Final Response to Reviewers

Manuscript: Soil moisture as the dominant driver of CO_2 efflux in Mediterranean urban green spaces: evidence for a Gaussian temperature response and mechanistic modelling of moisture and temperature interactions

We thank both referees for their careful and constructive reviews, which have substantially improved the manuscript over the course of the interactive discussion. Below we respond point by point to each comment raised during the open discussion and describe the corresponding changes made to the revised manuscript. Referee comments are shown in blue italics; our responses follow, with the exact confirmed changes to the manuscript text indicated for each point.

Referee 1

We thank the referee for the very positive evaluation of the manuscript and for the constructive suggestions, which strengthen the statistical robustness and interpretability of the proposed model.

Major comments

1. ***Referee comment:*** *The proposed mechanistic model is one of the main strengths of the manuscript. However, its performance would be easier to evaluate if it were quantitatively compared with more conventional approaches (e.g. the classical Q_{10} model). Including model performance metrics (RMSE, AIC, BIC, or cross-validation statistics) would provide stronger evidence that the proposed model offers a genuine improvement rather than simply a more complex fit.*

Response: We have fitted both a standard Q_{10} exponential model and the Gaussian model of Lellei-Kovács et al. (2011) — already cited in the manuscript — to the same dataset, and compared them against our proposed mechanistic model:

Model	AIC	BIC	RMSE	LRT vs. proposed model (p-value)
Q_{10}	4451.04	4465.18	3.602	$< 2.2\times10^{-16}$
Gaussian (Lellei-Kovács et al., 2011)	4365.95	4384.80	3.417	0.017
Proposed mechanistic model	4362.26	4385.83	3.405	— (reference model)

Our proposed model achieved the lowest AIC and RMSE of the three, and a BIC essentially equivalent to the Gaussian model of Lellei-Kovács et al. (2011) (difference < 1.5 points), despite BIC's stronger penalty for additional parameters. Likelihood-ratio tests confirmed that our model significantly reduces deviance relative to both alternatives (Q_{10}: $p < 2.2\times10^{-16}$; Gaussian: $p = 0.017$), providing direct statistical evidence that the improvement in fit reflects genuine explanatory gain rather than simply added flexibility.

We also note that our model is not substantially more complex than the alternatives: it can be written in closed form as $CO_2(SST) = A\cdot\exp(-a\cdot SST - B\cdot\exp(-b\cdot SST))$, which is an algebraically equivalent reparameterisation of Eq. 5 already presented in Section 2.4 (with A, a, B and b corresponding to combinations of C_0, c^T, c^M, M_0 and m^T). This model therefore has 4 free parameters — the same order of complexity as a standard Gaussian model — and these same parameters are already used in Section 2.4 (Eqs. 6–7) to analytically derive the biologically interpretable quantities Tc_{max} and Cc_{max} reported in Table 6, which the purely empirical Q_{10} and Gaussian formulations do not provide. No changes are needed to Section 2.4, as this derivation is already fully presented there; we will instead add the following paragraph and a new Table 7 at the end of Section 3.4, immediately before the Discussion.

Confirmed text (new paragraph, end of Section 3.4, inserted immediately after "...The optimum was at higher SST in AG-H and PH-M, the ones with higher SSM influence." and before "# 4 Discussion"): *"To further evaluate the performance of the proposed mechanistic model, we compared it against a standard Q_{10} exponential model and against the Gaussian model of Lellei-Kovács et al. (2011), fitted to the same dataset (Table 7). The proposed model achieved the lowest AIC and RMSE of the three models, and a BIC virtually indistinguishable from the Gaussian model, despite BIC's stronger penalty for complexity. Likelihood-ratio tests confirmed that the proposed model significantly reduced deviance relative to both the Q_{10} model ($p < 2.2\times10^{-16}$) and the Gaussian model ($p = 0.017$), indicating that the improvement reflects genuine explanatory gain rather than added flexibility."*

Finally, addressing the referee's specific mention of cross-validation statistics, we performed a leave-one-plot-out cross-validation of the proposed mechanistic model and compared it to the original in-sample fit:

	RMSE	MAE	Bias	R^2 (prediction)
Original fit	3.217	2.359	0.007	0.224
Leave-one-plot-out CV	3.287	2.417	0.011	0.189

The cross-validated metrics are very close to the original in-sample fit, indicating that the model is not overfitted and generalises well to plots not used in parameter estimation.

Confirmed text (new paragraph, end of Section 3.4, inserted immediately after the Table 7 paragraph confirmed above and before "# 4 Discussion"): *"To further assess predictive performance, we conducted a leave-one-plot-out cross-validation of the proposed mechanistic model, comparing the cross-validated RMSE, MAE, bias and predictive R^2*

against the original in-sample fit (Table 8). The cross-validated metrics were very close to the in-sample values, indicating that the model generalises well to plots not used in parameter estimation and is not overfitted."

Confirmed table caption (Table 8, placed above the table, in Section 3.4): *"Table 8. Leave-one-plot-out cross-validation of the proposed mechanistic model (Section 2.4), compared against the original in-sample fit. RMSE and MAE in μmol CO_2 m^{-2} s^{-1}. R^2 (prediction): coefficient of determination between observed and cross-validated predicted values."*

Confirmed text (Discussion, Section 4, inserted immediately after "...and is not at all recommended under urban conditions (Wu et al., 2021 highlighted the doubts about Q10 index in a changing climate)." and before "The urban heat island (UHI) effect..."): *"Consistent with this, our own model comparison confirmed that the Q_{10} formulation performed markedly worse than both the Gaussian and the proposed mechanistic model (Table 7)."*

2. ***Referee comment:*** *The identified breakpoint at 18.4 °C is central to the manuscript. It would be helpful to provide additional information on the uncertainty associated with this estimate (e.g. confidence intervals or standard error) and briefly discuss its robustness.*

Response: Agreed. We report the standard error and 95% confidence interval of the breakpoint, obtained via confint.segmented() (R package segmented, Muggeo, 2008): the estimated breakpoint of 18.4 °C has a 95% CI of (17.17, 19.79) °C. This has been added to Sections 2.3 and 3.2.1/3.3, together with a brief discussion of its robustness.

Confirmed text (Section 2.3): *"The standard error and 95% confidence interval of the estimated breakpoint were obtained using the confint.segmented() function of the R segmented package (Muggeo, 2008)."*

Confirmed text (Section 3.2.1, first statistical presentation of the breakpoint, replaces "...regression breakpoint of 18.4°C (SST)..."): *"...a positive correlation up to a regression breakpoint of 18.4°C (95% CI: 17.17–19.79 °C) (SST) and a negative effect above this value..." Subsequent references to the 18.4 °C threshold throughout Section 3.3 (subsection headers, figure captions) are not individually annotated with the CI, as they refer back to this single estimate. A brief discussion of the robustness of this estimate has been added to the Discussion (see our response to comment 6 below).*

3. ***Referee comment:*** *Measurements were collected repeatedly over the same plots during one year. Please clarify whether temporal autocorrelation was assessed or considered during model development, or explain why it was not necessary.*

Response: We agree this distinction is important. The random intercept per plot, already described in Section 2.3 ("To avoid lack of independence between communities within the same plot, a random intercept was included in all models taking the plot as grouping variable"), addresses spatial pseudoreplication (multiple plant communities nested within the same plot) but does not by itself account for temporal autocorrelation of the fortnightly repeated measurements within each plot over the course of the year. We explored this further for the mechanistic model described in Section 2.4.

Confirmed text (Section 2.4, new paragraph after the model description — methodological rationale only): *"The purpose of this analysis was not to estimate population-level or plot-level effects, nor to quantify between-plot variance, but to test whether the theoretically derived functional form could reproduce the observed relationship between soil respiration and soil temperature. To assess whether this conclusion was robust to potential temporal autocorrelation in the repeated measurements, we additionally fitted a nonlinear mixed-effects version of the model, including plot as a random effect and alternative temporal correlation structures (see Section 3.4)."*

Confirmed text (Section 3.4, new paragraph — results of the sensitivity analysis — inserted after the cross-validation paragraph and Table 8 confirmed above, immediately before "# 4 Discussion"): *"As a robustness check, we compared the original fit (Table 6, Fig. 12) against a nonlinear mixed-effects formulation of the same model incorporating alternative temporal correlation structures (Section 2.4). Models incorporating a first-order autoregressive structure (corAR1) converged and yielded parameter estimates broadly consistent with the original fit (Table S2, Fig. S1), with optimum temperatures (Tc_{max}) shifting modestly (+0.5 to +1.5 °C) while preserving the same ranking between plant communities. More complex correlation structures (e.g., corARMA) also converged but were less stable, with parameter estimates more sensitive to model specification. Residual temporal autocorrelation could not be fully eliminated under any structure tested; we consider this likely to reflect shared seasonal environmental drivers acting on soil respiration across all plots simultaneously, rather than a violation of the independence assumption. Because the mixed-effects formulation introduces variance components that are not part of the theoretical relationship being tested, and given the sensitivity of these components to model specification, we regard the mixed-effects results primarily as a sensitivity analysis (Supplementary Material, Table S2, Fig. S1) rather than as an alternative primary estimate; the original model is retained as our main result."*

We verified that the corAR1-adjusted parameter estimates preserve the same ranking of optimum temperatures between communities as the original model (PG-D < AH-D ≈ PH-M < AG-H in both cases), so the interpretive statements in the Discussion regarding PH-M and AG-H remain valid under this sensitivity analysis.

Minor comments

4. *Referee comment:* *I suggest slightly moderating a few statements referring to Mediterranean urban ecosystems in general, as the study was conducted in three urban green spaces within a single metropolitan area.*

Response: Agreed. We have revised the wording throughout Section 6 (Conclusions) to consistently refer to the specific study system (three urban green spaces in Madrid) rather than to Mediterranean urban ecosystems in general, consistent with the caveat already present in Section 5 (Limitations). The following four sentences have been revised:

(a) "Soil respiration in Mediterranean urban green spaces does not follow the conventional Q_{10} exponential model." → "*Soil respiration in the studied Mediterranean urban green spaces does not follow the conventional Q_{10} exponential model.*"
(b) "...suggests that Q_{10}-based approaches may systematically misrepresent soil CO_2 dynamics in Mediterranean cities." → "...suggests that Q_{10}-based approaches may misrepresent soil CO_2 dynamics in comparable Mediterranean urban green spaces."
(c) "This physical coupling between temperature and moisture generates the observed unimodal pattern and must be explicitly incorporated into models of urban soil carbon fluxes." → "...and should be considered when developing models of urban soil carbon fluxes in comparable Mediterranean urban settings."
(d) "These findings collectively highlight the need for moisture-integrated approaches to accurately model soil CO_2 fluxes in Mediterranean urban ecosystems, and underscore the importance of accounting for the drying role of the urban heat island when projecting urban carbon budgets under climate change scenarios." → "These findings collectively highlight the need for moisture-integrated approaches to accurately model soil CO_2 fluxes in Mediterranean urban green spaces such as those studied here, and underscore the importance of accounting for the drying role of the urban heat island when projecting urban carbon budgets, pending validation across a broader range of cities and hydrological years."

5. *Referee comment:* *The manuscript is very well written. I only noticed a few minor typographical and grammatical issues that can easily be corrected during the final revision.*

Response: We have conducted a full proofreading pass of the manuscript, including the typographical correction already identified in response to Referee 1 (Antrosols → Anthrosols), and checked terminology and unit consistency throughout. In addition, we identified two further issues during this pass:

(a) Discussion (Section 4): "...higher emissions than AG-D when the effect of SST was removed." — "AG-D" is not a valid plant community code (the four communities are AH-D, AG-H, PH-M and PG-D) and has been confirmed and corrected to read "*...higher emissions than AH-D when the effect of SST was removed.*"
(b) Discussion (Section 4): "(Chang et al., 2014).This suggests..." — missing space after the period; corrected to "(Chang et al., 2014). This suggests...".

6. *Referee comment:* *Figures are generally of high quality and effectively support the results. The only suggestion would be to ensure that all axis labels and symbols remain easily readable after journal formatting.*

Response: All figures have been reviewed and axis labels, tick labels, legends and statistical symbols (*, **, ***) enlarged where necessary, including Figure 2 (also raised by Referee 1). We additionally reviewed all remaining figures (Figs. 3, 5, 6, 7, 9 and 12) and confirmed that fonts and significance symbols are clearly legible at the current resolution; we will nonetheless verify legibility again at the galley-proof stage, as figures may be further resized during journal typesetting.

Referee 2

We thank the referee for the positive assessment of the study's relevance and for the detailed comments aimed at improving clarity, consistency and accuracy throughout the manuscript.

1. *Referee comment:* *Typo? Antrosols → Anthrosols*

Response: Corrected throughout the manuscript (Section 2.1: "...classified as Antrosols (ATu)" → "...classified as Anthrosols (ATu)"). We checked the rest of the text and found no further instances.

2. *Referee comment:* *Could you provide more details about the study sites? I am curious about the study areas (size) and how the four plant communities were distributed within each study site. How many replicate sampling points do you have?*

Response: The study was conducted across three urban green spaces in Madrid: Ciudad Universitaria, La Elipa and Pinar de Conde Orgaz. Within each site, three replicate sampling points were established for each of the four plant communities (3 replicates × 4 communities = 12 sampling points per site; total n = 36 sampling points across the study). We have added site extent, spatial distribution and replicate information to Sections 2.1 and 2.2, and the geographic coordinates of all sampling points are now provided in the Supplementary Material (Table S1, attached separately). The Table 1 caption has been corrected to n = 3, and the sentence in Section 3.1 referencing a value recorded at a site no longer part of the study ("Casa Campo") has been revised as follows.

Confirmed text (Section 2.1): *"The three study sites (Ciudad Universitaria, La Elipa and Pinar de Conde Orgaz) spanned approximately 1.9, 0.8 and 0.5 km along their longest axis, respectively. Within each site, the four plant communities occurred as naturally distributed ruderal patches rather than in a fixed spatial arrangement, and three replicate sampling points were established within each community's patch."*

Confirmed text (Section 2.2): *"At each of the three study sites, three replicate sampling points were established for each of the four plant communities (3 replicates × 4 communities = 12 sampling points per site; total n = 36 sampling points across the study). The geographic coordinates of all sampling points are provided in the Supplementary Material (Table S1)."*

Confirmed text (Table 1 caption, replaces "n = 4"): *"...(mean ± SE, n = 3) per plant community."*

Confirmed text (Section 3.1, replaces the original sentence referencing "Casa Campo"): *"...which was markedly higher in PG-D communities (53.25 ± 9.09%), reflecting the greater root biomass and organic matter accumulation typically associated with perennial grasslands."*

Note: *if desired, the minimum distance between replicate sampling points within each community can also be reported in Section 2.2; this is optional and does not block the current response.*

3. ***Referee comment:*** *Table captions should be placed above the table.*

Response: Agreed. All table captions (Tables 1–6 and any supplementary tables) have been moved above their corresponding tables, in line with Copernicus/SOIL formatting.

4. ***Referee comment:*** *Gas sampling time is important because temperature varies throughout the day. Could you clarify how the measurements were randomized? Were there differences in measurement timing among plant communities or study sites? Could you also provide the range of measurement times?*

Response: The order in which the three parks were visited was systematically rotated so that each sampling campaign began at a different park, ensuring broad coverage of the diurnal temperature range across sites. Within each park, the order of visiting the 12 sampling points (4 plant communities × 3 replicates) was randomized using a 12-sided die. Measurements were carried out between 08:00 and 18:00–19:00 h local time, depending on fieldwork duration on a given day. Because park visiting order was rotated and point order was randomized, no plant community or study site was systematically associated with a particular time of day.

Confirmed text (Section 2.2): *"To ensure adequate coverage of the diurnal temperature range and avoid systematic timing bias between sites, the order in which the three parks were visited was systematically rotated so that each sampling campaign began at a different park. Within each park, the order of visiting the 12 sampling points (4 plant communities × 3 replicates) was randomized using a 12-sided die. Measurements were carried out between 08:00 and 18:00–19:00 h local time, depending on fieldwork duration on a given day."*

5. ***Referee comment:*** *Could the higher CO2 emissions in PH-M be related to higher temperatures at the time of measurement?*

Response: This is unlikely. Soil and atmospheric temperature did not differ significantly among plant communities (Table 2: $p = 0.82$ for SST-related comparisons among communities), and, as described in our response to comment 4, the visiting order of parks and sampling points was systematically rotated and randomized across a wide daily window (08:00–19:00 h). We have added a clarifying sentence to the Discussion.

Confirmed text (Discussion): *"Because soil and atmospheric temperature did not differ significantly among plant communities (Table 2), and because the visiting order of parks and sampling points was systematically rotated and randomized across a wide daily window (08:00–19:00 h; Section 2.2), the higher emissions observed in PH-M cannot be attributed to a systematic measurement-time temperature bias, but rather point to an intrinsic community-level effect, likely related to soil disturbance and litter quality (see below)."*

6. ***Referee comment:*** *Given the large diurnal temperature variation in this region, it is unclear what the 18.4 °C breakpoint represents. Do the confounding effects of soil temperature and moisture occur mainly at the diurnal scale or at the seasonal scale?*

Response: The breakpoint is derived from a piecewise regression on the full pooled dataset, spanning a complete annual cycle. It therefore predominantly reflects seasonal variation in SST (which ranges roughly from 5 to 35 °C across the year) rather than diurnal variation within a single sampling window. Because sampling on any given day was confined to a defined time window and park visiting order was rotated (see response to comment 4), we interpret the SST–SSM confounding effect described in the Discussion to occur predominantly at the seasonal scale. We have added a clarifying paragraph.

Confirmed text (new paragraph inserted in the Discussion, immediately after the paragraph ending "...may therefore promote an exponential, rather than Gaussian, SR--SSM relationship." and before the paragraph beginning "In cold conditions (SST < 18.4 °C)..."): *"It should be noted that the 18.4 °C breakpoint was estimated from a pooled dataset spanning a full annual cycle, and therefore predominantly reflects seasonal rather than diurnal variation in SST, given the defined daily sampling window and rotated visiting order described in Section 2.2. The narrow 95% confidence interval of this estimate (17.17–19.79 °C) further supports its robustness as a meaningful threshold rather than a statistical artefact."*

7. ***Referee comment:*** *How did you measure the evapotranspiration level (mm day-1; A different unit is used in Figure 2)?*

Response: The referee is correct that there was an inconsistency between the unit reported in the text (mm day^{-1}) and the axis label in Figure 2 (mmol mol^{-1}). The correct unit is mmol mol^{-1}, which is measured directly by the LI-8100 chamber as the water vapor mole fraction at the soil surface; Figure 2 has been corrected accordingly. The unit is already defined once in Section 2.3 ("evapotranspiration (EV – in mmol mol^{-1}) were measured simultaneously with CO_2 using chamber equipment"); we therefore removed the incorrect, redundant unit from the General Linear Models sentence later in the same section rather than repeating it, and corrected the Table 2 legend to match.

Confirmed text (Section 2.3, replaces "...on CO2 and Evapotranspiration levels (mm day-1)."): *"General linear models were developed to assess the effect of ambient variables (AT, SST, AM, SSM and RWi) on CO2 and Evapotranspiration levels."*

Confirmed text (Table 2 legend, replaces "EV: Soil evapotranspiration."): *"EV: Soil evapotranspiration (mmol mol^{-1})."*

8. ***Referee comment:*** *Figure 2: Please increase the font size in the figure.*

Response: Figure 2 has been regenerated with larger axis titles, tick labels and legend text for improved readability.

9. ***Referee comment:*** *Figure 8: The graph on left and right looks the same.*

Response: Both panels of the original Figure 8 were mistakenly labelled "Partial Correlations (SST < 18.4°C)". The left panel, which shows full (Pearson) correlations, has been relabelled "Correlations (SST < 18.4°C)"; the right panel correctly shows partial correlations. The underlying values in the two panels do differ (e.g., the correlation between AM and SSM), but the identical titles made this difficult to notice.