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 QE 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 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 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 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 W m^{-2}; d = 0.60 vs. 0.59) (Table 1; Fig. 2e–f)."

The revised Fig. 2 is reproduced below for clarity:

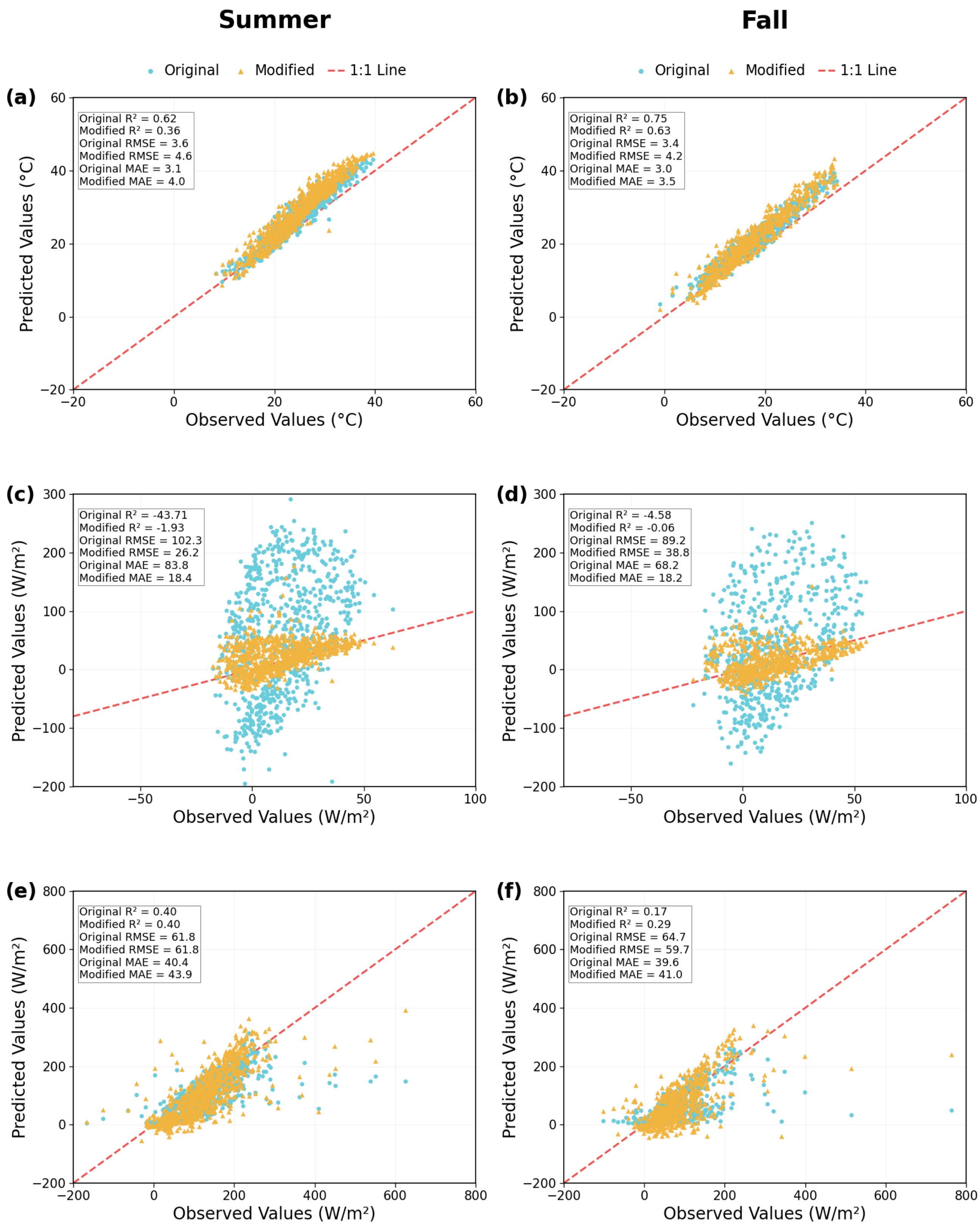

Figure 2. Comparison of the original and modified multi-layer green roof models against observations during July 1 to August 31, 2014 (Summer; left column) and September 1 to October 31, 2014 (Fall; right column). In each panel, the original model is shown with blue circles and the modified model with orange triangles. Panels (a, b) show surface temperature, Ts (°C); panels (c, d) show soil heat flux, QG (W m^{-2}); and panels (e, f) show latent heat flux, QE (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 R2 be negative in Figure 1 (c,d)? Please define R2, 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, $R^2 = 1 - \Sigma(Mi - Oi)^2 / \Sigma(Oi - \bar{O})^2$. 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 θcap 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 $\alpha(\theta 1) = 0.5[1 - \cos(\pi\theta 1/\theta\text{cap})]$ for $\theta 1 < \theta$cap and $\alpha = 1$ for $\theta 1 \geq \theta$cap. Therefore, α approaches 1 continuously as $\theta 1$ approaches θcap, 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 θcap.

> "$\alpha(\theta 1) = 0.5[1 - \cos(\pi\theta 1/\theta\text{cap})]$ for $\theta 1 < \theta$cap, and $\alpha(\theta 1) = 1$ for $\theta 1 \geq \theta$cap. Here $\theta\text{cap} = 0.323$ m^3 m^{-3} is the soil moisture threshold used in the model. With this formulation, α approaches 1 continuously as $\theta 1$ approaches θcap, 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:

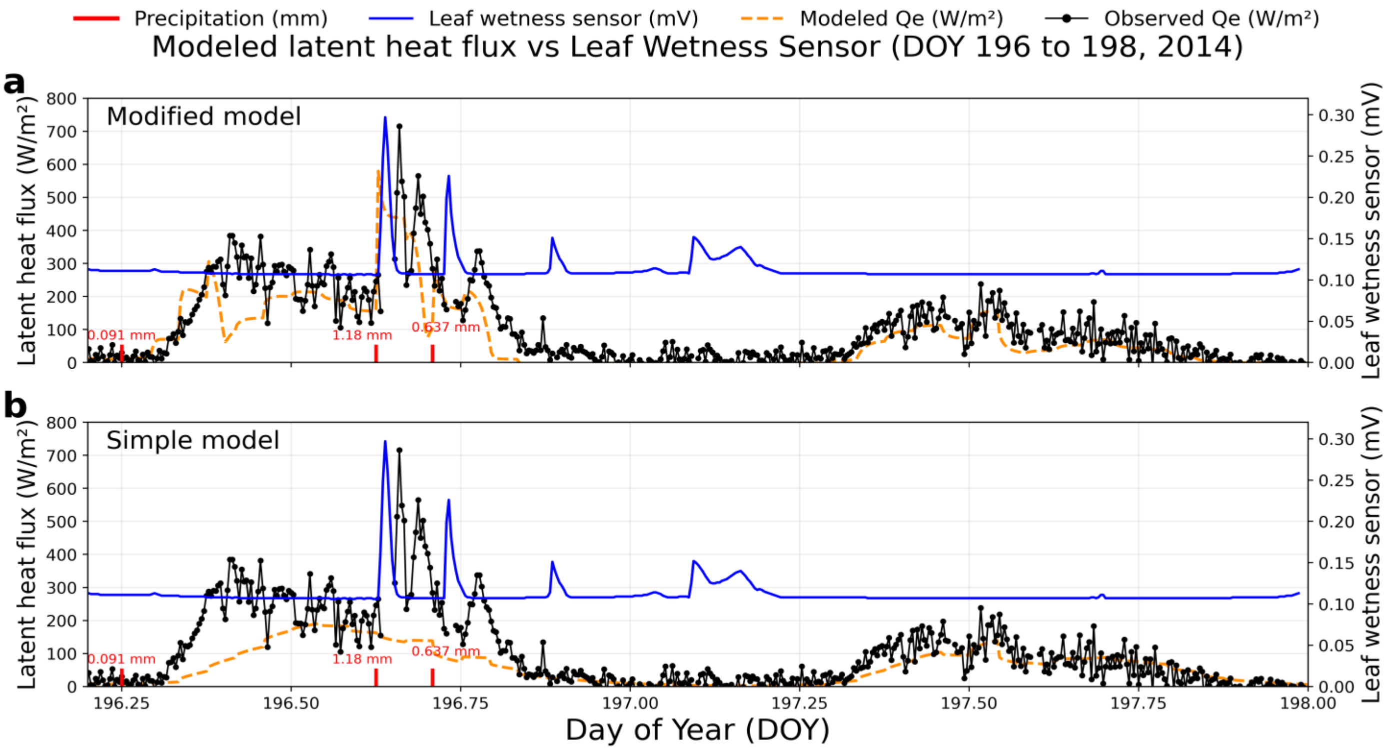

Figure 5. Event-scale comparison of modeled latent heat flux (left axis, W m⁻²) with leaf wetness sensor signal at the sedum surface (right axis, mV) and precipitation (red bars, mm) for DOY 196–198 (15–17 July 2014). The leaf wetness sensor reports the raw voltage output (mV) from the probe; higher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

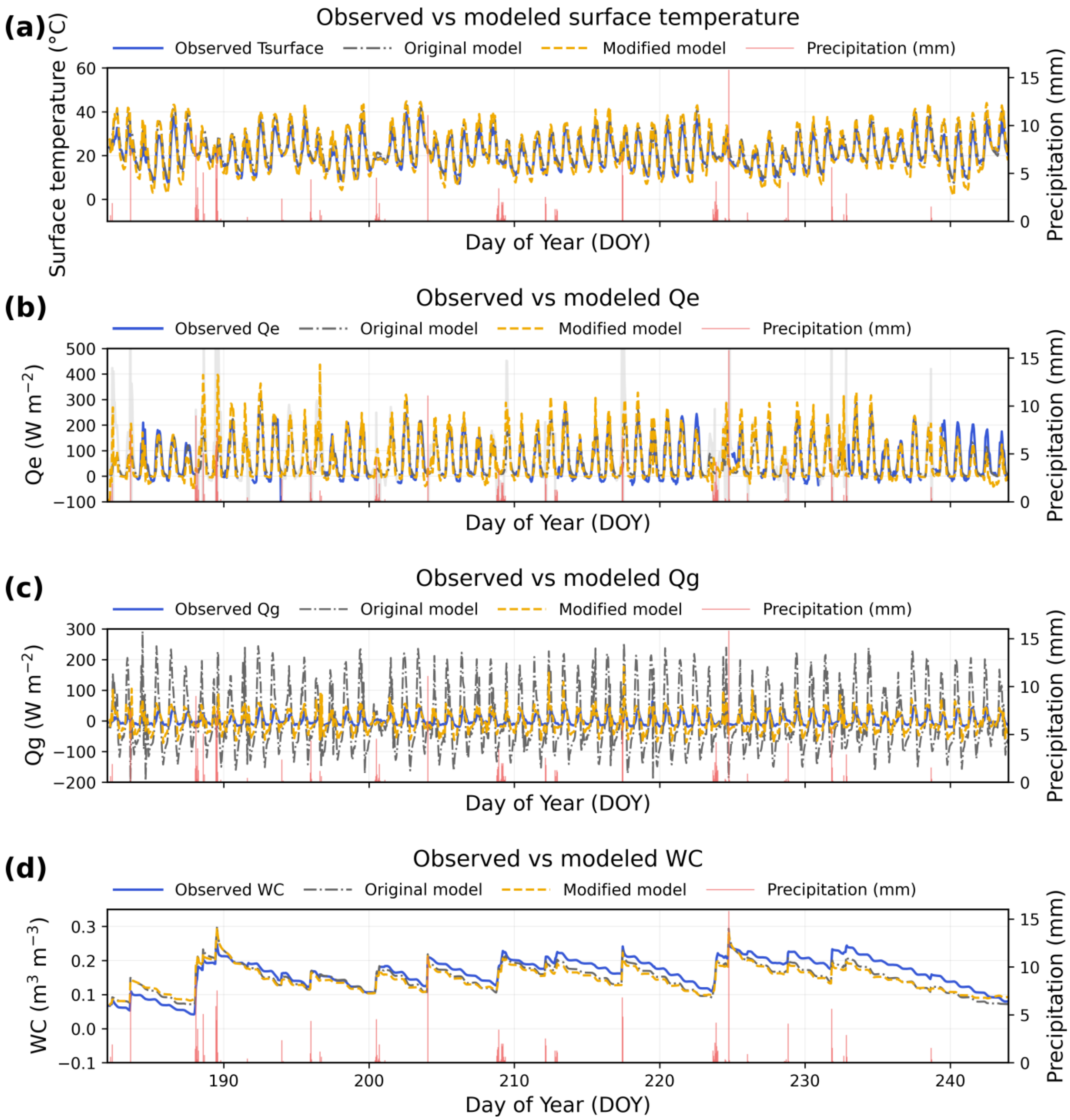

Figure S4. Combined summer 2014 time series of observations, the original WRF multi-layer green-roof model, and the modified WRF-MLGR v2.0 model. The same three series are shown together to allow direct comparison of timing and peaks.

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, 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.