Review of egusphere-2026-4828 - Evaluation of Partitioning Methods to Identify Timescale-Dependent Drivers of Light Use Efficiency in a Wet Tropical Forest

The manuscript presents 11 months (July 2022–May 2023) of eddy-covariance CO2 fluxes, with profile-based storage correction, from a 57 m tower at La Selva, Costa Rica. These are the first EC fluxes reported for the site since Loescher et al. (2003), which is quite valuable given the sparseness of these datasets in tropical regions. The authors compare different NEE partitioning methods: full daytime, morning daytime, afternoon daytime, nighttime and sundown. They show that morning data degrade daytime light-response fits, while the afternoon, nighttime and sundown approaches converge. They report a strong, temporally stable sink (i.e. with almost no seasonality) of about 947 g C m^{-2} yr^{-1} but acknowledge that this value could be influenced by la Niña conditions throughout most of the studiy period. Using afternoon-partitioned GPP, they train random forest models on LUE at half-hourly, daily and weekly resolution. PAR dominates at the half-hourly scale, and temperature and moisture variables gain importance at longer timescales.

The manuscript is well written, and the work is very valuable for the tropical flux community, where recent EC records and systematic comparisons of partitioning methods are scarce. The study is well thought out and scientifically sound. The multi-method comparison is a useful quality-assurance step that other tropical sites could explore. My main concerns are the structure of the paper, how gap-filling interacts with the partitioning methods, and how robust the driver attribution for LUE is. I recommend to accept the paper after minor to moderate revisions.

We thank the reviewer for taking the time to give thoughtful and constructive feedback on this manuscript. We also appreciate the reviewer's comments regarding the structure of the manuscript, the interaction between gap-filling and NEE partitioning, and the interpretation of environmental controls on LUE. We have revised the manuscript to improve the organization. We have also clarified our gap-filling rationale and addressed the potential influence of PAR normalization on the relationships between LUE, temperature, and vapor pressure deficit through an additional sensitivity analysis. Finally, we have refined the abstract and discussion to ensure that our conclusions accurately reflect the findings and limitations of this observational dataset.

Please note that line numbers are for the clean manuscript, not the tracked changes version.

Major points

1. Paper structure

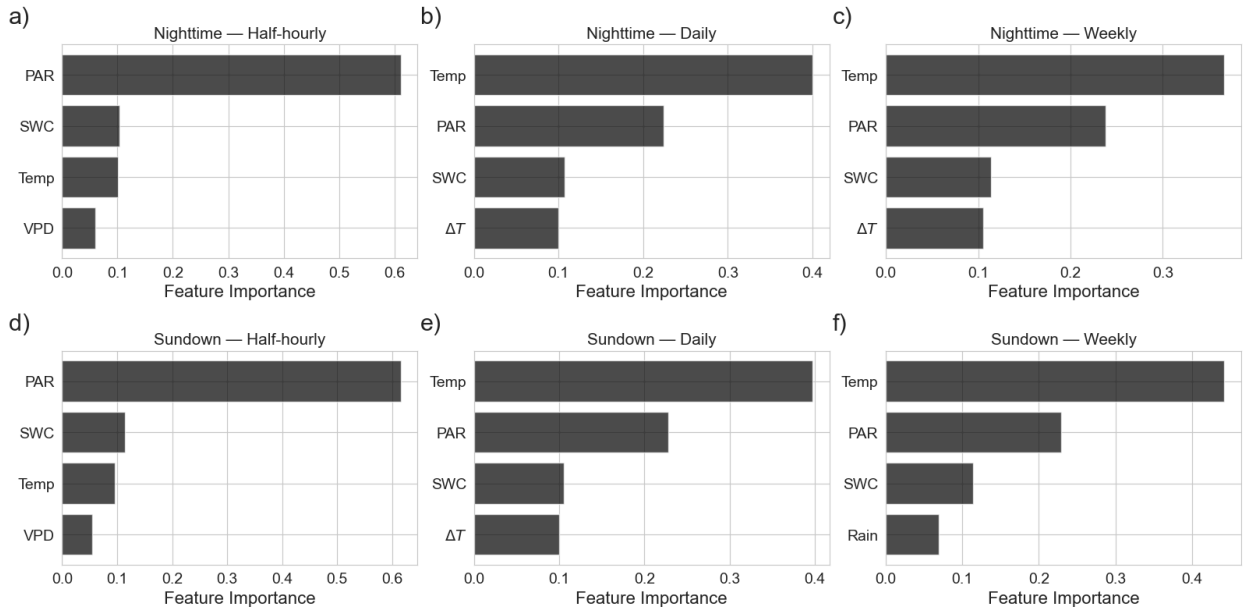
The manuscript pursues two goals: (i) evaluating partitioning methods and (ii) attributing LUE drivers across timescales. I think the authors could shorten the manuscript by more effectively communicating Section 3.1, 3.1.1 and 3.2, which precede one of the actual main results (i.e (i)). I suggest the following:

• Tighten Sects. 3.3.1–3.3.2 considerably, or move much of them (e.g. the detailed parameter comparison and Fig. 2d–e) to the Supplement. The key result, that morning data inflate α and γ while afternoon fits are stable, can be made in far less space.

We thank the reviewer for their time and contributions. This text has been shortened considerably to highlight key results more clearly. In addition, per the comment above, sections 3.1, 3.1.1, and 3.2 have also been edited to be more clear and concise.

• State the link between the two parts explicitly (partitioning and LUE drivers). In particular, say how much the LUE driver results would change if GPP_N or GPP_S were used instead of GPP_A. If the drivers are insensitive to the method, as the authors imply, that strengthens the paper. If they are sensitive, it is an important result in itself.

We thank the reviewer for this suggestion. We repeated the LUE driver analysis using GPP from all three partitioning methods and added this as a new supplemental figure (Fig. A8). The leading driver was the same for all three methods at every timescale: PAR at the half-hourly scale and temperature at the daily and weekly scales. The ranking of secondary predictors differed, with SWC ranking higher for the nighttime and sundown methods than for the afternoon method. The main conclusions about timescale-dependent drivers of LUE are therefore robust to the choice of partitioning method, which we now state explicitly as a link between the two parts of the study.



• Shorten the Conclusions. They currently restate most of the Results.

We thank the reviewer for this comment. Redundancy in the conclusions has been removed.

• Fix the section numbering: there are two "Environmental Variables" sections (2.1 title and 2.3),

and the conclusions are numbered "1. CONCLUSIONS”.

Thank you for your comment, this has been fixed.

2. Effects of gap-filling on the partitioning methods

According to the authors, 62% of half-hours remain after quality control (L146). The gap-filling scheme (using linear interpolation for less than 90 min, and a mean diurnal variation for up to 6 h, and longer gaps unfilled) affects the partitioning windows very differently (I guess..), but this is not addressed or referred to. Specifically:

• Were gap-filled values included when fitting the light-response curves, the night-time temperature regressions and the sundown NEE–PAR regressions? If so, what is the effect of the mean diurnal variation filling, with a +- 3-day / +-1 h window, on the morning–afternoon analysis and in the short sundown windows that the analysis focuses on. Please report the fraction of gap-filled data within each method's fitting window (night, sundown, morning, afternoon).

We thank the reviewer for bringing up this point. Gap-filled values were not used in any partitioning fit. Only quality-controlled NEE was used for partitioning. Gap-filled values (1.2% of half-hours) were used only to compute NEE flux totals. The fraction of half-hours with measured NEE was 78.1% at night, 81.4% in the morning, 79.2% in the afternoon, and 72.9% in the sundown window. We now state this explicitly in the manuscript (L208-210):

“Following gap-filling, 4.29% of daytime observations and 4.43% of nighttime observations remained missing. All partitioning fits used measured, quality-controlled NEE only; gap-filled values were used solely for computing NEE flux totals.”

• The u* threshold of 0.313 m s-1 (Fig. A2) is quite high and lies well above the nighttime u* values.
Thus, nighttime gaps must be very large, and the nighttime method would be based on a small and possibly unrepresentative subset (?). I suggest the authors showing how sensitive is each partitioning method to the u* filtering, particularly the nighttime method.

We thank the reviewer for highlighting this point. We clarify that no u* filter was applied: thresholds typical of other sites would have removed most nighttime data at La Selva (Fig. A2b). Instead, NEE included profile storage fluxes which remove any structural bias of datapoints (Fig. A2c). To assess residual bias from low-turbulence periods, we performed a sensitivity test restricting nighttime data to u* ≥ 0.1 and ≥ 0.2 m s^{-1}, which changed $R_{eco,N}$ by less than 2%. We have revised the text to state this explicitly (L582-585):

“A u* filter was not applied, as thresholds typical of other sites would have removed most nighttime data. Instead, we relied on profile storage fluxes, which limit this bias: in a sensitivity

test, $R_{eco,N}$ changed by <2% when nighttime data were restricted to u* ≥ 0.1 or ≥ 0.2 m s-1, although losses to drainage cannot be excluded."

• L530–532 state that restricting fits to afternoon periods "reduces bias in gap-filled fluxes". Could the authors elaborate on that? I was thinking that given that the gap-filling used here does not rely on light-response curves, how could this be? Please clarify, or show this.

We thank the reviewer for pointing this out. This text has been removed from the manuscript since the gap-filling does not rely on light-response curves.

• I recommend a sensitivity test using standard gap-filling methods (e.g. MDS and the nighttime and daytime methods in REddyProc or ONEFlux) and comparing the results with the current ones.

Thank you for this suggestion. We agree that comparing alternative gap-filling approaches could provide additional insight into uncertainties associated with estimating NEE and partitioning GPP. However, a comprehensive comparison of gap-filling methodologies is beyond the scope of this study, which focuses on evaluating NEE partitioning methods and their influence on estimates of GPP and its environmental drivers. Although approaches such as MDS, used in REddyProc and ONEFlux, are widely used, their performance can be challenging in tropical ecosystems characterized by high environmental variability, rapid changes in cloud cover and radiation, and relatively weak seasonal patterns. These conditions can limit the availability of comparable meteorological observations within the sampling windows used by MDS and potentially introduce additional uncertainty into gap-filled estimates. Similarly, alternative nighttime and daytime partitioning approaches involve different assumptions for estimating missing fluxes and partitioning NEE, and evaluating their performance would require a separate methodological comparison. Therefore, we retained our original gap-filling approach to maintain methodological consistency. We agree that evaluating the sensitivity of results to alternative gap-filling approaches would be a valuable direction for future work.

• For the morning and afternoon methods, both GPP and Reco are taken from the fitted lightresponse function, rather than as residuals of observed NEE. Nighttime GPP is set to zero and gaps > 6 h are left unfilled, so sums of NEE, GPP and Reco may cover different half-hours. In my experience at another tropical site, such inconsistencies can produce substantial mismatches, and even sign differences, at monthly integration. I recommend that, for every method, the reconstructed NEE (Reco − GPP) closes with the observed NEE. From the figures it looks like it does, but I suggest doing it more quantitatively.

We thank the reviewer for this comment. For all methods, only R_{eco} is taken from the fitted model; GPP is computed as the residual GPP = R_{eco} − NEE from observed NEE, and nighttime R_{eco} was set equal to observed NEE with GPP = 0. This approach of setting negative half-hourly

GPP (~4% of daytime half-hours) to zero broke this closure, biasing GPP upward by ~0.3 μmol m^{-2} s^{-1}. We have changed this to report all GPP values at their unclipped GPP (L253-256):

"During the day, R_{eco} was estimated from each method's fitted model, and GPP was computed as the residual GPP = R_{eco} − NEE using observed NEE. Negative half-hourly GPP values were retained rather than set to zero, so that reconstructed NEE (R_{eco} − GPP) closes with observed NEE."

3. Robustness of the LUE driver attribution

The results of the random forest analysis are interesting, but several design choices weaken the attribution:

• LUE = GPP_A/PAR, and PAR is also a predictor. Because GPP saturates at high light, LUE decreases with PAR by definition. The negative LUE–T and LUE–VPD relationships (Fig. 6a–c) and the dominance of PAR (Fig. 5d) could potentially be artifacts of the PAR normalization, instead of physiology? The authors could, for example, model GPP with PAR as a predictor, or use LUE at GPPmax and compare the results.

We agree that the dominance of PAR in the LUE models partially reflects the relationship between LUE and PAR, as LUE decreases when GPP approaches light saturation. To evaluate whether the negative LUE–temperature and LUE–VPD relationships could also result from PAR normalization, we modeled half-hourly GPP directly as a flexible function of PAR, with temperature and VPD included as additional predictors. Standard errors were clustered by day to account for temporal dependence (see table below; not added to manuscript). Temperature remained negatively associated with GPP across all scenarios, while VPD was significantly negatively associated with GPP in all but one scenario. These results support the negative temperature and VPD relationships that aren't fully explained by the mathematical effect of PAR normalization on LUE. However, the magnitude of these associations varied among partitioning methods and generally decreased when time of day was included as a predictor. For afternoon-method GPP, the strong covariance between VPD and time of day prevented us from independently resolving their contributions. Thus, although the additional analysis supports negative associations between temperature, VPD, and GPP, the relative contributions of these environmental drivers remain uncertain.

Partitioning Method	Controls	Temperature (per ºC)	VPD (per kPa)
Afternoon	PAR	−1.17 ± 0.19	−4.40 ± 0.82
Afternoon	PAR + hour	−0.86 ± 0.20	−0.80 ± 0.82 (ns)
Nighttime	PAR	−0.73 ± 0.17	−6.37 ± 0.75
Nighttime	PAR + hour	−0.38 ± 0.18	−2.73 ± 0.74
Sundown	PAR	−0.80 ± 0.17	−6.14 ± 0.73
Sundown	PAR + hour	−0.47 ± 0.18	−2.59 ± 0.74

Associations of half-hourly GPP with air temperature and VPD after controlling for PAR, for each partitioning method.

• Record length: with a single 11-month record that is not a full seasonal cycle, and the drying signal concentrated in the last two months, the weekly-scale importance of temperature and SWC is essentially driven by one event. This should be stated as a limitation more clearly, both in the abstract and in the conclusions.

We agree with this comment. This has been explicitly noted (L34-36; L790-795):

“However, the apparent importance of temperature and soil moisture at weekly timescales was influenced by a drying event concentrated in the final two months of the measurement period.”

“However, the interpretation of longer-timescale environmental drivers is limited by the duration of our measurement record. Consequently, the increased importance of temperature and SWC at weekly timescales was likely strongly influenced by the relatively short period of drying in the final two months of observations. Longer-term observations spanning multiple seasonal and climatic cycles are needed to determine whether these relationships persist and to better characterize the environmental constraints on tropical forest productivity.”

• Validation scheme: see the comments at L335–339 below.

We thank the reviewer for this comment. It is addressed accordingly (see below).

Minor points

L30: And Reco? Please also report mean Reco in the abstract. I know is not the main target of the
study, but perhaps is worth reporting there?

Thank you for this point. We now report R_{eco} in the abstract (L30-31):

“GPP averaged 11.5 μmol CO^2 m^{-2} s^{-1} and R_{eco} averaged 9.04 μmol CO^2 m^{-2} s^{-1}.”

L34–36: Is this at all surprising? I recommend that the authors end the abstract with a more specific message.

We agree this is not a strong ending. We have replaced the final sentence with a stronger take home message (L36-41):

“Partitioning method choice mainly affected the physical plausibility of fitted light-response parameters, while weekly GPP estimates agreed within ~4% and leading LUE drivers were consistent across methods. LUE declined most steeply near 27.7 °C, close to photosynthetic temperature optima of tropical trees. These findings highlight the importance of timescale when identifying environmental constraints on tropical forest productivity and suggest that increasing

temperatures and atmospheric dryness may impose growing limitations on carbon uptake, even in perennially wet tropical forests."

L42–43: I would not present this as a broad agreement in the GPP community. The size of global
GPP is under active debate. See:
- https://www.nature.com/articles/s41586-024-08050-3
- https://www.nature.com/articles/s41586-026-10562-z
- https://doi.org/10.1038/s41586-026-10561-0

We thank the reviewer for raising this point. This has been edited accordingly (L43-46):

"Tropical forests are disproportionately important to the global carbon cycle, accounting for a large share of global terrestrial gross primary production (GPP) and storing roughly 25% of terrestrial carbon stocks (Clark et al., 2017; Meister et al., 2012). However, the magnitude of both global and tropical GPP remains debated, with the largest discrepancy in the tropics (Lai et al., 2024; Tian et al., 2026; Lai et al., 2026)."

L51: The following study could be cited here as well, it gives an observationally constrained timescale for the transit time of carbon in tropical forests:
- https://bg.copernicus.org/articles/22/455/2025/

Done, thank you.

L61–62: "However, there are uncertainties associated with both approaches." This applies to all approaches. I suggest removing the sentence and going directly to "However, the nighttime method often…".

Thank you for this comment. This serves as a transition sentence to cue the reader as to what's coming next. We have edited it to be less generic (L66-67):

"However, each approach has methodological limitations that can introduce uncertainty into GPP estimates."

L68–69: What is the growing season at La Selva? Please rephrase or refer the reader to a part of
the year.

Thank you for making this point. This phrasing has been changed to avoid stating this as it was somewhat misleading since La Selva doesn't have a conventional growing season as a nearly aseasonal wet tropical forest.

L72: "More recently" for a 2016 paper? Please rephrase, or add more recent references.

We thank the reviewer for raising this point. This was phrased accordingly since it was developed more recently than the nighttime and daytime partitioning methods, but has been rephrased to avoid confusion (L82):

“A later developed approach, the sundown partitioning method,…”

L76–77: This sentence feels out of place: there is a sudden jump from Reco to global GPP.

We agree and this sentence has been cut.

L94: A citation is required here.

Thank you for this comment. We have added a citation to the FLUXNET site summary, which includes a figure from 2015 illustrating this quite nicely. Unfortunately, the site summary has not been updated since 2015, but the tropics are still disproportionately underrepresented from EC measurements.

L96–97: This statement is not entirely correct. In fact, the nighttime method, particularly the one presented by Reichstein et al., (2005), is not used given the very small temperature range in tropical forests, see:
- https://doi.org/10.1016/j.agrformet.2013.04.031
- https://onlinelibrary.wiley.com/doi/10.1111/gcb.15905
I suggest the authors to rephrase that sentence.

We thank the reviewer for raising this point. This statement has been removed.

L110: Research question 3 needs reformulating; the phrasing sounds odd.

We agree, this has been rephrased accordingly (L122-123):

“What are the rates of GPP and ecosystem respiration at this site, and how can these estimates inform subsequent analyses of environmental drivers of LUE?”

L152: Is 57 m the height of the tower? Please state it explicitly here and I guess this is where the gas analyzer is, right?

Yes, this is stated at the beginning of the section (L150-151):

"CO_2 and H_2O EC flux measurements were collected via an enclosed path infrared gas analyzer (Li-7200RS, LI-COR Biosciences) and sonic anemometer (CSAT-3b, Campbell Scientific) at the top of a 57 m tower."

L255–256: Solar noon is fixed at the campaign mean (11:45 CST). How sensitive are the morning and afternoon partitioning results to this choice?

We thank the reviewer for this constructive comment. We evaluated the sensitivity of the morning and afternoon partitioning results to the choice of solar noon and found that using the campaign-mean solar noon (11:45 CST) had little effect on the results. This is likely because solar noon varies relatively little throughout the year at our study site (10.43° N), given its proximity to the equator. Therefore, we believe that using a fixed campaign-mean solar noon provides a reasonable reference for consistently defining morning and afternoon periods throughout the duration of the study.

L268: Baseline respiration at 0 °C? This is far outside the observed range (about 22–27 °C). A reference temperature within the observed range would be more meaningful, right?

We agree. The intercept now refers to base respiration at the mean nighttime temperature of each 7-day window (22–27 °C), which lies within the observed range. Eq. 2 has been revised accordingly. In revising the nighttime method, we also now estimate a single temperature sensitivity from all nighttime data, as window-specific slopes were poorly constrained.

L277–278: Mahecha et al. (2010) show a large spread for tropical sites. Please check the tropical values rather than comparing with the global mean.

We thank the reviewer for highlighting this. We now compare our fitted Q_{10} with the site-specific value reported for nighttime NEE at La Selva (Q_{10} = 1.79; Loescher et al., 2003), rather than a global mean (L321-323).

"The fitted Q_{10} of 2.02 is comparable to the value of 1.79 previously reported for nighttime NEE at La Selva (Loescher et al., 2003), although temperature explained little nighttime variability in both studies."

L335–336 (rolling-origin validation with a 50/50 split): Please clarify if this validation strategy is a single temporal split, if so, then: the test period contains the end-of-campaign drying, which the model never sees in training, right?

We thank the reviewer for pointing out this inconsistency. We neglected to update the text after adjusting the methods, but the validation was not a single temporal split and instead an expanding-window rolling-origin scheme. In this scheme, following an initial training period in the first 50% of the record, the remaining data were divided into 10 consecutive test folds, each

predicted by a model trained on all preceding data. The folds covering the end-of-campaign drying (April–May 2023) were therefore predicted by models trained through the preceding months, including the earlier drying in late 2022. Even so, 20% of soil water content values in the final two folds fell below the minimum of their training data, and because random forests cannot extrapolate beyond the training range, the models could not fully represent soil moisture effects under the driest conditions. We note this limitation in the text (L382-386):

“For half-hourly and daily models, we used expanding-window rolling-origin validation: models were first trained on the initial 50% of the record, and the remaining 50% was divided into 10 consecutive test folds, with each fold predicted by a model trained on all data preceding it. Test observations therefore always came from periods after the training data.”

L337–339 (10-fold blocked CV, weekly model): With n ≈ 48 weeks and gaps (e.g. Aug–Sep 2022 in Fig. 4), each block holds about 4–5 weeks. With this I think the per-fold R2 will be unstable(?). Please state whether R2 is computed from pooled out-of-fold predictions or averaged across folds.

We thank the reviewer for this constructive comment. R^2 and RMSE were computed from the pooled out-of-fold predictions across all folds, not averaged across folds. We agree that per-fold R^2 would be unstable given 4–5 weeks per weekly block, which is why we report pooled values. We now state this explicitly in the Methods (L386-389):

“Given the smaller sample size of the weekly model (n = 48 weeks), 10-fold blocked cross-validation was used, in which each fold was a contiguous temporal block held out in turn. Half-hourly models additionally included time-of-day and month terms (sine and cosine) and VPD lagged by one and two half-hours as predictors. For all timescales, R^2 and RMSE were computed from the pooled out-of-fold predictions.”

L360: Remind the reader what that temperature range is.

We thank the reviewer for making this point. This part of the sentence has been deleted due to Reviewer 1’s comments of repetitiveness.

L403–407: There are precipitation deficits but lower VPD, attributed to La Niña(?). Is this for the same months? It doesn’t makes sense if it is.. no? Please report VPD for the same months with precipitation anomalies.

We thank the reviewer for pointing out this inconsistency. Daytime VPD was below average in 9 of 11 months, but above average in the two months with the largest rainfall deficits (December 2022 and May 2023), indicating that atmospheric dryness increased only under the strongest deficits. We have replaced the statement that VPD was consistently below average and

removed the attribution to La Niña plus reported VPD for the same months as the precipitation anomalies (L447-452):

“Daytime VPD was below its monthly climatology in 9 of 11 months (by 0.11–0.31 kPa), including during the moderate July deficit, but exceeded the long-term mean in the two months with the largest rainfall deficits (December 2022: +0.05 kPa; May 2023: +0.18 kPa). Atmospheric dryness therefore increased only under the strongest rainfall deficits, while humidity remained high during moderate deficits. Overall, the study period had near-normal temperatures and radiation, with below-normal rainfall and elevated VPD concentrated in December and at the end of the campaign.”

L430–431: The authors state that taller towers improve the robustness of EC flux estimates: not always. Larger footprints, heterogeneity, morning flush peaks and storage can also become larger issues. Please have this in mind in this part.

We thank the reviewer for highlighting this. We have revised the statement to acknowledge that although greater measurement heights can improve the vertical representativeness of flux measurements, they also increase the footprint area and may introduce additional uncertainties associated with landscape heterogeneity, horizontal advection, and below-sensor storage (466-469):

“Measurement height can influence the representativeness and uncertainty of EC flux estimates, with taller towers sampling larger areas but also potentially increasing sensitivity to landscape heterogeneity, horizontal advection, and uncertainties associated with below-sensor CO_2 storage (Munger et al., 2012).”

L442: Such strong uptake relative to NPP would imply a very high carbon use efficiency, which is difficult to reconcile with other wet tropical forests, but plausible I guess. Please discuss having Chambers et al., (2004) findings in mind: https://esajournals.onlinelibrary.wiley.com/doi/10.1890/01-6012

We thank the reviewer for this thorough review. This sentence has been removed since Reviewer 1 commented that it was not directly applicable for comparison purposes. We also recognize the importance of ecosystem respiration when interpreting net carbon uptake, which supports that high tropical forest productivity does not necessarily translate into equivalent net ecosystem carbon uptake. We have therefore avoided using the Russell et al. comparison.

Table 1 (L461): Please double check this study, as I don't think they used nighttime partitioning (i.e. Reichstein et al. (2005)). See my comment at L96–97.

We thank the reviewer for pointing out this inconsistency. This has been updated to the daytime partitioning method (L493).

L462 (Sect. 3.3.1): Please check for and cite other studies that have done similar comparisons in the tropics.

We thank the reviewer for this comment. We now compare our results with similar analyses in other tropical forests, including comparisons of respiration estimates from daytime light-response extrapolation, u*-filtered, bottom-up, and radon-based approaches at the Tapajós site (Hutyra et al., 2008), and the time-of-day dependence of the canopy light response observed in Amazonian forests (Doughty et al., 2006): (L514-519):

“Similar sensitivity of respiration estimates to methodology has been reported in other tropical forests. At the Tapajós site in Brazil, respiration estimated by extrapolating daytime fluxes to zero light differed from u*-filtered, bottom-up, and radon-based estimates, with lost nighttime flux and canopy storage identified as the main sources of bias (Hutyra et al., 2008). Time-of-day dependence of the canopy light response, with lower uptake at a given irradiance in the early morning and late afternoon, has also been observed in Amazonian forests and attributed to circadian regulation of gas exchange (Doughty et al., 2006).”

L511–512: No need to repeat this, as it is clear from the legend and caption.

Thank you for this comment. This has been cut.

L585: "Earth Science Data Systems, 2025": what is this citation? Is it a Reco study? Please cite the original authors of the LBA-ECO CD-08 dataset or the associated paper.

We thank the reviewer for this comment. This has been updated to cite the associated paper by Chambers et al., 2004 (L568).

L605–609: The claim that storage fluxes explain the agreement could in principle be tested, e.g. by repeating the nighttime partitioning without the storage term. This could be shown in the Supplement.

We thank the reviewer for this suggestion. We have substantially revised this paragraph in response to Reviewer 1’s comments and removed the previous interpretation that accounting for storage fluxes explains the agreement among partitioning methods. We now distinguish between accounting for CO_2 storage using profile observations and the potential for flux losses under low-turbulence conditions. We also clarify that agreement among partitioning methods does not necessarily indicate accuracy, particularly because the methods share measured NEE. Given this revised interpretation, we do not believe that an additional partitioning sensitivity test excluding the storage term is necessary.

L630–632: Is low turbulence the only concern? The nighttime method also assumes the same temperature dependence during day and night, ignores daytime inhibition of leaf respiration (Keenan et al., 2019) and extrapolates across temperature ranges.

We agree that low nighttime turbulence is not the only theoretical limitation of the nighttime partitioning method. We edited the prior section to discuss these additional limitations, including the assumption of equivalent temperature dependencies of respiration during daytime and nighttime conditions, extrapolation of nighttime temperature-response relationships to daytime conditions, and the omission of potential light inhibition of leaf respiration. To avoid repetition, we did not address this again.

L697: Did the authors test other predictors, e.g. cumulative water deficit (CWD)?

Yes, thank you for this suggestion. We evaluated several additional predictors during model development, including maximum and minimum daytime and nighttime air temperatures and cumulative water deficit (CWD), among others. However, these variables consistently showed low relative importance in the random forest models and contributed little additional explanatory value. We therefore excluded them from the final analysis and figures to focus on the environmental variables that were most influential across timescales.

L759: Other sites show lower temperature optima, and including them may change the interpretation, consider this paper: https://iopscience.iop.org/article/10.1088/1748-9326/aa6f97

We thank the reviewer for this comment. We have added a comparison to this study accordingly (L729-733):

“This is also within the range of ecosystem-scale optimum temperatures for light-saturated net photosynthesis reported for seven tropical forests (23.7–28.1 °C; Tan et al., 2017). Tan et al. (2017) further showed that when VPD was held constant, the ecosystem optimum at a per-humid site rose from ~26.5 °C to at least 30 °C, indicating that stomatal responses to VPD lower apparent temperature optima at the ecosystem scale, consistent with the covariation of temperature and VPD at our site.”

L761–763: Isn't this true by definition? VPD depends exponentially on temperature through saturation vapour pressure. Please clarify what is meant.

Yes, this sentence has been removed in edits.

L797–798: Did the authors try another ML method, to check that the results are not specific to random forests?

Thank you for this suggestion. We repeated the analysis with gradient boosting and a regularized linear model, using identical predictors and cross-validation folds. The leading predictor of LUE at all timescales for all three methods remained the same, indicating that this result is not specific to random forests. Model performance was similar across methods (CV R^2 = 0.32–0.42 at half-hourly and daily scales). Secondary drivers vary among methods and are interpreted with caution.

Timescale	RFM	Gradient Boosting	Linear (ridge)
Half hourly	0.359	0.320	0.388
Daily	0.418	0.341	0.390
Weekly	0.385	–0.065	0.526

Cross-validated performance of LUE for three machine learning methods at half-hourly, daily, and weekly timescales. Models were random forest (300 trees, maximum depth 12, minimum leaf size 5), gradient boosting (histogram-based; 300 iterations, learning rate 0.05, minimum leaf size 20), and ridge regression on standardized predictors (regularization selected by internal cross-validation). All methods used identical predictors and cross-validation folds: expanding-window rolling-origin validation for half-hourly and daily models, and 10-fold blocked cross-validation for the weekly model. CV R^2 was calculated from pooled out-of-fold predictions; negative values indicate predictions worse than the mean of the held-out data. The four highest-ranked predictors are shown with their permutation importance (mean decrease in held-out R^2, 10 repeats per fold). This has been incorporated into the manuscript text (L688-691):

“The leading LUE predictor at each timescale was consistent across random forest, gradient boosting, and ridge regression models, indicating that these patterns were not specific to the random forest approach. Although predictive performance was comparable across methods at the half-hourly and daily scales, it varied more substantially at the weekly scale, and secondary predictor rankings differed among algorithms.”