Referee comments 1 RC1

This study provides excellent insight into spatial heterogeneity of GHG fluxes in a fairly unprecedented sampling extent of chamber-based fluxes. The authors compensate well for the lack of temporal coverage in fluxes by leveraging the spatial information available through a hierarchical Bayesian modeling approach to NEE. The remote sensing analysis is thorough. The comparison of scaling models provides useful insight into important multivariate influences on GHG fluxes.

Response: Thank you for this feedback!

I have two areas of significant improvement to address. The first is a lack of information provided about the Bayesian hierarchical modeling. The authors reference Williams et al. 2006 for their model structure. This should be explicitly provided, along with the parameters that are being estimated, at least in the supplement. The authors mention using vague prior information for these parameters. The prior distributions and initial values used for the MCMC chains should be provided (at least in the supplement). There is no mention of posterior predictive checks or tests for convergence. These are necessary to ensure the model is appropriate for the data and that parameters are estimated correctly (without need for a longer burn-in for example). There should be some presentation of the posterior distributions for parameter estimates. There could be some additional discussion as well related to how much the random effect of plot contributed predictions, or how variable the random effect was within vegetation types, etc.

Response: Thank you for pointing this out. We will clarify that ER, maximum GPP, half-saturation constant, and an exponential air temperature response of ER were the parameters in the main text.

We will add the following details to the Supplement:

“We used weakly informative priors for our parameters, informed by those reported in Williams et al. 2006 and Happonen et al 2022. The means and standard deviations for the priors were 1 and 2 for the logarithm of the ER (posterior: 0.65 and 0.51), 0 and 1 for the temperature effect on logarithm of the ER (posterior: 0.02), 10 and 10 for maximum GPP (posterior: 8.54 and 5.51), and 6.2 and 0.3 for the logarithm of the half-saturation parameter (posterior: 5.92 and 0.48). We used logarithms for some parameters to normalize their error distribution.

The Bayesian R^2 of the model was 0.96, four out of five family- and population-specific mean parameters had an Rhat of 1.00, and posterior predictive draws matched well with the observations (Fig. X.), indicating model convergence and good predictive performance. For more details on the model structure, see Happonen et al. 2022 (section 2.4.1) and the code light_response_model.R in Virkkala et al. 2023.”

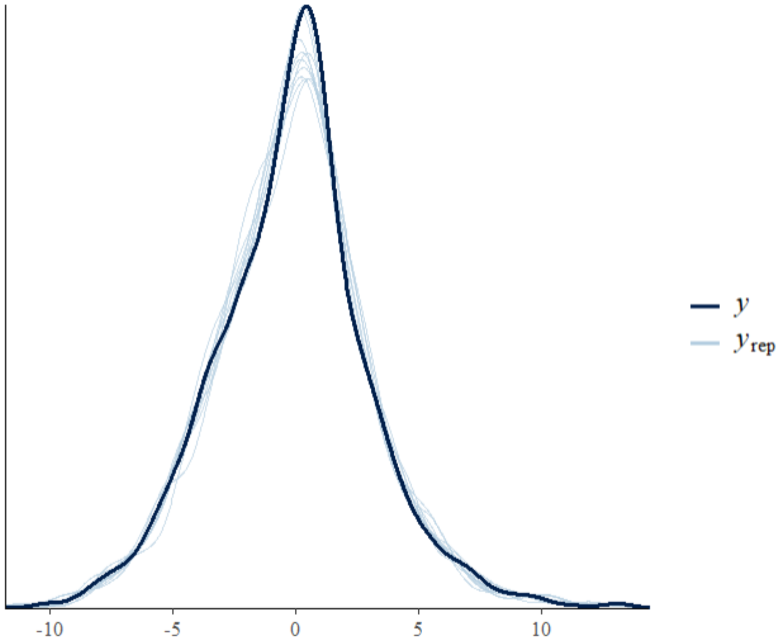

Fig. X. The distribution of NEE observations (y) and draws from the posterior predictive distribution (yrep) in micromoles CO_2 m^{-2} s^{-1}.

The second area for improvement is related to using back-transformations of log-transformed predictions. From my understanding, soil C, biomass, and soil moisture were log-transformed during their upscaling. Then they were back-transformed and subsequently used as drivers to predict GHG fluxes. Back transforming a prediction (from a non-affine transformation) will introduce bias that needs to be corrected. For a useful explanation of the problem, see this blogpost: https://florianwilhelm.info/2020/05/honey_i_shrunk_the_target_variable/. There are multiple methods available for correcting back-transformation bias, some of which are analytical such as in the case of simple linear regression. See this paper for a comparison of several bias correction methods for GBM models: https://kdd-milets.github.io/milets2022/papers/MILETS_2022_paper_0925.pdf. Since all three of the back-transformed variables rank as fairly high predictors, and are especially important at high soil C, high soil moisture, etc where the back-transformation bias is larger, it is critical to correct this bias. The CH4 flux scaling similarly needs a back transformation bias correction, since a cube-root transformation is also non-affine, and these predicted fluxes are subsequently back transformed for comparing to in situ fluxes and calculating carbon budgets.

Response: We agree that back-transformations are problematic because they impact both the error distribution and the shape of the regression. While this can certainly be an issue for the environmental datasets (soil C, biomass, and soil moisture), we think that the largest uncertainties in our study would be with the back-transformation of the CH4 flux variable as it was one of our main GHG flux response variables. We will analyze how these transformations impact our main conclusions as well as explore potential ways to correct for the biases in the following ways:

1) For the transformed variables, we will compare the range of values in the observations and predictions to understand if the summary statistics (e.g., maximum and minimum values) in each vegetation type are different to see if there are potential issues.

2) We will test how the CH4 flux predictions change if no CH4 flux transformations are being made. Technically, machine learning models should be more flexible with the distributions and assumptions than linear regressions and therefore might work fine without transformations. We will compare the distribution of residuals, model predictive performances, and the resulting predictions.

3) We will test correcting for the bias with a non-linear model.

Based on these tests, we will either 1) keep the transformed variables as they are with evidence highlighting that backtransformations did not impact our final conclusions, 2) apply a bias-correction as suggested by the referee, or 3) remove all transformations after a careful re-examination of the model performance of the models without transformations.

Minor comments:
There are numerous regressions demonstrating model performance (Fig 4, FigS3), with the r-squared reported. The slopes intercepts, and p-values should also be reported, as this would help assess performance and bias in the model predictions.

Response: We will add these statistics to the manuscript.

Font sizes for Fig 3 are too small.

Response: We will reorganize the figure and remove some less important boxplots to make the fonts bigger.

It is unclear what the 'Agency 2017' reference is in Fig 1. It is also unclear what the colored vegetation boxes correspond to in panel (c) of Fig 1.

Response: We will correct and clarify these.