other green-roof models. A separate WRF green-roof scheme has also been developed for the single-layer urban canopy model (Yang et al. 2015). This scheme is different from the multi-layer WRF-BEP+BEM green-roof scheme of Zonato et al. (2021) evaluated and modified in the present study. The single-layer scheme does not document a dedicated interception storage and evaporation process; however, it was shown to reproduce surface temperature and turbulent energy fluxes within a reasonable range when evaluated against observations over a 6-day period. Another approach is that the TEB-GREENROOF model directly couples the vegetated substrate to the roof layers without a canopy air space. When evaluated, it showed underestimated substrate moisture, overestimated drainage, and larger than observed temperature amplitudes, reflecting a warm bias and discrepancies in moisture dynamics (de Munck et al., 2013). In the following, we assess the behavior of the modified WRF multi-layer green roof scheme and its strengths, weaknesses, and the impacts of the applied updates.”*

The following paragraphs in Sect. 4.1 then quantify the improvement of the modified multi-layer scheme relative to its original version using the same observations (e.g.,  $Q_G$  RMSE decreases from 105.9 to 24.0 W m<sup>-2</sup> in summer and from 94.2 to 24.0 W m<sup>-2</sup> in fall; Table 1).

### Reviewer Comment 5

**Reviewer comment:** The figures shown in the current manuscript are quite unclear, which may be due to the resolution of the figures. In addition, in both Figures 2 and 3b, there are grey lines underlying the figures; what do those lines represent?

#### Response:

Thank you for this helpful comment. We agree. All manuscript figures have been replaced with higher-resolution versions, with clearer axes, legends, labels, and text. We also clarified the meaning of the gray/light traces in the figure captions. In both Figures 3 and 4b, the gray lines are observed  $Q_E$  during rainy days.

### Changes made in the manuscript:

The Fig. 3 caption was clarified (page 14, around lines 341–342) to state:

*“In panel (b), gray lines indicate observed latent heat flux during rainy days.”*

In addition, all figures throughout the revised manuscript were regenerated/replaced at higher resolution.

### Reviewer Comment 6

**Reviewer comment:** For the results, I have one major question. As shown by the current results, the storage heat is largely improved; the error improved from around 100W/m<sup>2</sup> to around 20 W/m<sup>2</sup>, but the surface temperature results are comparable. This means the energy partitioning in the previous and current/modified modules is quite different. This part of the results needs further and deeper explanation and discussion. I suggest a comprehensive analysis of each energy partitioning and an explanation of the reasons/possible reasons behind it. Figure. 6 has addressed this partially, but it is still a bit superficial and only describes the differences.

#### Response:

Thank you for this important comment. We agree that the previous discussion was too descriptive. We expanded the energy-partitioning discussion to explain why  $Q_G$  improves strongly while  $T_s$  changes much less. Net radiation ( $Q^*$ ) is very similar in the two simulations, so reducing excessive  $Q_G$  leaves more energy for  $Q_E$  and  $Q_H$ .  $Q_G$  can change

substantially without a similar change in  $T_s$  because it depends on both the soil temperature difference and the soil thermal conductivity. We also explain how soil evaporation, canopy interception, and multi-layer root uptake affect  $Q_E$  and therefore the remaining energy available for  $Q_H$ . The energy-partitioning figure referred to as Fig. 6 by the reviewer is now Fig. 7 after figure renumbering.

### Changes made in the manuscript:

We expanded Sect. 4.1 (pages 23–24, lines 482–498) with the following explanation:

*“Reducing the unrealistic daytime  $Q_G$  peak changes how the available energy is divided between the different heat fluxes. Net radiation ( $Q^*$ ) is very similar in the original and modified simulations. Therefore, when  $Q_G$  becomes smaller in the modified model, more energy is available for  $Q_E$  and  $Q_H$ .”*

*“The main reason for the large change in  $Q_G$  is the improved soil thermal conductivity. Ground heat flux depends on both the temperature difference within the soil and the soil thermal conductivity. Therefore,  $Q_G$  can change a lot even if the modeled surface temperature does not change very much. In the original model, the soil thermal conductivity became unrealistically high in the morning, which caused too much heat to move into the soil. In the modified model, the revised soil thermal conductivity and especially the vegetation-dependent surface conductivity reduce this excessive heat transfer. As a result, more of the available energy is directed to  $Q_E$  and  $Q_H$ , particularly earlier in the day, shifting the timing of the diurnal cycle of these fluxes.”*

*“The hydrological modifications further influence this coupled energy balance. Soil evaporation provides an additional pathway for water loss from the green roof, canopy interception allows stored rainfall to evaporate rapidly after rain, and multilayer root uptake allows plants to access water from deeper soil layers as the upper soil becomes dry. These processes generally enhance  $Q_E$  when water is available. The resulting changes in  $Q_E$ ,  $Q_G$ , surface temperature, and moisture conditions also affect  $Q_H$ . Therefore, the modified model produces a more realistic overall partitioning and diurnal evolution of the surface energy fluxes, even though changes in surface temperature remain relatively small.”*

*“The hydrological modifications also help explain the changes in latent heat flux ( $Q_E$ ).”*

This expanded discussion is supported by Fig. 7 and by the one-at-a-time sensitivity analysis in Figs. 8–9, which separately shows the effects of the moisture-related and thermal-conductivity modifications.

## References

Martinez, M., Saeedi, A., Voogt, J., and Krayenhoff, S.: Process-based evaluation of green roof models for assessment of heat mitigation efficacy in WRF (v4.3.1) and EnergyPlus (v8.6.0), Geosci. Model Dev., under review, 2026.

We thank the reviewer for these comments and hope the revisions are satisfactory.

# Response to Reviewer #3 Comments

WRF-MLGR v2.0 / EGUsphere Green-Roof Manuscript

*All page and line numbers refer to the revised manuscript submitted with these responses.*

## General Response

We sincerely thank Reviewer #3 for the careful and constructive review. The reviewer identified several important issues in the statistical evaluation, model description, water balance diagnostics, and figure presentation. These comments led us to re-check the observational data used for the  $Q_E$  evaluation, recalculate the affected statistics, clarify the model configuration, add a quantitative water mass balance diagnostic, strengthen the discussion of the experimental limitations, and substantially revise the figures and Results section.

## Reviewer Comment 1

**Reviewer comment:** Please verify the statistical calculations in Table 1. Modified and Original RMSE for  $Q_E$  for the Fall is showing 316.2 and 369.7 corresponding to the MAE values. It is anomalously high in multiple places, while the scatter plots show something else. Please check this.

### Response:

Thank you very much for identifying this issue. We agree that the  $Q_E$  statistics in the original submission were inconsistent, and the reviewer's comment prompted us to audit the complete  $Q_E$  evaluation workflow.

During this re-check, we found two separate data selection mistakes. For the  $Q_E$  scatter plot, we inadvertently used an observational  $Q_E$  file that had been shared within our group for a separate dry period model test, rather than the final observational series used for this manuscript. For Table 1, we used a  $Q_E$  observational series before the final quality control/noise filtering step. These two mistakes explain why the scatter plot and Table 1 did not agree and why the fall  $Q_E$  RMSE values were unrealistically large.

We corrected both problems and regenerated Table 1 and Fig. 2 using the same quality-controlled hourly  $Q_E$  observational dataset. We also rechecked the affected calculations and revised the text so that it reflects the corrected results. The corrected results show that the original and modified models have similar overall  $Q_E$  performance in summer, while the modified model has a lower RMSE and higher  $R^2$  in fall.

### Changes made in the manuscript:

Table 1 was recalculated and corrected on pages 11–12. The revised  $Q_E$  interpretation is given in Sect. 3.1, page 12, lines 290–296. Figure 2 was regenerated using the same observational  $Q_E$  data as Table 1; its revised caption is on page 13, lines 335–339.

“For latent heat flux ( $Q_E$ ), the original and modified models show similar overall performance in summer, with identical RMSE and  $R^2$  values ( $61.8 \text{ W m}^{-2}$  and  $R^2 = 0.40$ , respectively). However, the original model has a slightly lower MAE and higher refined index of agreement ( $d$ ) than the modified model ( $40.4$  vs.  $43.9 \text{ W m}^{-2}$ ;  $d = 0.67$  vs.  $0.64$ ). In fall, the modified model shows a lower RMSE and higher  $R^2$  than the original model ( $59.7$  vs.  $64.7 \text{ W m}^{-2}$ ;  $R^2 = 0.29$  vs.  $0.17$ ), while the original model has a slightly lower MAE and slightly higher  $d$  ( $39.6$  vs.  $41.0 \text{ W m}^{-2}$ ;  $d = 0.60$  vs.  $0.59$ ) (Table 1; Fig. 2e–f).”

The revised Fig. 2 is reproduced below for clarity:

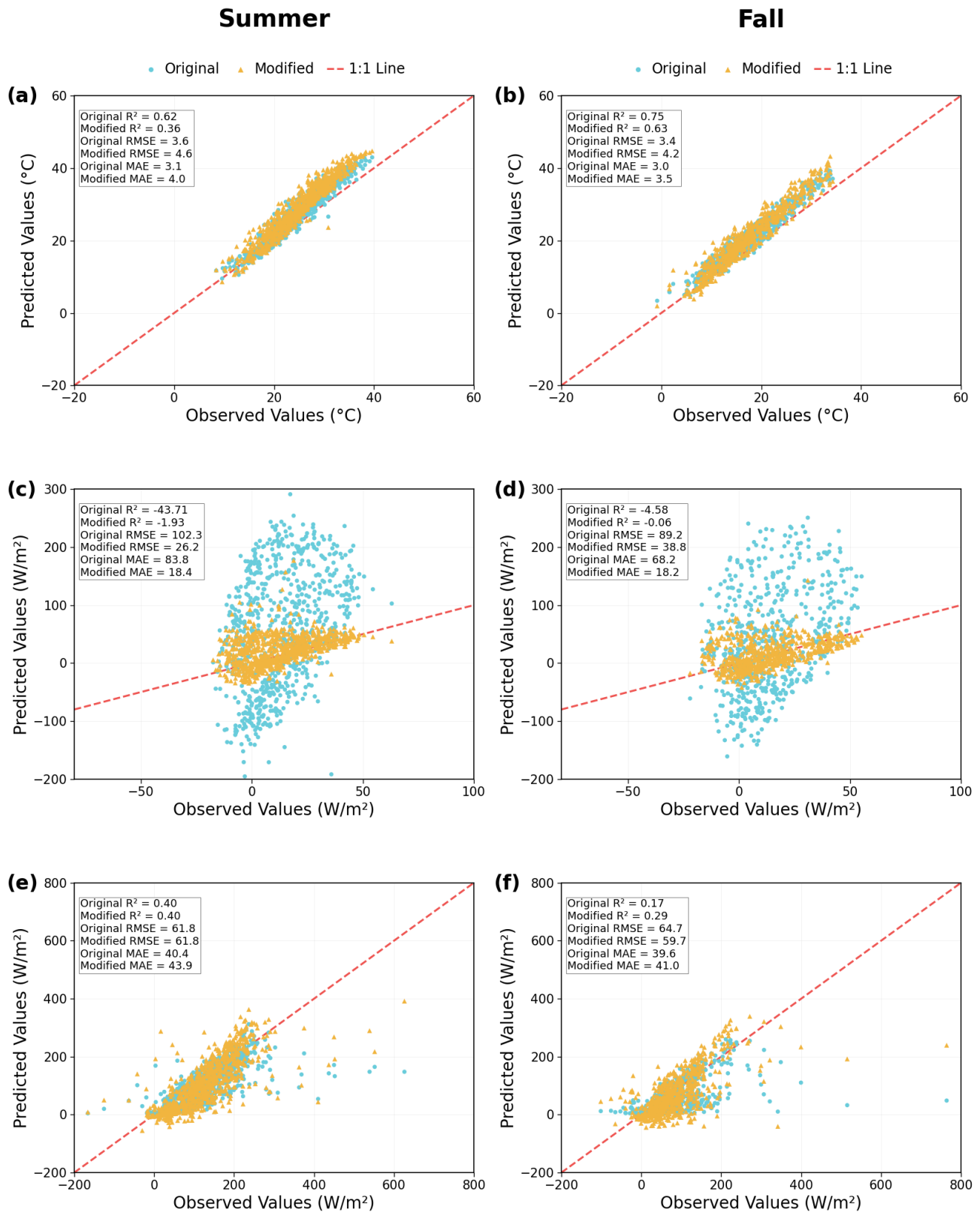


Figure 2. Comparison of the original and modified multi-layer green roof models against observations during July 1 to August 31, 2014 (Summer; left column) and September 1 to October 31, 2014 (Fall; right column). In each panel, the original model is shown with blue circles and the modified model with orange triangles. Panels (a, b) show surface temperature,  $T_s$  (°C); panels (c, d) show soil heat flux,  $Q_G$  ( $W\ m^{-2}$ ); and panels (e, f) show latent heat flux,  $Q_E$  ( $W\ m^{-2}$ ). The dashed red line indicates the 1:1 line, and  $R^2$ , RMSE, and MAE are reported in each panel.

## Reviewer Comment 2

**Reviewer comment:** How can  $R^2$  be negative in Figure 1 (c,d)? Please define  $R^2$ , if it is the squared of the correlation coefficient it cannot be negative.

### Response:

Thank you for pointing this out. The  $R^2$  used in our scatter plots is not the squared Pearson correlation coefficient. We use the residual-based coefficient of determination,  $R^2 = 1 - \Sigma(M_i - O_i)^2 / \Sigma(O_i - \bar{O})^2$ . With this definition,  $R^2$  can be negative when the model error is larger than the error obtained by simply using the observed mean as the predictor. We have now defined this explicitly in the Methods so that the negative values in the  $Q_G$  panels are clear.

### Changes made in the manuscript:

We added the definition to Sect. 2.1.3, page 4, lines 110–117:

“The coefficient of determination ( $R^2$ ) shown in the scatter plots was calculated using the residual-based form  $R^2 = 1 - \Sigma(M_i - O_i)^2 / \Sigma(O_i - \bar{O})^2$  (Onyutha, 2022), where  $M_i$  and  $O_i$  are the modeled and observed values, respectively, and  $\bar{O}$  is the mean of the observations. This metric evaluates model performance relative to the observed mean as a reference predictor; values below zero indicate that the model performs worse than simply using the observed mean (Nash and Sutcliffe, 1970; Onyutha, 2022).”

## Reviewer Comment 3

**Reviewer comment:** The experimental modules have a total substrate depth of 0.15 m (lines 81–84), whereas Section 2.2 describes ten model layers spanning approximately 0.30 m. Section 2.1.2 also states that layer depths were prescribed from site measurements. Please explain the discrepancy in this data, this will have effect on all the results.

### Response:

Thank you for identifying this ambiguity. There is no 0.30 m green roof soil depth in the simulations. The earlier wording did not clearly distinguish the green roof soil layers from the underlying structural roof layers. The model has 10 thermal layers in total: six layers represent the green roof soil and together have a depth of 0.15 m, matching the experimental modules; the remaining four layers represent approximately 0.15 m of underlying roof construction. Therefore, the full thermal column is approximately 0.30 m, but only the upper 0.15 m is green-roof soil. Temperature is calculated in all 10 thermal layers, whereas water content is calculated only in the six soil layers. This is a clarification of the model structure, not a change to the simulations or results.

### Changes made in the manuscript and Supplement:

We revised Sect. 2.2, pages 4–5, lines 120–126, to state the layer structure and depths explicitly.

“Six layers represent the green-roof soil and have a combined depth of 0.15 m, consistent with the observed substrate depth of the experimental modules, while the remaining four layers represent approximately 0.15 m of underlying roof construction. Thus, the complete modeled thermal column is approximately 0.30 m deep, but only the upper 0.15 m represents the green-roof soil. Temperature is calculated in all thermal layers, while volumetric water content is calculated only for the six soil layers.”

## Reviewer Comment 4

**Reviewer comment:** Line 70 there is a mention of energy and mass conservation, however I do not see any diagnostics of mass conservation, perhaps that needs to be included to support the results. The claim of mass conservation needs to be quantified.

### Response:

Thank you. We agree that the statement about mass conservation needed a quantitative diagnostic. We therefore calculated the seasonal water mass balance for both model configurations. The diagnostic accounts for precipitation, irrigation, net evaporative water loss from the modeled green-roof system, drainage, and the change in modeled water storage. Here, net

evaporative water loss represents the total water transferred from the modeled green-roof system through the latent water flux, including evaporation and transpiration, with dew formation treated as a water input to the system. No irrigation was applied in these simulations.

The original model has water-balance closure errors of 1.95% of precipitation in summer and 1.31% in fall. The modified model has much smaller closure errors of 0.13% and 0.15%, respectively. Thus, both configurations are close to seasonal closure, and the modified model closes to within 0.2% of precipitation in both periods.

### Changes made in the manuscript and Supplement:

We added the water-balance result to Sect. 3.2, pages 19–20, lines 405–419. We also added Sect. S3 and Table S1 to the Supplement (Supplement, page 3), where the full balance and residual are reported.

“As an additional conservation diagnostic, the seasonal water-balance closure errors were 1.95 % and 1.31 % for the original model in summer and fall, respectively, compared with only 0.13 % and 0.15 % for the modified model, after accounting for precipitation, net atmospheric water loss, drainage, and changes in water storage (Sect. S3 and Table S1).”

## Reviewer Comment 5

**Reviewer comment:** Eqn (6) explain the importance of  $\theta_{cap}$  and its implications. Because as  $\theta_1 = \theta_{cap}$  the value suddenly/abruptly jumps to 1. This introduces an abrupt discontinuity. Is this intentional, and could it affect numerical behavior or simulated evaporation?

### Response:

Thank you for identifying this important issue. The reviewer is correct that Eq. (6) as written in the original manuscript would introduce an abrupt jump at  $\theta_1 = \theta_{cap}$ . During our re-check, we found that this was a manuscript transcription error. We had inadvertently reproduced the original Viterbo and Beljaars (1995) expression containing the factor 1.6, but that factor is not used in the WRF-MLGR v2.0 code used for the simulations in this study.

In the implementation used here, the wetness factor is  $\alpha(\theta_1) = 0.5[1 - \cos(\pi\theta_1/\theta_{cap})]$  for  $\theta_1 < \theta_{cap}$  and  $\alpha = 1$  for  $\theta_1 \geq \theta_{cap}$ . Therefore,  $\alpha$  approaches 1 continuously as  $\theta_1$  approaches  $\theta_{cap}$ , and there is no abrupt jump in the substrate wetness factor or in the soil evaporation term at the threshold. The factor 1.6 in the original Viterbo and Beljaars formulation was associated with their treatment of moisture averaged over a thicker upper soil layer; our upper green roof soil layer is thinner and its water content is used directly. This correction changes only the manuscript description; the simulations and reported results are unchanged.

### Changes made in the manuscript:

We corrected Eq. (6) and its explanation in Sect. 2.3.3, page 9, lines 230–239, so that the equation shown in the manuscript matches the code used in the simulations. The revised explanation states that the wetness factor is continuous at  $\theta_1 = \theta_{cap}$  and explains the role of  $\theta_{cap}$ .

“ $\alpha(\theta_1) = 0.5[1 - \cos(\pi\theta_1/\theta_{cap})]$  for  $\theta_1 < \theta_{cap}$ , and  $\alpha(\theta_1) = 1$  for  $\theta_1 \geq \theta_{cap}$ . Here  $\theta_{cap} = 0.323 \text{ m}^3 \text{ m}^{-3}$  is the soil moisture threshold used in the model. With this formulation,  $\alpha$  approaches 1 continuously as  $\theta_1$  approaches  $\theta_{cap}$ , so no discontinuity is introduced into the simulated soil evaporation.”

## Reviewer Comment 6

**Reviewer comment:** Figure 4, why is the observed  $Q_E$  not added in this figure? Please add that and discuss the implications.

### Response:

Thank you for this suggestion. We agree that the observed  $Q_E$  should be shown directly. After the addition and renumbering of figures, the event scale figure referred to by the reviewer is now Fig. 5. We added the lysimeter-derived observed  $Q_E$  as a black line with markers in both panels and expanded the text to discuss the comparison.

### Changes made in the manuscript:

We expanded the event discussion in Sect. 3.1, page 13, lines 324–334, and revised Fig. 5 and its caption on pages 16–17, lines 355–359.

“The lysimeter-derived observed  $Q_E$  also increases rapidly following rainfall and shows close temporal agreement with the modeled response, with only a small difference in the timing of the peak. In fact, observed  $Q_E$  is derived from changes in lysimeter mass over each measurement interval, which can contribute to small differences in the apparent timing of the responses.”

The revised Fig. 5 is reproduced below for clarity:

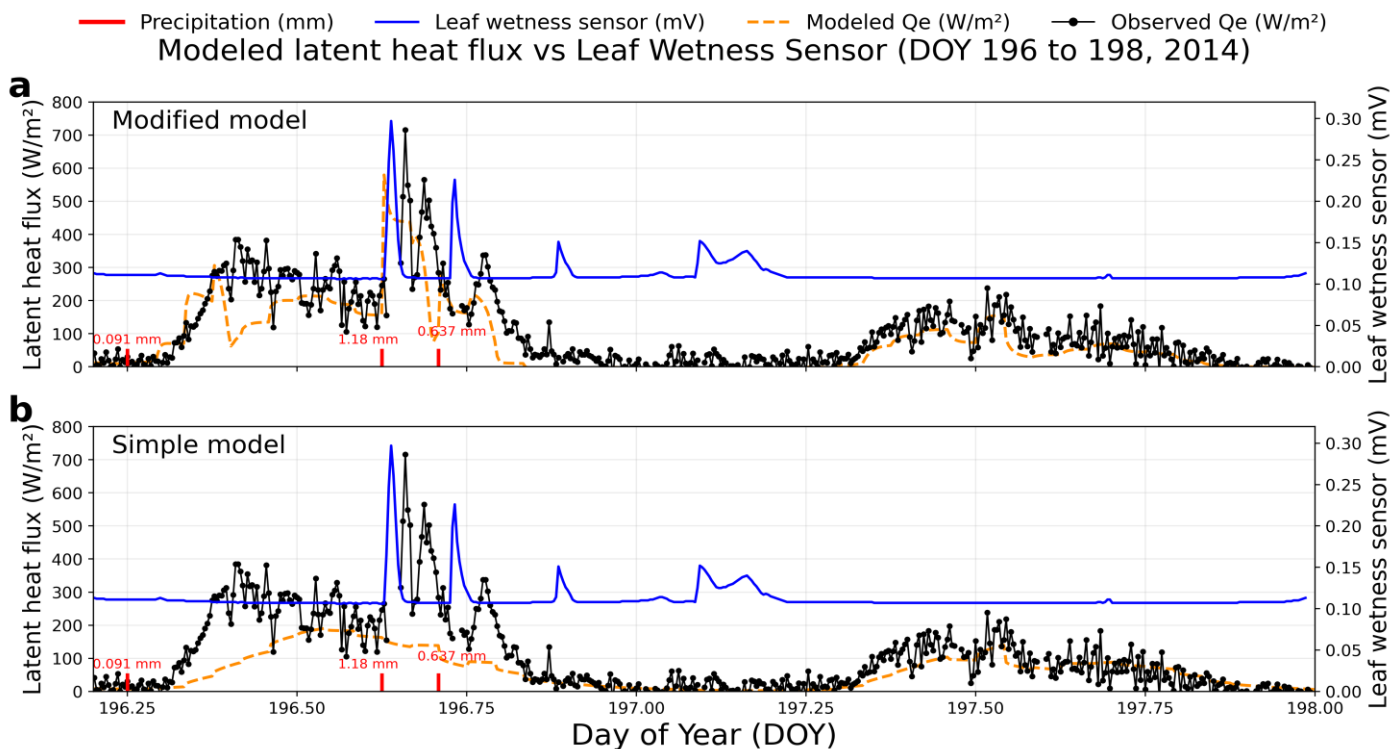


**Figure 5.** Event-scale comparison of modeled latent heat flux (left axis,  $W m^{-2}$ ) with leaf wetness sensor signal at the sedum surface (right axis, mV) and precipitation (red bars, mm) for DOY 196–198 (15–17 July 2014). The leaf wetness sensor reports the raw voltage output (mV) from the probe; higher values indicate a wetter surface. Modeled  $Q_E$  is shown in dashed orange for (a) modified model and (b) original model. Observed  $Q_E$  is shown as the solid black line with markers.

### Reviewer Comment 7

**Reviewer comment:** Simulation details are missing especially WRF model version and details, whether it is fully online atmosphere coupled simulation or offline simulation.

### Response:

Thank you. We agree that these simulation details were missing. We have now stated the WRF version, the urban canopy configuration, and whether the evaluation was online or offline. The simulations use the multi-layer green roof scheme from WRF v4.3.3 with BEP+BEM in offline mode. The full atmospheric WRF model was not run; instead, the green roof model was driven directly by observed meteorological forcing.

### Changes made in the manuscript:

We added this information in Sect. 2.1.2, page 4, lines 97–105:

“The simulations were performed using the multi-layer green roof scheme from WRF v4.3.3 with BEP+BEM in offline mode. The full atmospheric WRF model was not run. Instead, the green roof model was driven directly by observed meteorological forcing.”

## Reviewer Comment 8

**Reviewer comment:** Line 245: Discussion about amplitude and timing and peak would require referring to actual time series of the figures rather than the scatter plots. I suggest presenting time series first and then present the scatter plot i.e. make Figure 2 as Figure 1. Also Figure 2 and Figure 3 could simply be merged by using different line styles. 3 lines in a single plot is not really a lot. This would make it easier for the reader to easily compare the old and the new method than scrolling down to see if there is a mismatch in the timing of the amplitude etc.

### Response:

Thank you for this helpful suggestion. We agree with the main point that timing and peak behavior should be evaluated from time-series or diurnal information, not from scatter plots. We therefore removed the earlier timing/peak interpretation from the scatter plot discussion. The revised text associated with Fig. 2 now focuses on quantitative performance metrics such as RMSE, MAE,  $R^2$ , and  $d$ , while timing and peak behavior are discussed using the time-series, event-scale, and diurnal figures.

We also tested the reviewer’s suggestion to overlay the observed, original model, and modified model time series in the same panels. This direct comparison is useful, but for several variables the curves overlap strongly. We therefore retained Figs. 3 and 4 separately in the main manuscript for readability and added the overlaid comparison as Fig. S4. This gives the reader both a clean individual view and a direct same-panel comparison. We kept the scatter plot before the time series as a compact statistical summary, but it is no longer used as evidence for timing or peak behavior.

### Changes made in the manuscript and Supplement:

The revised statistical discussion is in Sect. 3.1, page 12, lines 290–296. The direct-comparison explanation is on page 13, lines 315–320. Figs. 3 and 4 are retained separately on pages 14–15. We added Sect. S4 and Fig. S4 to the Supplement (Supplement, page 4).

“For direct visual comparison of the timing and magnitude of the original and modified model responses, the corresponding time series are overlaid in Fig. S4 of the Supplement.

## Reviewer Comment 9

**Reviewer comment:** Figure 1: Scatter plot instead of plotting summer and fall images together. Considering plotting original and modified methods for summer in the left column and fall in the right column.

### Response:

Thank you. We reorganized the scatter plots as suggested. Summer is now shown in the left column and fall in the right column, and the original and modified model results are plotted together within each seasonal panel. We also use different marker shapes in addition to colors: circles for the original model and triangles for the modified model. This makes direct comparison much easier.

### Changes made in the manuscript:

The reorganized scatter plot is now Fig. 2 because a new model schematic was added as Fig. 1. The revised Fig. 2 caption is on page 13, lines 335–339. The revised Fig. 2 is reproduced above under Reviewer Comment 1.

## Reviewer Comment 10

**Reviewer comment:** Line 265 -- 270: So does that mean the original model was doing a good job in terms of soil water content even with simplifications? Discuss the reasons for this.

**Response:**

Thank you for this observation. The original model does reproduce the bulk soil water content reasonably well, especially in summer, despite its simpler hydrological representation. We have added an explanation so that this result is not over-interpreted.

Both model configurations are driven by the same precipitation forcing and start from same initial soil moisture conditions, which strongly constrain the seasonal evolution of storage. In addition, compensating differences among water balance components can produce similar soil-water storage even when the underlying flux partitioning is different. For example, in summer the original model has lower atmospheric water loss but higher drainage than the modified model, so these differences partly offset each other. The modified model also has much better seasonal water balance closure. Therefore, the slightly better summer soil water content statistics of the original model do not necessarily mean that its simpler hydrology is more physically realistic. We also note that this advantage is not consistent across seasons: the modified model performs slightly better in fall.

**Changes made in the manuscript and Supplement:**

We revised the soil-water-content discussion in Sect. 3.1, page 12, lines 307–313. We added the detailed water balance interpretation in Sect. 3.2, pages 19–20, lines 413–419, supported by Sect. S3 and Table S1 of the Supplement (Supplement, page 3).

“The relatively good summer performance of the original model does not necessarily indicate that its simpler hydrological representation is more physically realistic, because compensating differences in atmospheric water loss and drainage reduce overall errors. This behavior is examined further in Sect. 3.2 based on the seasonal water-balance analysis presented in Sect. S3 and Table S1.”

**Reviewer Comment 11**

**Reviewer comment:** Line 270--275: Comparing the timing and peaks of different models would require that they be plotted in the same figure.

**Response:**

Thank you. We agree that direct comparison of timing and peaks is clearest when the model versions can be viewed together. We therefore prepared a combined summer time-series figure with observations, the original model, and the modified model overlaid in the same panels. Because the lines overlap strongly for several variables, placing this combined figure in the main text reduced readability. We retained the clearer separate time series as Figs. 3 and 4 and added the combined comparison as Fig. S4, with an explicit reference in the main text. This allows readers to compare the model timing directly without making the main figures difficult to read.

**Changes made in the manuscript and Supplement:**

The new reference to the direct comparison is in Sect. 3.1, page 13, lines 315–320. We added Sect. S4 and Fig. S4 to the Supplement (Supplement, page 4).

“For direct visual comparison of the timing and magnitude of the original and modified model responses, the corresponding time series are overlaid in Fig. S4 of the Supplement.”

The new combined comparison is reproduced below for clarity:

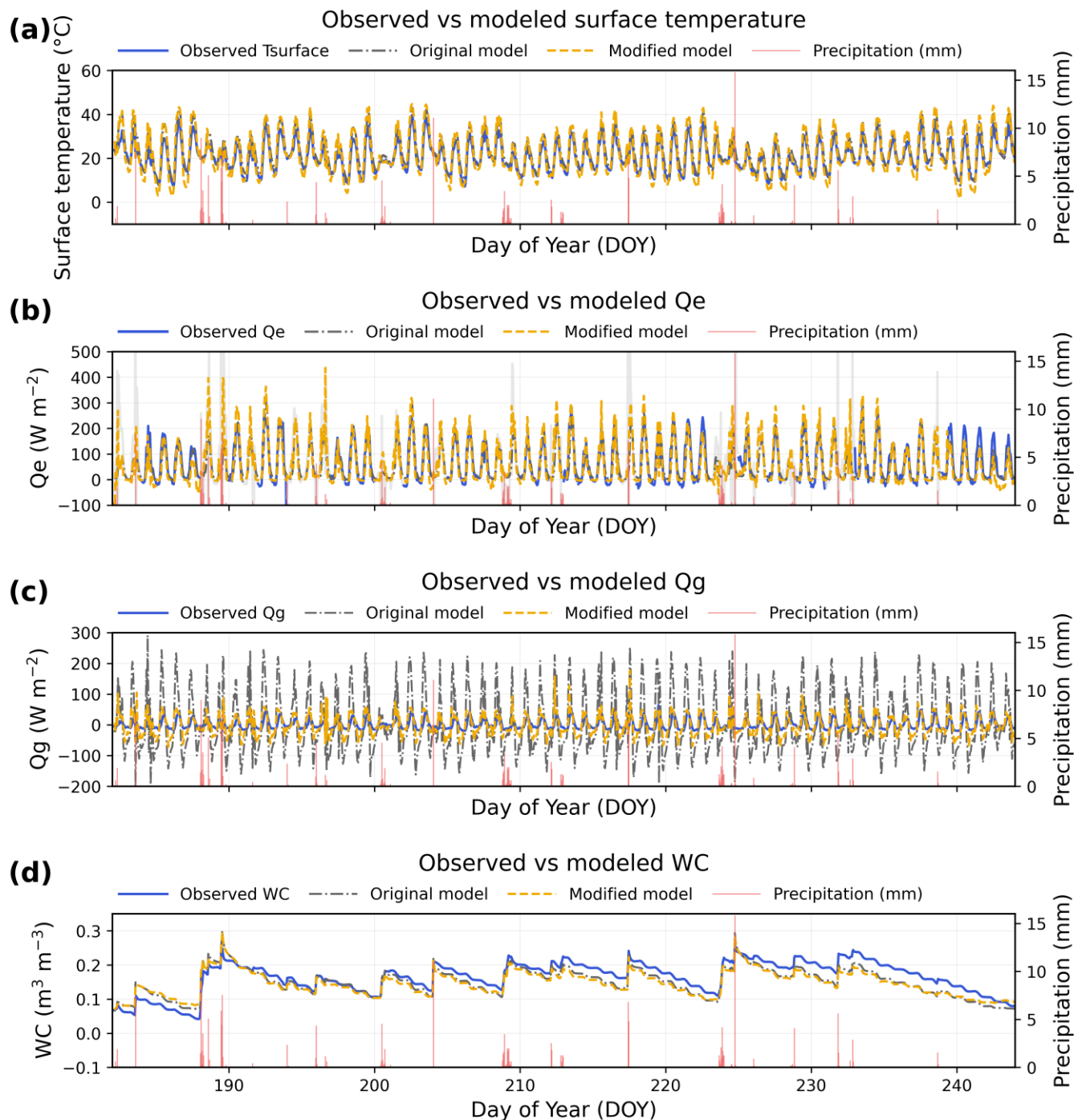


Figure S4. Combined summer 2014 time series of observations, the original WRF multi-layer green-roof model, and the modified WRF-MLGR v2.0 model. The same three series are shown together to allow direct comparison of timing and peaks.

## Reviewer Comment 12

**Reviewer comment:** Line 260--265: I agree that  $Q_G$  is substantially improved, but the modified model still produces pronounced negative nighttime peaks (Fig. 2c), while the observations remain close to zero. Moreover, the air gap and exposed lateral boundaries could influence measurements during both daytime and nighttime. Please discuss their implications across the complete diurnal cycle.

## Response:

Thank you. We agree that the negative nighttime  $Q_G$  peaks remain an important limitation and that the experimental air gap and exposed module sides can influence measurements during the whole diurnal cycle, not only at night. We revised the manuscript to state this explicitly.

The 2014 modules were elevated to accommodate the lysimeters and drainage units, leaving an air space beneath the green roof system, and the individual modules had exposed lateral boundaries. These conditions allow additional heat exchange through the bottom and sides that is not represented in the one-dimensional model whose substrate and roof layers are contiguous. The effect appears strongest at night, when observed  $Q_G$  becomes nearly flat and approaches zero, but daytime  $Q_G$  can also be affected. We therefore report daytime only  $Q_G$  statistics as an additional diagnostic while clearly stating that the daytime observations are not completely free of this limitation. However, during daytime these model-observation differences in flux magnitude are smaller relative to the magnitude of the energy balance forcing (i.e.,  $Q^*$ ) than at night. We also use the October 2025 comparison, when the module was placed directly on the roof deck, to provide a configuration that more closely matches the model lower boundary.

## Changes made in the manuscript:

We revised the Results discussion in Sect. 3.1, page 12, lines 296–302, and expanded the limitation in Sect. 4.2, page 25, lines 553–563.

“The experimental configuration can influence measured  $Q_G$  throughout the full diurnal cycle because the elevated modules leave an air space beneath the green roof system and the individual modules introduce exposed lateral boundaries. These effects appear to be strongest at night due to smaller energy balance forcing (i.e.,  $Q^*$ ) at this time but may also influence daytime measurements.”

## Reviewer Comment 13

**Reviewer comment:** Readability of results section can be further improved by dividing it into sub sections based on energy or mass balance analysis.

## Response:

Thank you for this suggestion. We reorganized the Results section into two subsections so that the energy-related and water-related results are easier to follow.

## Changes made in the manuscript:

Section 3 now contains Sect. 3.1, “Surface energy balance and temperature,” beginning on page 11, line 280, and Sect. 3.2, “Water balance and hydrological processes,” beginning on page 19, line 405. This separates the surface-energy and temperature evaluation from the drainage and water-mass-balance analysis.

## Minor Comments

### Minor Comment 1

**Reviewer comment:** In general image quality is quite low. Please use vector image formats or high DPI images.

## Response:

Thank you. We replaced the figures with higher resolution versions and improved the overall visual quality. Axes, legends, labels, and annotations were also revised for clearer reproduction in the manuscript.

## Changes made in the manuscript:

The revised figures appear throughout the manuscript. Figs. 2–10

## Minor Comment 2

**Reviewer comment:** In some figures (Fig 2 for example), the tick sizes of the figure are small, thumb rule should be that the font sizes in the figures would be approximately same as the text size in the manuscript.

### Response:

Thank you. We increased the tick label, axis label, legend, and annotation font sizes in the revised figures so that they are easier to read and more consistent with the manuscript text.

### Changes made in the manuscript:

This change was applied throughout the revised figures, including Fig. 2 (page 13) and the time-series figures on pages 14–17.

## Minor Comment 3

**Reviewer comment:** Perhaps include the statistical metrics in Figure 1, which would make it easier for the reader to compare the values directly. Mention these metrics under R2.

### Response:

Thank you for the suggestion. We added the main statistical metrics directly to each scatter-plot panel.  $R^2$ , RMSE, and MAE are now shown within each panel so that the original and modified model performance can be compared without moving back and forth between the figure and Table 1.

### Changes made in the manuscript:

The statistical annotations are included in the revised Fig. 2, and the caption explicitly states that  $R^2$ , RMSE, and MAE are reported in each panel (page 13, lines 335–339).

## Minor Comment 4

**Reviewer comment:** Instead of just colours, please use different markers (circles, squares or open circles) to denote summer and fall. Considering the low quality images it is really difficult to differentiate between different points.

### Response:

Thank you. We revised the scatter plot so that the comparison does not depend on color alone. Summer and fall are separated into the left and right columns, respectively, and the two model configurations use different marker shapes: circles for the original model and triangles for the modified model. This improves readability and makes the figure easier to interpret when colors are difficult to distinguish.

### Changes made in the manuscript:

The marker and layout changes are shown in revised Fig. 2; the caption describing them is on page 13, lines 335–339.

## Minor Comment 5

**Reviewer comment:** In general please do a thorough revision w.r.t readability of the figures and presentation of the sections in general.

### Response:

Thank you. We carried out a broad revision of the Results presentation and figures. We reorganized the Results into energy- and water-related subsections, regenerated the figures at higher quality, increased figure text sizes, added statistical metrics directly to the scatter plots, used different marker shapes in addition to colors, reorganized the seasonal

scatter panels, added observed  $Q_E$  to the event-scale figure, and added a direct original versus modified time series comparison. We also revised the associated text and figure references for consistency.

### **Changes made in the manuscript:**

These revisions are visible throughout Sect. 3 (pages 11–21), with the new subsection structure beginning at lines 280 and 405, and in revised Figs. 2–5 and Supplementary Fig. S4.

### **References**

Martinez, M., Saeedi, A., Voogt, J., and Krayenhoff, S.: Process-based evaluation of green roof models for assessment of heat mitigation efficacy in WRF (v4.3.1) and EnergyPlus (v8.6.0), *Geosci. Model Dev.*, under review, 2026.

Onyutha, C.: A hydrological model skill score and revised R-squared, *Hydrol. Res.*, 53, 51–64, <https://doi.org/10.2166/nh.2021.071>, 2022.

Viterbo, P. and Beljaars, A. C. M.: An improved land surface parameterization scheme in the ECMWF model and its validation, *J. Climate*, 8, 2716–2748, [https://doi.org/10.1175/1520-0442\(1995\)008<2716:AILSPS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1995)008<2716:AILSPS>2.0.CO;2), 1995.

Willmott, C. J., Robeson, S. M., and Matsuura, K.: A refined index of model performance, *Int. J. Climatol.*, 32, 2088–2094, <https://doi.org/10.1002/joc.2419>, 2012.

**We thank the reviewer again for the careful review and constructive suggestions. We believe these revisions have substantially improved the accuracy, transparency, and readability of the manuscript.**