After more than three decades of global EC measurements, there is still a large data gap in the tropics. As tropical forests contribute largely to carbon and water fluxes, this data gap is a large source of uncertainty in estimating global carbon and water dynamics. This paper adds to closing this gap with EC data from a wet tropical forest stie in Costa Rica, even though the data coverage is only +- one year. I think the dataset is very valuable, however the data processing is not very thoroughly explained in the paper, e.g. info about u* filtering, data left after QC, shape and value of storage correction, …. In addition it seems like several references are not on the right place and/or are not the main reference for the method or results mentioned in the text. This might give the indication that the methods and claims are not well supported. The partitioning methods explained in the text are missing details, which leaves the reader in confusion and limits the understanding of the results and the re-usability of the methods. Some large conclusions about partitioning methods are drawn and in my opinion this conclusions are not yet well supported by the analyses carried out in the study. While the set-up and ideas behind the study are very interesting, such as comparing different partitioning techniques, looking into drivers on different timescales, … the study itself still raises more questions than answers and needs some more work before publishing.

Overall I would advice to check all references, to more accurately describe the analyses on the NEE as this is the strong basis for all following analyses, to more thoroughly describe the partitioning methods used and more clearly indicate (dis)similarities with more common methods and to strengthen the comparison between partitioning methods.

We sincerely thank the reviewer for their thoughtful and detailed review of our manuscript. We appreciate their recognition of the value of this dataset. We agree that clearly documenting the processing of EC measurements and the implementation of NEE partitioning methods is essential, particularly given the importance of these methodological choices for subsequent estimates of GPP and their environmental drivers. In response to these comments, we have substantially revised the Methods section to provide a more detailed description of the data processing, quality control, u* evaluation, and profile-based storage correction. We have also expanded the descriptions of the NEE partitioning methods, including their underlying assumptions, implementation, and differences from more commonly used approaches. Additionally, we have carefully reviewed the manuscript references to ensure that methodological descriptions and interpretations are supported by appropriate literature. We have also revisited the comparisons and conclusions drawn from the different partitioning approaches, refining our interpretation to better distinguish between differences in fitted light-response parameters and differences in resulting GPP estimates. Finally, we have clarified the limitations of our observational record and the extent to which our analyses support conclusions about environmental controls on LUE. We believe these revisions have substantially improved the methodological documentation, clarity, and interpretation of the manuscript. We have addressed each of the reviewer's comments individually below and thank them again for their suggestions, which have helped strengthen the study.

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

Comments:

1 Introduction

Line 57 to 69: rewrite this section, this reads very unstructured and contains dubious information.

➔ Introduce NEE as the sum of GPP and RECO. The sentence “partition net ecosystem exchange (NEE) into its component fluxes”, is not entirely clear.

We thank the reviewer for noticing this. We have now improved the definition of NEE.

Defined NEE more clearly (L59-62):

“Gross primary production (GPP), the largest carbon flux from the atmosphere into terrestrial ecosystems, is not directly measured at the ecosystem scale and must therefore be estimated by partitioning net ecosystem exchange (NEE) into gross carbon uptake through photosynthesis and carbon release through ecosystem respiration (R_{eco}). We adopt the convention that positive NEE denotes net CO_2 release to the atmosphere, NEE = R_{eco} − GPP.”

➔ “the nighttime-based method to extrapolate respiration's temperature response to daytime periods” rephrase this to state that it is the nighttime temperature-respiration relation that is extrapolated to the daytime periods. Now it is not clear which respiration temperature response is meant.

We thank the reviewer for this suggestion. We have rephrased this for clarity (L67-69):

“Nighttime-based partitioning assumes that the temperature dependence of respiration observed at night adequately represents its behavior during the day, but nighttime flux measurements can be affected by weak atmospheric turbulence and insufficient vertical mixing.”

➔ “The nighttime method often corresponds to the period of day when …”. “Method corresponds to a period” is not clear. Better write that the data used to fit the temperature-respiration relation is nighttime data, and then mention the associated uncertainties with nighttime data. Uncertainties in nighttime data are indeed accumulation, but also heavy data filtering is a source of uncertainty and the assumptions underlying the methos. I would argue that if there is respiration underestimation, than this is an issue of your NEE measurements, and not inherently a downside of the partitioning technique. If this is accurately corrected for, using filtering and storage corrections, this underestimation would not influence the partitioning method too much., so I would definitely include additional other uncertainties next to the underestimation, such as heavily filtered nighttime data, light inhibition of leaf respiration, the uncertainty due to the corrections, ..

Thank you for this comment. This has been rephrased for clarity and edited points of weakness for the nighttime partitioning method (L69-73):

"This may cause CO_2 accumulation below the eddy covariance (EC) measurement height (Aubinet et al., 2012). Quality filtering can exclude a substantial fraction of nighttime observations, reducing the data available to constrain the temperature–respiration relationship. Uncertainties in flux corrections and potential differences between nighttime and daytime respiration, including light inhibition of leaf respiration, can further affect partitioning results."

➔ "Meanwhile, limitations associated with the daytime partitioning method include potential overestimation of morning CO2 fluxes from buildup throughout the night, as well as strong sensitivity of GPP estimates to variability in incoming radiation". The overestimation of morning CO2 from build up, should already been corrected in the NEE, so should not influence the partitioning method, similarly as the previous comment. The uncertainty or variability due to corrections can play a part of course. I am not sure how the "strong sensitivity of GPP estimates to variability in incoming radiation" introduces a bias. Do you mean that the model does not capture fast changing environmental conditions?

We thank the reviewer for this comment. However, it is quite challenging to completely correct the morning storage fluxes. We have edited points the of weakness for the daytime partitioning method to include temperature and clarify this point (L74-79):

"Meanwhile, daytime-based partitioning relies on the assumption that a light-response model adequately represents the relationship between NEE and PAR over the fitting period. Although fitting over relatively short time windows allows model parameters to vary with changing environmental conditions, the resulting GPP estimates remain sensitive to the ability of the fitted relationship to capture changes in ecosystem responses to fast changing environmental conditions such as light, temperature, and the potential overestimation of morning CO_2 fluxes from buildup throughout the night."

➔ "Moreover, the daytime method is subject to biases introduced by seasonal changes in phenology, leaf area index, and water stress (Keenan et al., 2019; Lasslop et al., 2010), which alter both the shape and magnitude of ecosystem light response curves over the growing season."

The shape of the light response curve is allowed to change, as parameters are fitted on small windows, so I don't see how seasonal changes can introduce strong biases. Maybe here add the light inhibition of leaf respiration, this is an issue with both the night and daytime partitioning method? Also I don't see how the two references here explain the possible biases.

We thank the reviewer for pointing this out. This has been addressed along with the previous comment, in L74-79.

Line 71-72: More recently, the sundown partitioning approach has been proposed as an alternative to better constrain GPP and Reco estimates (Wehr et al., 2016).

Wehr et al 2016 used simultaneously carbon and isotope EC measurements to partition NEE. This paper is not at all related to the sundown method. Are you referring to van Gorsel et al. 2009?

We thank the reviewer for highlighting this. Yes, van Gorsel et al. 2009 is the appropriate citation. This has been fixed along with other misplaced references throughout the manuscript.

2 Methods

Line 137: closed path needs to be enclosed path, because the LI7200 is an enclosed analyzer.

Done.

Line 140 Webb-Pearman-Leuning (WPL) corrections for density fluctuations (Webb et al., 1980) are not needed for closed and enclosed analyzers, however this will not have influenced your data as eddypro will automatically deactivate the setting.

Thank you for this comment. This has been removed from the text to avoid confusion.

Line 144: The enclosed-path design of the Li-7200RS minimizes rain contamination of the optical path, and remaining rain-affected periods were removed through EddyPro's standard quality flagging

➔ There are not issues with rain contamination of the optical path in an enclosed path.

Thank you for highlighting this. This has been removed from the text to avoid confusion.

Maybe add a diel cycle of non-storage corrected, storage corrected NEE and the storage correction term itself in supplementary to indicate the effect of storage.

We agree this would be a useful addition. This has been added to the supplement (Fig. A2):

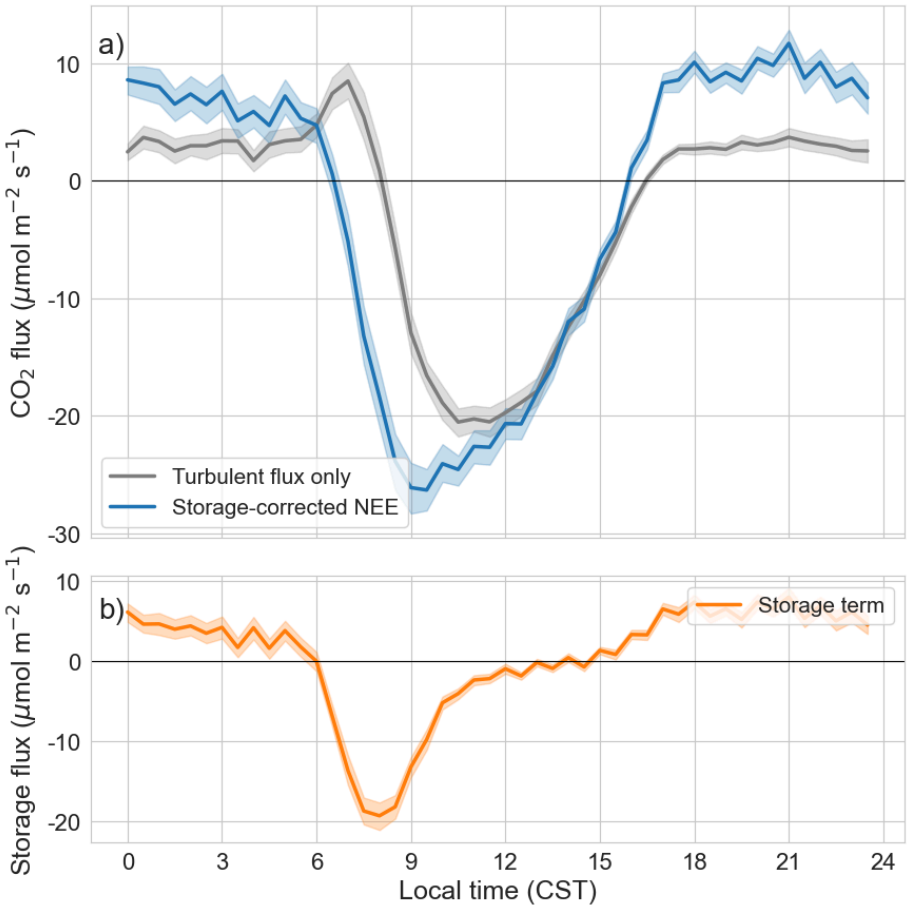


Figure A2: Mean diel cycles of (a) turbulent CO2 flux without storage correction (gray) and storage-corrected net ecosystem exchange (NEE; blue), and (b) the CO2 storage flux below the EC measurement height, calculated from the nine-level CO2 profile (1–57 m). Positive values indicate net CO2 release to the atmosphere. Shading shows 95% confidence intervals of the half-hourly mean from bootstrap resampling of days (1000 iterations).

Line 165: closed path sonic anemometer

Using the enclosed path analyzer, sonic anemometer and tower profile system…

Done.

Line 168– 171 The u* threshold reflects the potential bias in nighttime NEE due to low turbulence conditions.

➔ Mention the u* threshold that you used to filter out low turbulent conditions, mention also how many data points are removed and how many data points were left after the u* filtering.

We thank the reviewer for pointing out this lack of clarity. We did not use the u* threshold for filtering since it would have excluded a substantial portion of observations (supported in Fig. A2). This has been clarified in text (L179-185):

"However, we did not apply this threshold to filter the final NEE record because doing so would have excluded a substantial proportion of observations (Fig. A2). Instead, we retained observations that passed the EddyPro quality-control criteria and calculated NEE by combining turbulent CO_2 fluxes with profile-derived storage fluxes. This approach accounts for temporal changes in CO_2 stored below the EC measurement height, which are particularly relevant under low-turbulence conditions. We evaluated the sensitivity of NEE estimates to the u* threshold as part of the uncertainty analysis, rather than using the threshold to exclude observations from the final flux record."

➔ Uncertainty due to U* is reflected in the difference between the 5th and 95th percentile of u* thresholds, mention these thresholds maybe in supplementary.

We thank the reviewer for this comment. We added these details to Fig. A3 caption:

"Block-bootstrap analysis yielded 5th and 95th percentile threshold estimates of 0.263 and 0.413 m s^{-1}, respectively, characterizing uncertainty in the estimated u* threshold."

➔ In supplementary Fig A2 there is stated: "Points below the threshold would be excluded due to insufficient turbulence, while points above the threshold would be retained. Storage and turbulent flux components are shown." With the "would be" construction it seems like the data points below the threshold were not removed?

We thank the reviewer for pointing this out. This has been addressed in the above comment and clarified in manuscript text (L179-185).

"However, we did not apply this threshold to filter the final NEE record because doing so would have excluded a substantial proportion of observations (Fig. A3). Instead, we retained observations that passed the EddyPro quality-control criteria and calculated NEE by combining turbulent CO_2 fluxes with profile-derived storage fluxes. This approach accounts for temporal changes in CO_2 stored below the EC measurement height, which are particularly relevant under low-turbulence conditions. We evaluated the sensitivity of NEE estimates to the u* threshold as part of the uncertainty analysis, rather than using the threshold to exclude observations from the final flux record."

Line 188: If gaps are left unfilled, how did you calculate the budgets mentioned in Line 420 later on? Are these gaps equally distributed between day and night measurements? If more nighttime gaps are present, this could result in a underestimation of the ecosystem respiration

We thank the reviewer for highlighting this. Missing NEE observations were addressed using two steps: first, linear interpolation was applied to gaps of up to three consecutive missing observations. Second, the mean diurnal variation (MDV) method was used to fill remaining gaps of up to 12 consecutive missing observations (6 hours), using mean flux values for the corresponding half-hour of the day calculated across the measurement record. Gaps longer than 6 hours were left unfilled. Before gap-filling, 4.29% of daytime NEE observations and 5.19% of nighttime NEE observations were missing. Following gap-filling, only 4.29% of daytime

observations and 4.43% of nighttime observations remained missing. The small percentage of missing data has a negligible impact on the results from our study. This has been clarified in text (L209-211).

“Following gap-filling, 4.29% of daytime observations and 4.43% of nighttime observations remained missing.”

Line 193 -194: how was this gap-filling done? Jut putting the values of the other meteorological station into the gaps? Did the instruments compared well?

We compared overlapping measurements from the flux tower and the La Selva meteorological station to assess agreement between the instruments. After confirming that the measurements were comparable, we used the meteorological station measurements to replace missing values in the tower record for air temperature, rainfall, and PAR. This has been clarified in text (L218-220):

“Before gap-filling, overlapping measurements from the two sites were compared to assess agreement between the instruments. Following agreement between sites, meteorological station measurements were used to replace missing values in the tower record.”

Which software did you use to do the partitioning, the gap-filing, ..?

Flux partitioning and gap-filling were performed using Python scripts rather than a software package. We have clarified this in the Methods and provided a citation for the code. Code is currently available for reviewers and will be made publicly available upon publication.

Line 233: formula (1)

➔ The method described in Lasslop et al. 2010 used a rectangular hyperbolic light response curve, which has a different formula that is used here. Mention that this is different. Of course the non-rectangular hyperbolic fit can also be used, however then the formula is missing a term, because now the square root of $(\alpha PAR + \beta)^2$ just returns the absolute value of $\alpha PAR + \beta$. Check the formula again.

Thank you for identifying this error. We corrected the equation and clarified that our analysis uses a non-rectangular hyperbolic light-response function, which differs from the rectangular hyperbolic function described by Lasslop et al. (2010). We revised the text accordingly (L267-269).

“We fitted a non-rectangular hyperbolic light-response function, which differs from the rectangular hyperbolic formulation described by Lasslop et al. (2010) and allows the curvature of the light-response relationship to be explicitly parameterized (Gilmanov et al., 2003).”

➔ “The y-intercept γ is the base respiration level in the absence of light, which is then used to calculate Reco following the assumptions described in Sect. 2.4”

So respiration is constant during day and night and only changes when a new curve is fitted, i.e, once every week? This is not similar to Lasslop et al. 2010 as he used a temperature dependance for respiration. This method seems closer to the method by Gilmanov et al. 2003

We corrected this citation to Gilmanov et al. 2003 accordingly (see previous comment).

Line 242-243 “Using the base respiration rate from the non-rectangular hyperbolic fit, GPP can be derived using the equation $GPP = NEE - R$eco To reliably fit this curve, a sufficient density of data points is required. This limited the temporal resolution to a minimum of one week. As a result, two timescales were used: daily and weekly. On the weekly timescale, light-response curves were fitted to obtain stable model coefficients, including the base respiration rate. These weekly respiration estimates were then applied to the corresponding daily NEE observations within each week to calculate daily GPP values.”

➔ “using the equation $GPP = NEE - R$eco”. I think this statement makes things difficult because now there can be confusion: do you mean NEE(PAR) with the NEE in the formula here, or the original measured NEE values? Likely the first

We thank the reviewer for this comment. We used the original measured NEE values, not NEE(PAR). This has been clarified in the manuscript (L282-283):

“Using the base respiration rate from the non-rectangular hyperbolic fit, GPP can be derived using the equation GPP=NEE-R_{eco}, where NEE is the observed value.”

➔ “These weekly respiration estimates were then applied to the corresponding daily NEE observations within each week to calculate daily GPP values.” So this means that you do not calculate half-hourly GPP values? I am confused here with the daily temporal resolution. You fit the curves on half-hourly data?

We thank the reviewer for highlighting this. The light-response curves were fitted using half-hourly daytime NEE and PAR observations, grouped into weekly periods to ensure a sufficient number of data points for robust parameter estimation. Nighttime observations were excluded from the light-response curve fitting. The resulting weekly base respiration estimates were then applied to the corresponding half-hourly observed NEE values within each week to calculate half-hourly GPP. We revised the text to clarify the distinction between the weekly timescale used for curve fitting and the half-hourly temporal resolution of the resulting GPP estimates (L283-287).

“To obtain robust parameter estimates, light-response curves were fitted to half-hourly daytime NEE and PAR observations grouped into weekly periods. Nighttime observations were excluded from curve fitting following Lasslop et al. (2010). The weekly base respiration estimates derived from these fits were then applied to the corresponding half-hourly NEE observations within each week to calculate half-hourly GPP.”

➔ Is this curve fit on day and nighttime data?

The curve is fit only on daytime data, following Lasslop et al., (2010). This has been clarified in text (L284-285):

“Nighttime observations were excluded from curve fitting following Lasslop et al. (2010).”

Section 2.4.2 In the Morning and Afternoon Daytime Partitioning Methods: no nighttime data is used to fit the curves?

No, see above comment. Nighttime data would occur at a PAR of 0.

Section 2.4.3:The method by Reichstein et al. 2005 uses a exponential regression model, not a linear regression. You explain why, but just maybe rephrase.

This has been rephrased accordingly (L302-303):

“This method assumes that $R_{eco,\ N}$ is primarily driven by temperature and models its temperature dependence using an exponential function. In this study, however, we used a simplified linear model:”

Line 269: “and GPPN is calculated from Reco, N”.

Mention that here you subtract the estimated Reco from the measured NEE to result in GPP.

We have clarified this in text (L312-313):

“GPP_N is calculated from $R_{eco,\ N}$ by subtracting the estimated $R_{eco,\ N}$ from the measured NEE to result in GPP_N.”

Line 272: “despite evidence that foliar respiration is often suppressed in high light (Cavaleri et al., 2008).”

I agree that light indeed suppresses foliar respiration, however the reference used here mainly talks about dark respiration and there is no mention of light inhibition of leaf respiration.

Thank you for this comment. This reference has been replaced with Keenan et al., 2019 (L315-317):

“Additionally, the model assumes that nighttime and daytime respiration are equivalent, despite evidence that foliar respiration is often suppressed in high light (Keenan et al., 2019).”

Section 2.4.4

You only mention Lasslop et al 2010 here, but this paper references the daytime partitioning method, not the sundown partitioning method. Do the methods somehow rely on some result from Eva van Gorsel 2009? You relate the Reco to temperature, is this with a linear relation?

We thank the reviewer for addressing this. We have revised Section 2.4.4 to cite van Gorsel et al. (2009) as the primary reference for the sundown partitioning approach. We did use a linear

relationship to relate Reco to temperature, and this has been re-worded in the manuscript for clarity.

Line 310: “To assess the relative importance of environmental drivers on GPP, a random forest model (RFM) was trained on observed light use efficiency (LUE) data calculated from GPPA divided by PAR analyzed on half hourly, daily, and weekly timescales (LUE = GPPA / PAR) (Breiman, 2001; Jung et al., 2011).”

➔ Data users are frequently confused on what EC measures and what we estimate of model, so try to be consistent in wording: LUE is not observed when it is calculated

We thank the reviewer for this comment. This has been fixed.

➔ It’s a complex sentence

We split this into two sentences to make it more understandable (L353-355):

“To assess the relative importance of environmental drivers on GPP, a random forest model (RFM) was trained on light use efficiency (LUE). LUE was calculated from GPP_A divided by PAR and analyzed on half hourly, daily, and weekly timescales (LUE = GPP_A / PAR).”

➔ I am confused with the reference to Jung et al.. This is a paper about upscaling FLUXNET observations of carbon dioxide, water, and energy fluxes to the global scale. Not related to LUE.

This reference has been removed.

3 Results

Line 447: “Several factors may contribute to this elevated sink strength including methodological choices such as tower configuration and storage flux treatment”

➔ How so, references? Most of the time storage flux reduces the sink strength, I would guess

We agree that this would reduce the sink strength. We have rephrased this sentence accordingly and supported it with the reference Schmid et al., 2003 (L477-480):

“Compared with other intact wet tropical forests, our estimated NEE also indicates relatively strong carbon uptake (Table 1). Reported values range from approximately 95 to −800 g C m^{-2} yr^{-1}, although comparisons are complicated by differences in measurement periods, tower configurations, and flux processing methods, particularly the treatment of CO_2 storage and weak turbulent mixing (Schmid et al., 2003).”

Line 449: Although a variety of GPP partitioning approaches have been used, the majority of previous studies relied solely on the nighttime partitioning method.

➔ You are talking about NEE here, partitioning methods can not influence the NEE measurements

Thank you for raising this point. The manuscript text has been clarified to correct this issue (L484-488):

"Resolving these potentially opposing responses requires partitioning NEE into GPP and R_{eco}. Notably, most previous studies included in Table 1 relied on nighttime partitioning, whereas our study evaluates daytime afternoon, nighttime, and sundown approaches. These methodological differences provide an opportunity to assess how partitioning choices influence estimates of tropical forest carbon exchange and the environmental controls on productivity."

Check the values again in Table 1: for example the values of French Guiana are very misleading, as these are values mentioned for different seasons and therefore can not be used as values for sink or source per year. Only use accumulated values of a whole year to make comparisons.

We have checked the table and removed French Guiana. We have updated the text accordingly to state that our NEE range is lower than the reported values from previous studies (L477-478):

"Compared with other intact wet tropical forests, our estimated NEE indicates relatively strong carbon uptake (Table 1)."

Line 464: "The base respiration parameter γ spans a physiologically unrealistic range, varying from 7.5 to 15.9 µmol m-2 s-1 (Cavaleri et al., 2008; Loescher et al., 2003)."

➔ Why is this a physiologically unrealistic range? This is explained further down, so maybe just restructure the two paragraphs

We agree this should be clarified. This whole section has been rewritten for clarity (see Sect. 3.3.1).

Line 472 – 496: I think this is an interesting analyses, however the two paragraphs can be better structured (see comment above)

We thank the reviewer for this comment. This has been rewritten (see comment above).

Line 474: "This variability is particularly pronounced during morning periods at low PAR (<500 µmol m-2 s-1), where high NEE values are observed (Fig. 2d).

➔ could this also be evening hours?

We thank the reviewer for raising this point. The evening hours show less scatter, likely because the boundary layer is more well mixed. However, we have removed figures 2d and 2e per Reviewer 2's comments so the associated text in the manuscript has been removed.

Line 489 – 495: In Fig. 2d, the initial slope is 0.0407 µmol C µmol PAR-1, whereas in Fig. 2e it is more than double at 0.0927 µmol C µmol PAR-1. Correspondingly, the estimated base respiration term in Fig. 2d is 8.22 µmol m-2 s-1, compared to 14.8 µmol m-2 s-1 in Fig. 2e, which typically averages around 10 µmol m-2 s-1 based on previous studies from the La Selva, Costa Rica rainforest (Cavaleri et al., 2008; Loescher et al., 2003). In tropical forests, documented responses of GPP and Reco to short-term variations in humidity or VPD are typically on the order of 10–30%, not the>100% change observed here (Restrepo-Coupe et al., 2013; Saleska et al., 2003).

➔ The two periods compared are a month away, so I don't completely see how the statement 'short-term variations in humidity or VPD …' is linked to the situation here, because a month is not short term and we have no clear indication of what the difference in meteorological conditions are during these two periods.

We thank the reviewer for raising this point. We removed this comparison and instead evaluate the change in γ directly against the observed meteorological conditions (Fig. 1, L499-503):

"Between the two periods, mean air temperature decreased from 26.6 to 26.1 °C and nighttime temperature from 24.8 to 22.6 °C (Fig. 1), which for Q10 ≈ 2 implies a ~10% decrease in respiration, opposite to the ~80% increase in γ. Soil water content remained above 0.40 m3 m-3 throughout. The simultaneous increase in α and γ, contrary to their expected environmental responses and consistent with their positive co-variation across the time series (Fig. 2a, b), indicates a fitting artifact rather than a physiological change."

Line 521 – 523: The coefficient of determination for the NEE–PAR relationship in the afternoon (R2 = 0.62) is almost double that of the morning (R2 =0.34), indicating that NEE variability is more strongly explained by PAR in the afternoon than the morning.

➔ Slightly rephrase: indicating that NEE variability in the afternoon is more strongly explained by PAR than the variability in the morning.

We thank the reviewer for this comment. We have revised the text accordingly (L531-533):

"The afternoon NEE–PAR relationship was substantially stronger (R^2 = 0.62) than the morning relationship (R^2 = 0.34), indicating that afternoon observations were more consistently explained by PAR."

Line 527 – 534: Many gap-filling methods rely on fitting light-response curves to daytime NEE data under the assumption that PAR is the dominant driver of variability (Reichstein et al., 2005; Lasslop et al., 2010).However, the elevated scatter and reduced R2 of PAR during morning periods indicates that including these data can reduce the model's overall performance. Restricting fits to the afternoon periods when atmospheric mixing is stronger and NEE variability is more tightly coupled to PAR can improve the stability of fitted parameters and reduce bias in gap-filled fluxes.

➔ I would leave Reichstein out of this, because the method does not use PAR.

We thank the reviewer for this comment. This reference has been removed.

➔ I am not convinced that we can make this conclusion yet. Did you test the models' overall performance by comparing the estimate NEE of all three models with the measured NEEs. Because the models performance on all NEE values does not directly depend on the R² relation between the subset of NEE and PAR values. Also it seems like these R² values are only based on te month of Juli, looking at the title of the graph, so is this month representative for all R² valus? You should actively test the overall performance of the different models by calculating R² or root mean squared errors between the predictions and all the NEE measurements. Because

you can have a very tight fit in the afternoon, but if this fit can not predict morning values, then this fit can not be used to predict overall NEE values. So I am not saying that the conclusion is wrong, I think you just need more tests to be able to claim this

We thank the reviewer for raising this important point. We agree that goodness of fit to the subset of observations used to parameterize a light-response curve does not necessarily indicate how well that curve represents NEE across the full diel cycle. To investigate this concern, we conducted an additional cross-validation analysis comparing all-day, morning-only, and afternoon-only fitting approaches. We used five-fold cross-validation by day within 21-day moving windows, evaluating predictions against all held-out, quality-controlled NEE observations. The results confirmed that afternoon-only fits, although providing a good representation of afternoon observations (R^2 = 0.66), did not capture the full diel variability in measured NEE, particularly morning uptake. All-day fits had the lowest overall RMSE. However, we recognize that measured NEE is itself subject to uncertainty, including uncertainties associated with storage flux estimation and turbulent transport, and therefore these comparisons do not independently establish the accuracy of any partitioning approach. We have consequently removed the original claim that afternoon-only fits perform best and have revised our interpretation to distinguish between the goodness of fit within the parameterization period and the ability of fitted relationships to represent NEE across the full diel cycle. We emphasize that our objective is to evaluate how partitioning choices influence estimates of GPP and R_{eco}, rather than to identify a universally superior fitting approach.

562-565 In the context of daytime partitioning, this can bias the fitted intercept upward, potentially overestimating base respiration if the light-response curve interprets these early-morning spikes as baseline flux rather than short-term spikes. Together, these effects suggest that enhanced morning fluxes arise from a combination of physical storage dynamics and ecosystem responses.

➔ Will the method not also use all nighttime data to fit this base respiration? will these morning data points have a stronger influence than all evening and night datapoints?

γ is estimated as the intercept of the light-response curve fitted to daytime data only; nighttime data are not used to fit base respiration. We found that because many of the low PAR events occur in the early morning, these points exert disproportionate influence on γ. The text has been revised to address this (L550-552):

“Because γ is determined largely by the low-PAR portion of the light-response curve, these observations exert disproportionate influence on its estimation.”

The original range was 7.5 to 15.9 µmol m-2 s-1, which was said to be physiologically unrealistic. The new one, from the afternoon methods, is 5.0 µmol m-2 s-1 to 11.6 µmol m2 s-1. In absolute terms the range is smaller, but the values themselves are also smaller, so

in the original, from the smallest to the largest is a multiplication with 2.1, while in the the new values there is a multiplication with 2.3. So I just wonder why the first range is unrealistic, but the second one is not. More the absolute values then?

We agree that the min-to-max range alone does not distinguish the two methods, as both span approximately a 2X range. We have revised the text to base this mainly on the fact that the morning and all day fits result in a physically unplausible values of the initial slope. All-day and morning fits produced slopes up to 0.13 and 0.24 μmol C μmol PAR^{-1}, respectively, exceeding the maximum leaf-level quantum yield of C_3 photosynthesis (~0.05–0.06 mol CO_2 mol $photon^{-1}$; Ehleringer & Björkman, 1977; Skillman, 2008), whereas afternoon fits remained within this limit throughout the record (max 0.06). Because γ is estimated by extrapolating the fitted curve to zero light, physically implausible slopes propagate directly into the intercept. We have revised the text to base the comparison on this rather than on the range of γ (L563-568):

"Although $R_{eco,A}$ still varies approximately two-fold across the time series, the afternoon fits initial slopes remained within 0.025–0.06 μmol C μmol PAR^{-1} throughout the record, at or below the maximum leaf-level quantum yield of C3 photosynthesis (~0.05–0.06 mol CO_2 mol $photon^{-1}$; Ehleringer & Björkman, 1977; Skillman, 2008). In contrast, all-day and morning fits reached unrealistic values of 0.13 and 0.24 μmol C μmol PAR^{-1}, respectively (Fig. 3b). $R_{eco,A}$ values are also consistent with the ~10 μmol m^{-2} s^{-1} typically reported for La Selva (Cavaleri et al., 2008; Chambers et al., 2004)."

Line 597-599: While this approach has been noted to face challenges in complex tropical forests with heterogeneous canopy structure (Hutyra et al., 2008; van Gorsel, 2009), our results suggest it performs well at La Selva compared to other partitioning methods.

➔ You base this conclusion solely on weekly GPP and RECO, and on comparison with other methods that are themselves not fail-proof. Did you check day-to-day variation, diel cycles, light-inhibition, compared Reco values to soil chambers, …? How much did the a and b parameter of the nighttime method fluctuate?

We agree that agreement among partitioning methods more so demonstrates consistency. Soil chamber measurements were not available at our tower during the study period, so we have added the following analyses: 1. Parameter variability: the nighttime model parameters were strongly correlated across windows (r = −0.89), so $R_{eco,N}$ was more stable (CV = 11%) than either parameter alone. 2. u* sensitivity: $R_{eco,N}$ changed by less than 2% when restricted to u* ≥ 0.1 or ≥ 0.2 m s^{-1}, but stricter thresholds left too few data for a robust test. 3. Daily comparison: GPP agreed closely among methods (r = 0.91–0.93), but daily R_{eco} did not (r ≤ 0.26), indicating that the GPP agreement largely reflects shared measured NEE. 4. Diel cycles: composite diel cycles show a decline in nighttime NEE through the night, suggesting residual underestimation under calm conditions. We have changed the text accordingly and now also note that the nighttime method does not account for light inhibition of leaf respiration (L582-589):

"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. GPP_N agreed closely with the daytime-based estimates (Fig. 4d), but this likely reflects their shared measured NEE, so agreement among methods indicates consistency rather than accuracy. Finally, because the method extrapolates a nighttime temperature response to daytime conditions, it does not account for light inhibition of leaf respiration, which may introduce uncertainty into daytime R_{eco} estimates and consequently influence GPP estimates (Wehr et al., 2016; Keenan et al., 2019)."

Line 601: Under these conditions, u* is frequently low (Fig. A2), indicating weak turbulent mixing that can lead to CO2 accumulation near the forest floor, within canopy gaps, and in low-lying areas (Araújo et al., 2008; Goulden et al., 2006). However, the inclusion of profile storage fluxes in our NEE calculations likely mitigates some of the potential underestimation that would otherwise occur when respired CO2 accumulates below sensor height (Montagnani et al., 2018). This may explain why the nighttime method shows strong agreement with the daytime afternoon and sundown approaches (Fig. 4d), despite the theoretical concerns about low nighttime turbulence.

➔ If you performed the u* filtering, this also likely took out all low turbulent moments, hence minimizing underestimation

We thank the reviewer for this comment. 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 (see comment above).

➔ How many nighttime data points were still present after all quality checks? Because for most tropical sites, the reason why the method performs badly, is because there is too little data to fit the curves with high confidence.

Following quality control, a median of 140 nighttime half-hourly observations (range: 22–174) remained per 7-day fitting window. Therefore, the number of available observations was generally sufficient for fitting the nighttime respiration model, likely in part because a u* filter was not applied, as explained above. However, the temperature sensitivity parameter was statistically significant in only 23% of fitting windows, indicating that the relationship between nighttime respiration and temperature was not consistently well resolved at the individual-window scale, despite a mean nighttime temperature range of 5.3 °C.

Line 620 – 621: The slightly lower values for GPPN and RecoN were expected, as this method tends to underestimate fluxes (Keenan et al., 2019).

> Keenan et al 2019 states “The nighttime method will thus overestimate daytime total ecosystem respiration, and consequently apparent photosynthesis”
>
> However the nighttime tends to underestimate the fluxes IF the nighttime respiration is underestimated, so use other reference, eg Lasslop

We have updated the citations and manuscript text to address this (L600-602):

“The relatively low estimates of GPP_N and $R_{eco,\,N}$ may reflect underestimation of nighttime CO_2 fluxes under weak-turbulence conditions, which can bias nighttime-based flux partitioning (Aubinet, 2008; Reichstein et al., 2005).”

Line 621-622 Weekly differences among partitioning methods for the daytime afternoon, nighttime, and sundown partitioning methods were assessed using a one-way ANOVA

➔ Mention all statistics in methods section

This has been updated to a more appropriate method than ANOVA (see next response) and updated in the methodology (L258-263):

“Agreement among partitioning methods was assessed using weekly mean GPP and R_{eco}, computed from half-hourly values over 24 h. Weeks with < 100 valid daytime half-hours were excluded, leaving 30 weeks. For each daytime-based method, paired weekly differences relative to the nighttime method were summarized by mean bias, RMSE, Pearson correlation, and 95% limits of agreement (mean difference $\pm$1.96 SD; Bland & Altman, 1986). Confidence intervals on the mean bias were obtained by bootstrap resampling of weeks (2000 iterations). Because all methods share the same measured NEE, they were treated as paired rather than independent samples.”

Line622-626: For GPP, no statistically significant differences were detected in 79.5% of weeks ($p > 0.05$), while Reco showed no significant differences in 81.8% of weeks, indicating strong agreement among methods at weekly timescales. During our 11-month observational period, none of the three partitioning methods differed significantly for either GPP or Reco ($p > 0.05$).

➔ ANOVA with time series might be dangerous as you violate the independence assumption, so I don’t think you can easily compare the different methods with anova.

Thank you for raising this point. We have removed the ANOVA and instead report paired agreement statistics on weekly values (mean bias, RMSE, correlation, and Bland–Altman limits of agreement), with confidence intervals from a bootstrap over weeks. Daytime-based methods yielded R_{eco} 3.2–5.0% higher than the nighttime method on average, with limits of agreement of approximately $\pm$2 µmol m^{-2} s^{-1}. Because GPP is derived from the same NEE in all methods, GPP differences are equal to R_{eco} differences, which we now state explicitly (see previous comment).

Line 755 – 756 At the daily timescale, predicted LUE declined with increasing temperature, with the strongest temperature sensitivity occurring at PAR>700 µmol m-2 s-1)

And Line 763 – 766 The interaction between temperature and light availability on LUE suggests that warming impacts on tropical forest carbon uptake may be most pronounced during lower-light periods, though further work isolating direct thermal effects from covarying VPD would help clarify this mechanism.

➔ These to sentences seem contradictory. When the largest difference (sensitivity) is at high PAR, then the warming impacts will be most pronounced at high PAR, no?

Yes, thank you for this correction. However, this text has been removed as Reviewer 2 commented that much of the conclusion restated the results and discussion section.

Line 758 – 759: This temperature threshold is lower than observed photosynthetic temperature optima of 28 °C for lowland tropical tree species (Slot & Winter, 2017) and temperature thresholds for GPP in tropical forests (Pau et al., 2018).

➔ Your temperature optima is an averaged daily temperature, while I think this is more frequently determined on a more instantaneous temperature effect. Likely if you would use maximum daytime temperature or something like that, you will get a higher value of temperature threshold. Because likely the photosynthesis is reduced only for a part of the day with high temperatures, above 28/30 degrees, but then the daily average temperature is likely only 25. So it might be more useful to calculate a temperature threshold on half hourly data.

We agree that this would be useful. We repeated the analysis at the half-hourly timescale using the same random forest predictors (Fig. A7; out-of-bag R^2 = 0.34). At this scale, predicted LUE declined most steeply near 27.7 °C, consistent with reported photosynthetic optima of ~28 °C (Slot & Winter, 2017), confirming that the lower daily value (25.5 °C) largely reflects averaging over the diel temperature cycle. We have revised the text accordingly (L727-738):

“At the daily timescale, predicted LUE declined with increasing temperature, most steeply above a daily mean air temperature of ~25.5 °C (Fig. 6d). Because daily means average over the diel

temperature cycle, we repeated the analysis at the half-hourly timescale (Fig. A7), where LUE declined most steeply near 27.7 °C, consistent with photosynthetic temperature optima of ~28 °C reported for lowland tropical tree species (Slot & Winter, 2017) and temperature thresholds for GPP in tropical forests (Pau et al., 2018). The lower daily value therefore largely reflects temporal averaging. Because these predictions hold VPD at its median, the decline reflects a temperature effect beyond co-varying atmospheric dryness, although the strong covariation of temperature, VPD, and PAR at the half-hourly scale limits a full separation of these drivers."

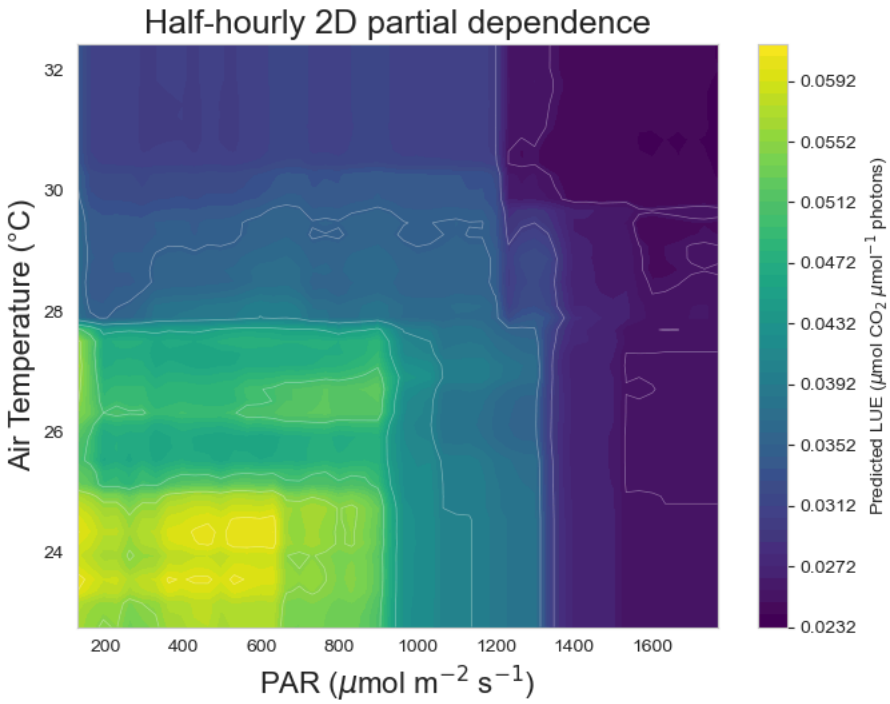


Figure A7. Predicted half-hourly LUE as a function of PAR and air temperature, from a RFM trained on half-hourly data (n = 4205; out-of-bag R^2 = 0.34). Predictors were air temperature, VPD, ΔT, and PAR. Predictions were made across the 5th–95th percentile range of PAR and air temperature, with all other predictors held at their median values. White lines show contours of predicted LUE.

This narrower daily range likely explains why VPD shows limited apparent influence at the daily scale since the model rarely encounters daily mean VPD values approaching the ~1.5 kPa threshold typically associated with significant stomatal downregulation in tropical wet forests (Grossiord et al., 2020).

➔ I can not seem to find the threshold of ~1.5 kPa in the mentioned reference

We thank the reviewer for this comment. We have replaced this with Grossiord et al. (2019), who reported VPD thresholds for sap flux saturation of 0.6–2.0 kPa across 34 neotropical canopy tree species. We rewrote this paragraph for clarity (L748-758):

"VPD showed a pattern broadly similar to temperature (Fig. 6f), with lower predicted LUE when daily mean VPD exceeded ~0.4 kPa at PAR > 700 μmol m^{-2} s^{-1}, consistent with the negative VPD–LUE relationship at the half-hourly timescale (Fig. 6c). In neotropical canopy trees, stomatal regulation of water loss has been observed to begin at a VPD of 0.6–2.0 kPa, where sap flux no longer increases with VPD (Grossiord et al., 2019). Because daily averaging smooths over the diurnal VPD cycle, most daily values fall below this range. The daily decline in LUE at moderate VPD is therefore more likely to reflect covariation with temperature and light on warmer days, when midday VPD exceeds these thresholds, as opposed to direct stomatal limitation at the daily mean VPD. This is consistent with the low importance of VPD in the daily model (Fig. 5e), and findings at other tropical wet forest sites, where VPD effects on carbon uptake become more apparent under more extreme drought conditions (Restrepo-Coupe et al., 2013). VPD also naturally covaries with temperature and other drivers (Bonal et al., 2008; Novick et al., 2016), which may further obscure its independent daily-scale signal."

Line 790: Our results demonstrate that instantaneous environmental limitations (particularly VPD and temperature) regulate LUE at the half-hourly timescale,

➔ They regulate on the daily timescale, consistent with Fig. 5

We thank the reviewer for this comment. This text has been removed per Reviewer 2's comments on conciseness.

Line 795 – 798 This underscores that temporal aggregation itself can obscure sub-daily controls, and scale dependent feature importance should be interpreted alongside the range of each predictor presented to the model at that scale. These findings highlight both the value and limitations of machine learning approaches for understanding ecosystem functioning.

➔ "This underscores that temporal aggregation itself can obscure sub-daily controls" I think this sentence is obsolete as temporal aggregation in essence looses sub-daily information. I think the conclusion here is just that looking into drivers on different time-scales is important. I don't think this is a limitation of machine learning approaches, as this is exactly what was done in the study

We have removed these sentences and replaced them with a conclusion emphasizing the importance of assessing drivers across multiple timescales (L765-768):

"Together, these results show that the dominant drivers of LUE are timescale-dependent, and that assessing environmental controls at multiple timescales is necessary to capture both sub-daily physiological constraints and longer-term environmental variability."

4 Conclusions

Line 800: Our results demonstrate that carefully selecting an EC partitioning method is essential for generating realistic and defensible estimates of tropical forest carbon exchange.

➔ I completely agree here, but in this study all methods were based on similar assumptions, namely: respiration is same during day and night, and only T dependent or event constant and GPP is fitted using light response. So you could argue that this is a rather small selection of methods from which you chose one, and that choice for the afternoon methods was not very strongly supported.

We have revised the sentence to state that method choice within this family mainly affected the physical plausibility of the fitted parameters and identify approaches that don't rely so much on the shared assumptions (Keenan et al., 2019; Wehr et al., 2016) as a stronger test of method choice (L770-774):

"Our results show that the choice of method and fitting window strongly affected the physical plausibility of fitted light-response parameters. However, all methods evaluated here assume that daytime respiration follows the same temperature response as nighttime respiration, without light inhibition. Approaches that relax this assumption, such as partitioning methods that account for light inhibition of daytime respiration (Keenan et al., 2019) or isotopic partitioning (Wehr et al., 2016), could provide additional tests of method choice in tropical forests."

Line 817 – 819 The growing influence of temperature at longer timescales indicates the cumulative effects of sustained thermal and atmospheric conditions on canopy photosynthetic efficiency.

➔ I think this is too strong. The importance of temperature only grows from half hourly to daily, which is still far away from sustained thermal conditions or cumulative effects

We agree this statement was too strong. However, we have cut this from conclusions so as not to restate the results and discussion section per Reviewer 2's comments.

Line 838 – 841: First, the pronounced difference between morning and afternoon light response parameters and substantially higher R2 and lower parameter uncertainty of afternoon fits suggests that the daytime partitioning approach may systematically overestimate base respiration at wet tropical sites where morning CO2 accumulation produces high early morning fluxes.

➔ I don't think you can conclude this. It seems like there is no consistent overestimation of the base respiration as in Fig. 3 c) the orange curve is frequently located above or right on the black curve.
➔ In addition, you never compared the effect of the differences in parameters, on the actual estimated NEE
➔ The fit might be less uncertain, but it is fitted on a smaller amount of data, so the result on the larger data set, might therefore not be better (see previous comment on that specific part)

We agree that the fitted curves in Fig. 3c do not consistently indicate higher base respiration in the morning. We also acknowledge that we did not directly evaluate how differences in fitted parameters affect estimated NEE across the full dataset. Finally, the higher R^2 values and lower parameter uncertainty of afternoon fits should be interpreted cautiously because these fits are based on fewer observations. We have therefore removed the statement suggesting systematic overestimation of base respiration and revised the discussion to emphasize the sensitivity of daytime partitioning parameters to the temporal selection of observations (L800-804):

"First, the pronounced differences between morning and afternoon light-response parameters highlight the sensitivity of daytime partitioning to the selection of observations used for model fitting. Afternoon fits exhibit higher R^2 values and lower parameter uncertainty than morning fits. These results highlight the importance of considering both the temporal distribution of observations and the uncertainty associated with fitted parameters when applying daytime partitioning methods in wet tropical forests."

<u>Small additional corrections:</u>

Line 33: VPD is not yet mentioned, introduce full word.

Done (L33).

Line 76 "Improving GPP estimates is essential for quantifying, understanding, and predicting the fate of tropical forests as net carbon sinks or sources.

Lost sentence (has no contribution here)

Removed.

Line 85: is PAR effect on GPP linear?

Only partially. We have rephrased so that it doesn't seem like the effects are entirely linear (L95-97):

"Light use efficiency (LUE) normalizes carbon uptake by light availability. Because GPP

saturates with increasing PAR, LUE still declines at high light, but normalization reduces the dominant scaling of GPP with light, allowing other environmental controls on carbon uptake to be more clearly resolved."

Line 96: Like most studies in wet tropical forests, this one relied on the nighttime method (Loescher et al., 2003), but comprehensive evaluations comparing multiple partitioning approaches remain limited.

> It is the partitioning methods that relied on the nighttime method
>
> I don't think they use the nighttime method in the paper, because they relate daytime NEE to PPFD

Thank you for this comment. This has been fixed (L108-109):

"Although these measurements provided important insights into carbon exchange at La Selva, comprehensive evaluations comparing multiple partitioning approaches remain limited."

Might include instrumentation list with brand, and measurements height in the supplementary.

We thank the reviewer for this suggestion. We have added a table to the Appendix addressing this.

Line 311-312 LUE values were computed directly from observed half-hourly GPP and PAR measurements prior to model training.

➔ GPP is not observed, is modeled.

Thank you for this comment. This has been fixed.

Line 359: Environmental conditions over the study period provide context for interpreting both GPP partitioning and the environmental controls on LUE, ~~particularly given the relatively narrow temperature range for this field site.~~

Thank you for this comment. This has been fixed

Line 444-446: Compared to other tropical ecosystems, our NEE values fall within the range of NEE estimates found for other wet tropical forest sites (ranging between 95 and –2,534 g C m-2 yr-1; Table 1), with our estimated NEE falling toward the high end of the range.

> Be consistent with minus signs for sink. Also the phrase "our estimated NEE falling

toward the high end of the range" is confusing, because the estimate here is very low (negative).

We thank the reviewer for catching this. We have changed the language surrounding our studies sink to say 'lower'.

Line 456: which dataset?

Thank you for this comment. We have since removed this line from a comment from Reviewer 2.

Line 646 – 647 remove, this is already mentioned

Done.

Line 652 – 654: The variation of GPP as a function of soil moisture and VPD motivates further examination of which environmental drivers most strongly regulate photosynthetic efficiency across timescales, addressed in Sect. 3.4.

➔ This variation is not mentioned yet.

Thank you for highlighting this. This has been fixed (L631-632):

"Given the potential influence of soil moisture and VPD on photosynthetic activity, Sect. 3.4 examines how these and other environmental drivers regulate photosynthetic efficiency across timescales."

Line 656 – 660 remove, already in methods section

This has been removed. The second half of the paragraph was moved to the methods section.

Line 681 – 683: Feature performance metrics at each temporal resolution matched the cross-validated R2 and RMSE values reported above (Fig. 5a–c), since permutation importance was computed using the same cross-validation folds as the model performance evaluation.

➔ Somehow strange formulation, especially feature "performance"

This has been rephrased accordingly (L649-651):

"The cross-validated R^2 and RMSE values associated with each temporal resolution are consistent with those reported above (Fig. 5a–c), as permutation importance was calculated using the same cross-validation folds as the model performance evaluation."

Line 699: or light saturation?

Thank you for highlighting this. This should be light saturation and has been changed accordingly.

Line 700 – 703: This aligns with well-established sensitivities of tropical photosynthesis and ecosystem respiration to sustained warm temperatures, which influence stomatal behavior and multi-day vapor pressure gradients (Lloyd & Farquhar, 2008; Slot & Winter, 2017).

➔ Complex sentence

This has been rewritten (L668-670):

“This is consistent with the sensitivity of tropical photosynthesis and R_{eco} to sustained increases in temperature (Lloyd & Farquhar, 2008; Slot & Winter, 2017).”

Line 720 – 730 move to methods

Thank you for your comment. This has been integrated into the methods section.

Line 799: conclusions should have different number

Thank you for catching this, it has been fixed.