Response to RC2

This study focuses on improving the estimation of Amazonian carbon fluxes, particularly the Net Biome Exchange (NBE), which encompasses biogenic and wildfire fluxes. The authors highlight the challenges in quantifying Amazonia's carbon balance due to anthropogenic disturbances and the need for more reliable long-term data. To address this, they utilize solar-induced fluorescence (SIF) data from NASA's OCO-2 satellite and other observations to enhance the Vegetation Photosynthesis and Respiration Model (VPRM). They compare different VPRM versions and the Simple Biosphere 4 (SiB4) model and further optimize these models using OCO-2 CO2 column observations. The study reveals that SIF-based VPRM versions, especially VPRM_SIFg, outperform traditional ones in capturing CO2 fluxes across various timescales and moisture conditions. The researchers underscore the importance of SIF in improving carbon flux estimations and understanding Amazon's response to environmental changes.

We thank Reviewer 2 for the extensive comments, including ways to make both this current work and future follow-on work more robust. Where changes in the paper have been made, comments have been included as "RC2.#".

There are some issues that, if the authors address, can make the study more robust:
RC2.1 The VPRM model calibration relies on a limited dataset from eight eddy flux sites, potentially hindering the model's ability to represent the diverse and heterogeneous Amazonian ecosystems accurately. (if possible) expanding the calibration dataset to include more sites and diverse vegetation types would improve the model's applicability and robustness.

We agree but unfortunately, at the time of this work, the eddy flux data set was limited in time and location to the sites as described in the paper. However, this is a topic of ongoing development: in progress VPRM-SIFg development will incorporate newly available data sets representing interior Amazon forest (to supplement K34) and lowland Amazon forest where model-observation mismatch is highest at least as suggested by the March 2016 case study. We now note this in Section 3.2.1 and also in the conclusions:

Sec 3.2.1: *The optimization suggests that all prior models in wet season 2016 tended to underestimate carbon uptake in the western Amazon and particularly in the Amazon lowlands;. The regions of underestimated uptake correspond to the anomalously wet region identified through the Self-calibrated Palmer Drought Severity Index (ScPDSI) by Jimenez-Muñoz et al. (2016), but also in the case of the VPRM to regions under-constrained by eddy flux calibration sites (Fig. 2a; Fig. 6a).*

Conclusions: *Currently, Br-K34 is the only site representing the interior Amazon; additional eddy flux data for calibration and/or validation such as from Amazon Tall Tower Observatory (ATTO) would be beneficial. In addition, eddy flux data from the Amazon lowlands would* *provide additional constraints to the higher model-observation mismatch observed in that region.*

RC2.2 The available eddy flux data used for calibration spans a period from 2001 to 2015, which is offset from the study period of 2014-2020. This temporal mismatch might affect the model's accuracy in capturing recent carbon dynamics, as ecosystem states and environmental conditions could have changed over time. Therefore, it becomes essential to incorporate uncertainty in VPRM parameters in inversion runs. These uncertainties can be obtained from the hessian of non-linear optimization procedure.

Agreed, and we had noted this limitation in the text below. We also fixed the 2014 typo in the paper to 2010 (fluxes are generated from 2010 not 2014)

Second, while the available eddy flux data can be offset up to thirteen years prior to the study period of 2010–2020, the major drivers of hourly ecosystem flux variations are provided by T_{air}, PAR, surface reflectance indices, and SIF. That is, while the static ecosystem parameters of λ, PAR_0, α, β, and γ would benefit from tuning to eddy flux data seasonally and/or over the entire study period to reflect concurrent ecosystem states most accurately, the real time variation is dominated by T_{air}, PAR, surface reflectance indices, and SIF (Dayalu et al., 2018).

However, we have also expanded the conclusions accordingly to note the concerns from RC2.2:
Future work will optimize GPP and R_{eco} separately; in that case, the VPRM can be optimized in parameter space (e.g., Matross et al., 2006) which will also account for the uncertainty associated with using carbon dynamics from 2000-2010 to describe carbon dynamics from 2010-2020.

RC2.3 Assessment of transport model errors is required to understand their influence in determining the coherence between observations and fluxes. This can be done through ensemble runs. The authors can also convolve the posterior uncertainty to see the envelope of uncertainty surrounding true observations and convolved observation that is forward operator*posterior fluxes.

We agree with this and note that a full model optimization (for the entire 2010-2020 time period and for an in-development expansion from 2003-2023) is being conducted. This will include a formal assessment of model transport errors. The scope of the current work was flux model development, with a small performance-testing case study. We had noted as such in Section 2.4:

In our application we also neglect transport model errors, which are difficult to quantify but can be further explored in the future using transport ensembles. Errors in WRF-STILT transport are discussed in Rastogi et al., (2021) in terms of calculating X_{CO2}^{sim} for OCO-2 XCO_2 retrievals in North America. They found that low partial column bias relative to *independent vertical profile CO_2 data show that errors in WRF-STILT transport contribute very minimally to bias in X_{CO2}^{sim}.*

RC2.4 While the study briefly touches upon the impact of fires on carbon fluxes, a more comprehensive evaluation of fire influences, including the effects of fire severity and frequency on ecosystem recovery and carbon balance, would provide a deeper understanding of the Amazon's response to fire disturbances. This can be done by OSSE kind of run to see its impact on the flux.

Agreed. Once again we note that the scope of this work was to develop a sufficiently representative biogenic flux model that could then be reliable combined with a fire emissions model to complete the picture of impacts on carbon fluxes. The test done here was to find which biogenic flux model provides a realistic representation of the vegetation carbon dynamics (absent confounding influences of fire emissions). In-progress work is developing a CO2 fire emissions inventory from an amazon-specific CO-fire emissions inventory.

RC2.5 Finally (this is the biggest issue), the study primarily focuses on a wet season case study in 2016 and lacks a long-term validation of the improved VPRM model against independent observations. Conducting a multi-year validation using additional data sources, such as atmospheric CO2 measurements or biomass inventories, would strengthen the confidence in the model's performance and its ability to capture interannual and seasonal carbon trends.

Agreed, and this was a big gap in the analysis that was also identified by RC1. We have rectified this by including a new section and an additional figure:

We have included additional analysis – namely, we have created a new Figure 10 and compared with the most relevant and recent work by Gatti et al. (2021) (Amazon basin as a whole) and the Cerrado and Caatinga region (also incorporating eddy flux site results from Mendes et al. 2020 and Alves et al. 2021). We have included a summary of these new results in an additional Section 3.3.3. Comparison with interannual observations. The section and figure are reproduced below. Note that we display the 95% CI in Figure 10b as displaying the larger IQR drowns out the median signal on the plot. Also note that Gatti et al. (2021) report fluxes as Total Flux – Fires ≈ NEE (but not exactly, as the total flux would also include river efflux). However, it's a fair quantity for comparison as NEE dominates that signal.

3.3.3 Comparison with interannual observations

We assessed the performance of VPRM_SIFg and SiB4 from 2010 to 2019 for the Amazon basin (Amazon mask; Fig. 10ab) and separately for the region containing the Cerrado and Caatinga biomes (Cerrado+Caatinga mask; Fig. 10ac) and compared against available observations.

For the Amazon mask, the VPRM-SIFg prior tends to estimate interannual net release while the SiB4 model tends to remain closer to neutral (Figure 10b). In addition, the VPRM-SIFg describes greater ecosystem heterogeneity relative to SiB4: the interquartile range (IQR) over the Amazon for the VPRM-SIFg is -0.47 to 0.83 g C m^{-2} d^{-1}. In contrast the SiB4 IQR is 0.06 to 0.07 g C m^{-2} d^{-1}. Meanwhile, the Gatti et al. (2021) mass balance approach using aircraft vertical profiles tends to estimate net fluxes closer to neutral that generally track SiB4 interannual estimates with a few notable exceptions: in 2016, corresponding to the tail of the severe 2015-2016 El Niño; aircraft profiles suggest a regional net release of 0.1 g C m^{-2} d^{-1} in agreement with VPRM_SIFg, while the following year shows a net regional uptake of -0.2 g C m^{-2} d^{-1}. We note that the VPRM_SIFg model agrees with the trajectory of the Gatti et al (2021) post-El Niño fluxes in that there is more net uptake implied between 2016 and 2018. Furthermore, we note that the 2010-2011 El Niño corresponds to a VPRM_SIFg estimate of net release, while Gatti et al. (2021) and SiB4 estimate carbon fluxes that are net neutral to uptake. Given the severity of the associated 2010 drought across the Amazon, particularly as it was only five years after the previous severe drought, it is worth exploring whether the VPRM_SIFg is better able to capture the regional carbon effects and impacts of antecedent environmental stressors.

The performance in the Cerrado and Caatinga region suggests that the ecosystem heterogeneity exhibited in the VPRM_SIFg model is realistic. The IQR for the VPRM_SIFg in the Cerrado and Caatinga region captures the site diversity exhibited by the Mendes et al. (2020) northern Caatinga eddy flux site and the Alves et al. (2021) southern Cerrado/converted pasture site. In contrast, the IQR of SiB4 remains closer to neutral. Note that the Gatti et al. (2021) analysis did not include an assessment of the Cerrado and Caatinga regions.

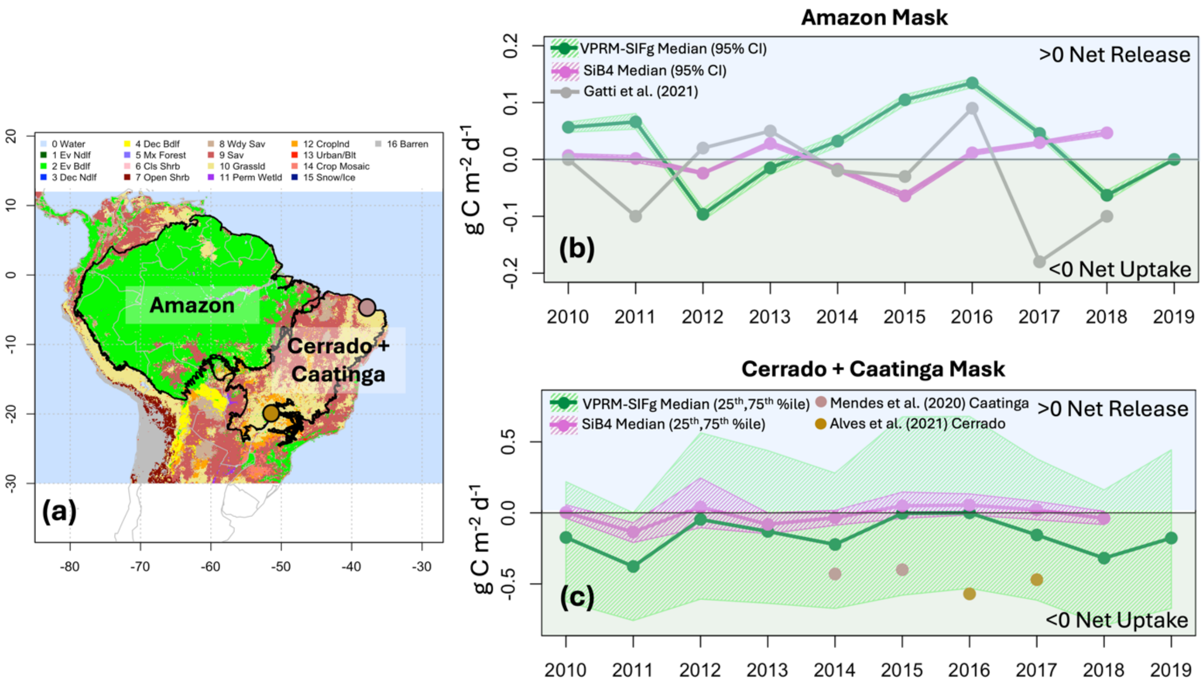

Figure 10. Interannual performance of VPRM_SIFg and SiB4 NEE (g C m^{-2} d^{-1}) relative to available observations for the decade beginning in 2010. (a) IGBP land use map overlaid with the Amazon mask, Cerrado + Caatinga mask, and two Cerrado and Caatinga eddy flux sites used for comparison; (b) VPRM_SIFg and SiB4 median annual NEE (95% CI of the median) for the Amazon mask along with estimates from Gatti et al. (2021); (c) VPRM_SIFg and SiB4 median annual NEE (25th, 75th percentiles) for the Cerrado + Caatinga mask along with annual estimates from two eddy flux sites.

Other comments:

RC2.6 How are the authors dealing with negative SIF values in their models? If they keep them, GPP will become positive in SIF-based equations. The authors technically are not completely replacing EVI and other scalars by SIF as CSIF is itself derived from MODIS reflectance. It would be good to know how does SIF directly obtained from OCO-2 perform in the VPRM models. Why rely on CSIF when SIF is directly available from OCO-2? Note that OCO-2 SIF also comes with uncertainty, whereas CSIF does not include uncertainty estimates. If there are cloud cover issues, then the distribution of CSIF can be compared with OCO-2. Also, if there are cloud cover issues, then CSIF is mainly influenced by MODIS. I suggest a few things authors can do:

- **Replace CSIF (derived from OCO-2) with OCO-2 SIF (native OCO-2 SIF)**
- **Compare the ECDF of CSIF with OCO-2. Check if they are similar or not. Use two samples, Anderson-Darling or other statistical tests, to ensure they carry the same information. If they are statistically different, then making any conclusions about SIF improving VPRM estimates would be difficult.**
- **Run this analysis with CSIF, GOSIF, and other SIF products. Please also check the annual variability in CSIF.**

All this is required to make sure that acceptance of the new model is not an artifact of CSIF.

OCO SIF was not available prior to 2014, and our study was looking at a decade of biogenic model performance from 2010-2020. For model consistency and to provide the basis for a SIF-based flux model that can be extended to the early 2000s and therefore enable multi-decade carbon trend analyses (this is currently in progress work all the way back to 2003), we wanted a consistent SIF field. The work you mention has already been done in the CSIF paper by Zhang et al. (2018a). Future work will explore a wider range of SIF products, and we have included this statement in the conclusions:
Future work will continue development of the VPRM_SIFg formulation, including further investigating the model structure as it relates to SIF and PAR as well as exploring the direct use of SIF satellite products rather than derived products such as CSIF.

RC2.7 Evaluation of VPRM models against observation is OK, but it depends on uncertainty. Clarify this. Show the posterior flux match of each of them against observations and whether they are within each other uncertainty bounds or they are outside uncertainty bounds, in which case they can be rejected outright.

Apologies, but we do not understand this comment and were unable to follow-through with a response. To what section are you referring? Are you referring to the optimization? To the aircraft vertical profile calculation? To the eddy flux data?

RC2.8 The study could benefit from adding a flowchart to provide a clear visual representation of its methodology. This would be particularly helpful in understanding the complex workflow and interconnections between the different components of the study, such as data processing, model calibration, regional inversion, and model evaluation.

We have created a flowchart and added it to the SI as Figure S4. The flowchart is reproduced below.

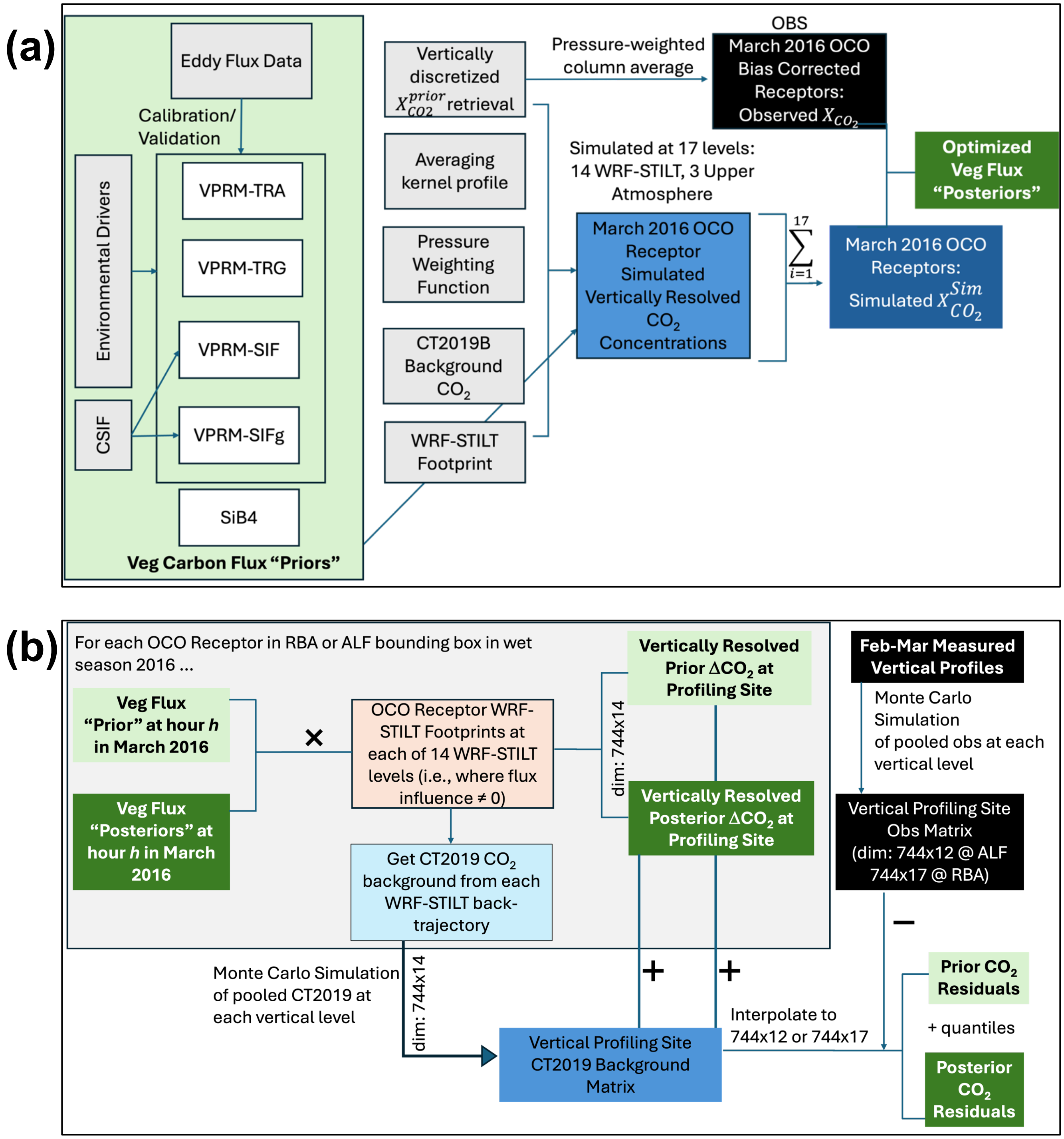

Figure S4. Flowchart of methodology. (a) Overall methodology; (b) aircraft vertical profile site simulation and comparison.

RC2.9 Many of the steps the authors took for their assessment need to be formalized to understand better what is being done. For example, "We bootstrap CT2019 background concentrations and vertical profile measurements at each vertical level to dimensions that enable merging with the month of hourly ....." *I was utterly lost here. I do not know what is being done. Line 360 to the end of the methodology section requires a significant rewrite for clarity. Have a clear flowchart + equations + Jupiter notebook. How is this all connected to equation 10. What kind of bootstrap is it?*

We have clarified and re-written Section 2.5, encompassing the area of confusion beginning in Line 360 (reproduced below). Thank you – the section now reads much better. In addition, we have included and equation in the main text for clarity, and also added in a flowchart as a Figure S4 in the supplemental information.

2.5 Model Aircraft Vertical Profile Simulation

We calculate prior and posterior modelled vertical profiles for SiB4 and each VPRM formulation for our March 2016 optimization period at locations roughly corresponding to the vertical profiling sites RBA and ALF displayed in Fig.1a. As the available OCO-2 receptors are not identical in space and time to locations of RBA and ALF profiling sites, a direct comparison of simulated prior and posterior CO_2 vertical profiles derived from convolving multi-level OCO-2 receptor footprints and vegetation flux models is not possible. In addition, RBA and ALF vertical profiles are typically obtained once or twice a month such that robust measured monthly averages for a single month are not available. To allow for direct comparison between modelled (prior and posterior) and measured vertical profiles for all 744 hours of March 2016, we construct pooled datasets that occasionally combine February and March 2016 measurements and/or modelled fields to develop a dataset that adequately represents a typical 2016 wet season month. Given the seasonal similarities across the Amazon in February and March, combining data across these months to create a representative "typical wet season month" is reasonable. Our method is detailed below.

We first construct a pooled dataset of "typical wet season 2016" measured vertical profiles at each of RBA and ALF sites. For RBA, we combine all measurements obtained between February and March 2016 (2016-02-08 at 1630UTC; 2016-02-27 at 1645UTC; and 2016-03-17 at 1700UTC), resulting in three measurements at each of 17 vertical levels between 300 and 4500 m asl. For ALF, we combine all measurements obtained between February and March 2016 (2016-02-23 at 1630UTC; 2016-02-29 at 1540UTC; 2016-03-13 at 1600UTC; and 2016-03-30 at 1540 UTC), resulting in four measurements at each of 12 vertical levels between 450 and 4500 m asl. For each site, we conduct a Monte Carlo simulation of measurements at each vertical level to obtain measured concentration matrices of 744 h x 17 levels (RBA) or 744 h x 12 levels (ALF). Second, we assess the availability of OCO-2 footprints in the vicinity of RBA and ALF sites from February to March 2016. Our goal was to obtain footprints sufficiently close to each profiling site to be representative of the near-field influences on the site, but also have a large enough bounding box so that at least two OCO-2 receptors and their footprints were present for transport uncertainty calculations. Figure 2a displays the final selected 5x5 km bounding boxes around each of RBA and ALF and the representative OCO-2 receptors. RBA and ALF simulated vertical profiles are then *derived from the OCO-2 receptor footprints bounded in each box (Fig. 2a). Third, we use NOAA's web-based HYSPLIT model to assess land surface influences on each of the measured vertical profile dates and compare them with the land surface influences on each of the selected OCO-2 receptors. Figures S2-S3 show that the land surfaces influencing both the simulated and actual profiles at RBA and ALF are comparable with the airmass trajectories representing typical seasonal prevailing winds in the February/March 2016 time frame and annually from 2010–2018; the average air mass trajectories are displayed in Fig. 2a. Fourth, we obtain CT2019 background concentrations associated with the vertical level of each airmass back trajectory before it enters the study domain. We capture background uncertainty by pooling all CT2019 concentrations at each vertical level, and conducting a Monte Carlo simulation resulting in a CT2019 matrix of 744 hours x 14 WRF-STILT vertical levels. Next, we use the March 2016 prior and posterior hourly fluxes to estimate a "typical wet season 2016 month of fluxes" and convolve them with WRF-STILT footprints from each of the six OCO-2 receptors in the RBA and ALF bounding boxes to obtain a spread of enhancements or depletions relative to the incoming CT2019 background CO_2 from 10 days prior. This results in a simulated CO2 concentration at each of 14 WRF-STILT vertical levels for each of 744 hours in March 2016 (Eq. 12). Finally, we linearly interpolate all components from WRF-STILT vertical grids to each of the RBA and ALF measured vertical profiles which enables us to calculate hourly model-observation residuals at each vertical level and extract means, 25th, and 75th percentiles. Ultimately, vertical profiles of modelled and measured residuals for the simulated month typical of the 2016 wet season incorporate uncertainties in transport, background, vertical profile measurements, and flux fields. Figure S4 displays the overall methodology in a flow chart.*

$$Simulated\ CO_{2,level=z,hour=h}(ppm) = Footprint_{z,h} \times Flux_h + CT2019_{z,h} \qquad (12)$$

RC2.10 Line 330 "The footprint domain is outlined in Figure Error! Reference source not found.b". Correct this.

Fixed.

RC2.11 In Figure 3. VPRM Model-Observation (Night-time) Respiration residuals (I think this is a boxplot). It would also be good to see this as a frequency or histogram plot, as clearly, the bars are not uncertainty estimates. Therefore, we need to know the proportion of residual per/quantile. The authors should explain the relevance of these results in the caption.

This has been fixed. We have re-displayed the boxplot as violin plots so that both the data distribution and deviation from normal is also apparent (i.e., combining the information in histograms and standard boxplots in one plot). We have also edited the caption and the text to summarize the key take-aways/relevance:

Sec 3.1:

Overall, we find that annually and by season, the SIF-based VPRM formulations—and VPRM-SIFg in particular—have less skewed distributions and lower overall bias than the traditional formulations (Fig. 3, Table S3). On average, the dry season respiration bias is lower across all model versions than in the wet season. In the wet season, while all models tend to underestimate respiration, VPRM-TRA and VPRM-TRG display the greatest bias, with VPRM-TRG displaying the greatest skew. In both seasons, the VPRM-SIFg formulation exhibits the lowest respiration bias with a residual distribution closest to normal. The underestimate in wet season night-time respiration also implies that the general underestimate of peak daytime wet-season drawdown in NEE occurs through underestimating GPP rather than through over-estimating R_{eco}. Of the VPRM formulations, the SIF-based formulations have more instances of overestimating respiration, especially in savanna ecosystems (Table S3).

Associated figure edit:

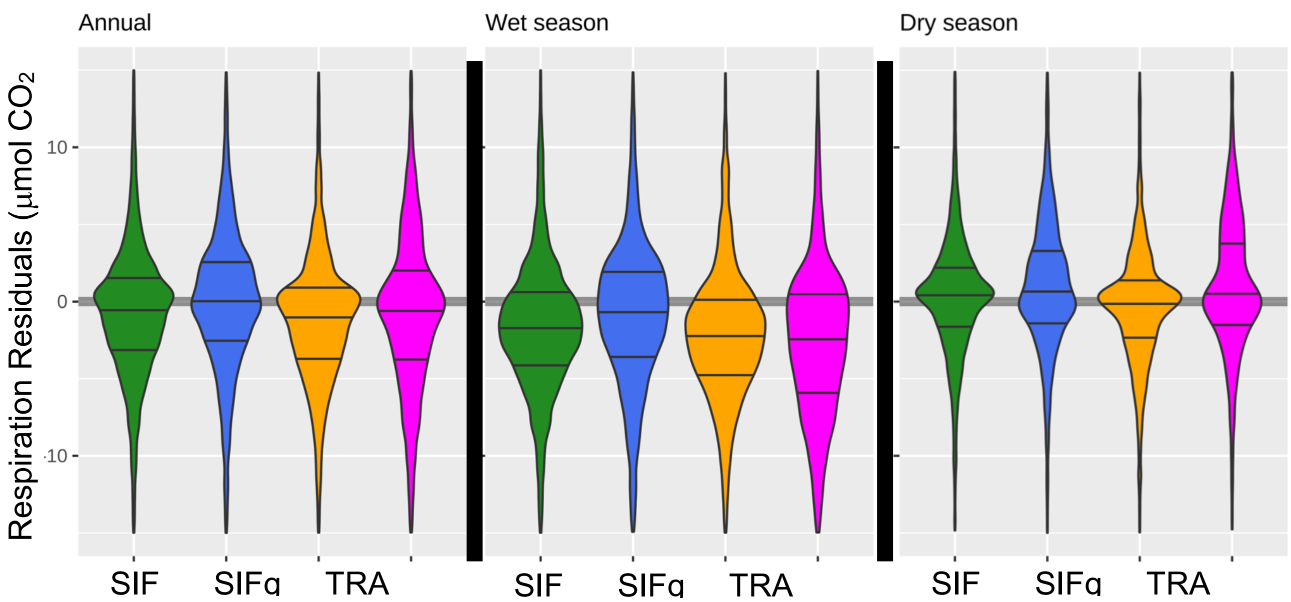

Figure 1. Violin plots of VPRM Model-Observation (Night-time) Respiration residuals, annually and by season at eddy flux sites (μmol CO_2 m^{-2} s^{-1}). Nighttime NEE (where GPP = 0) is used to approximate respiration. Lines are 25^{th}, 50^{th},, and 75^{th} quantiles. All models tend to underestimate wet season respiration and overestimate dry season respiration. Data skew suggests that VPRM_SIFg respiration residuals are the closest to a normal distribution.

RC2.12 In Figure 4. Panel(a) bars are not uncertainty estimates. They incorrectly imply uncertainty when it is something else. Clarify and, if possible, plot them in a way so that people, by just looking at them, do not think that these are estimates of uncertainty

Apologies for that; we have clarified that they are the 1-s standard deviation from the NLS fitting of parameters.