Response to reviewer's comments

Editor

We have received two reviews. Please note that one referee was unable to review the manuscript a second time. A new (expert) reviewer was able to review the manuscript. Both referees agree the work was greatly improved with respect to the former comments. Referee #3 has a major comment that should be evaluated. I think there is value in the study's detailed comparison between model data and field observations at the interannual scale. However, they do have a point about which datasets we can draw physical insight from in interpreting these interannual results. Before publication, we recommend major revisions to address this point from Referee #3. Please also consider the other detailed points from both reviewers.

We sincerely thank the Editor for the opportunity to publish our study in *Biogeosciences*, and we are grateful to Referee #3 for agreeing to review our manuscript. The reviewer's suggestions offer valuable insights into the interannual variability of CO_2 fluxes. However, they diverge from the direction and focus emphasized by the first two reviewers. While we acknowledge the relevance of these comments, we believe they extend beyond the scope of our paper, which primarily aims to evaluate model performance and its complementarity with observational data and to demonstrate seasonal shifts in the relationship between climate and CO_2 fluxes at the interannual timescale.

Furthermore, the limited number of flux tower sites in European forests, combined with the relatively short duration of the available datasets, currently limits our capacity to fully address the reviewer's suggestions within the framework of this study. That said, we have made efforts to incorporate Referee #3's remarks, notably by expanding the discussion and rewording certain sections of the manuscript to better reflect these points.

Please find below our detailed, point-by-point responses to the reviewer's comments. Kindly note that the line numbers mentioned in our responses refer to the clean version of the manuscript (without track changes).

Anonymous Referee #2 (Report #2)

This is a second review report of the manuscript "Climate impact on mean annual cycle and interannual variability of CO2 fluxes in European DBF and ENF forests: insights from observations and state-of-the-art data-driven and process-based models". I found the authors have largely improved their work by adding comprehensive site measurements and data-driven and process-based products, as well as using those data at their best spatial resolution to more accurately compare with site data. The authors also implement multiple hydro-climate drivers in understanding the model behavior. I have a few more points and would like to suggest a publication after minor revision.

1. In Figure 3, why do all winter months in European sites have positive soil moisture, especially some months the temperature is below -5 or -10?

ERA5-Land soil moisture represents the total volumetric water content in the soil, including both liquid water and ice. This is why soil moisture values remain above zero even when temperatures fall below freezing. This clarification has been added in Section 2.2.4 (l281-283).

2. Can the authors give any hypothesis that why all products have systemic underestimated NEE interannual variability while except for CarbonSpace?

The potential reasons that could explain the better performance of the CarbonSpace model in capturing the observed NEE interannual variability are now discussed. It now reads (l620-627): “The interannual variability in the models is weaker than in the observations, consistent with previous studies (e.g., Nelson et al., 2024). However, the CarbonSpace data-driven model proved to be the only model tested that does not underestimate the NEE interannual variability. The reasons may involve its high spatial resolution (few hectares) and the use of a Lagrangian particle dispersion model, which allows it to closely align with the flux tower footprints. This results in more precise flux localization, which may improve its response to fine-scale variability. They may also involve the use of an ensemble tree method for regression. This method offers greater flexibility in capturing nonlinear interactions between environmental variables and NEE. Further studies are needed to evaluate these hypotheses”.

3. How about these model performance in representing monthly anomalies when removing the seasonal cycles? This is also very relevant to guide product users and further model development.

Thank you for the suggestion. We added an analysis (new Fig. 9), which reveals weak impact of climate on NEE when all months are analyzed together. This apparent lack of influence arise from (and hides) distinct seasonal patterns. We modified Section 3.3.2 accordingly (l569-586).

Anonymous Referee #3 (Report #1)

In this revised manuscript, the authors have made substantial efforts to improve the original manuscript by incorporating most of the comments from previous reviewers. The key improvements are 1) shifting the research objective from evaluating models' performance to examining the climate impact on the mean annual cycle and interannual variability; 2) increasing the number of eddy covariance sites from three to 19; and 3) including more process-based and data-driven models. However, there is still one major issue regarding the interannual variability that I would like to see the authors address.

Thank you for taking time to review the two versions of the paper.

The major comment:

The authors have clearly shown that all process-based and data-driven products struggle to capture the magnitude and variance of the interannual variability of NEE (Figures 6 and 7).

This is a long-standing and known issue at least for data-driven products (Jung et al., 2020; Nelson et al., 2024). But they still claimed that these products can be used to analyze the climate impacts on the interannual variability of NEE. It is hard for me to follow this logic.

The rationale is that the magnitude of CO_2 flux interannual variability (as measured by the standard deviation) is largely independent from its temporal co-variability with climate (as measured by correlation). In other words, a model may exhibit a realistic magnitude of variability, while still failing to capture the correct climate-driven year-to-year fluctuations. Conversely, a model might underestimate (or exaggerate) the observed magnitude of interannual variability, but successfully capture the correct climate-driven year-to-year fluctuations. Therefore, we treat the magnitude (standard deviation) and temporal co-variability (correlation with climate) as distinct, uncorrelated aspects. It now reads:

- l311-313: "Two complementary metrics are used for model evaluation: the bias (model minus observation), which assesses errors in magnitude, and the Bravais-Pearson correlation coefficient (R), which evaluates temporal co-variability. These metrics capture distinct aspects of model performance and are not necessarily correlated."
- l463-465: "Importantly, a biased magnitude of CO_2 flux interannual variability (as measured by the standard deviation) does not preclude the models to capture their temporal co-variability (as measured by correlation) with observed CO_2 fluxes and climate."

Furthermore, models remain essential tools for assessing the relationship between climate and CO_2 fluxes at continental (e.g., Europe) to global scales. Although these models carry uncertainties related to input data, parameterization, and structural assumptions, they still provide valuable insights into the climate–CO_2 flux relationship. They are useful for identifying large-scale patterns, though they are not intended to accurately reproduce flux magnitudes or seasonal dynamics at specific sites. At the local scale, we acknowledge that direct climate measurements and site-specific observations offer more accurate information, provided that long-term data are available.

I agree with the authors that the length of some eddy covariance sites is probably not long enough to allow examining the interannual variability of NEE, but I do not think using long-term but problematic model products could help here. Highly inconsistent patterns among products in Figure 9 acknowledge this point.

This is true that patterns in Fig. 9 (now Fig. 10) are not identical across models, which is better acknowledge in the new version of the manuscript. While a single model cannot comprehensively capture the pattern, a model community (an ensemble of multiple models) at least capture the range of possible patterns and gives a common sense on patterns. In this study, several consistent features emerge between the multi-model ensemble and the observations. In winter and fall, NEE tends to be positively correlated with temperature, soil moisture, and VPD. In spring, NEE generally shows negative correlations with temperature and VPD, but a positive correlation with soil moisture. During summer, most models exhibit a negative correlation between NEE and soil moisture, as in the observations. Overall, the direction of the NEE–climate relationships is fairly consistent across the datasets.
We modified Section 3.3.2 to highlight common patterns and uncertainties (l569-586).

Nevertheless, examining the climate impact on the interannual variability of CO2 fluxes at the monthly scale is an interesting question and could improve our understanding of the ecosystem carbon cycling. Therefore, I would suggest that authors focus on eddy covariance data and completely give up using models, at least for the analysis of the interannual variability, which, of course, deviates a lot from the authors' initial idea, but would gain more scientific insights. In fact, the required length to represent the interannual variability is site-specific (Chu et al., 2017), therefore, authors could use the methods in Chu et al., 2017 to identify the sites with adequate temporal representativeness. If not enough sites are available, authors could extend the vegetation types beyond DBF and ENF, and/or extend the research area beyond Europe to increase the number of long sites.

This paper is part of a broader project aimed at evaluating the ability of models to capture CO_2 fluxes between European forest ecosystems and the atmosphere. In response to previous reviewer suggestions, we have made a huge effort to include additional data-driven and process-based models in this study. While uncertainties remain, the model results reveal clear and coherent signals for European forests (and uncertainties are a result *per se*). For these reasons, we have chosen to focus specifically on interannual variability within European forest ecosystems and do not extend the analysis to other ecosystems or regions.
To account for the reviewer remark, we have:

- modified the Abstract, Section 3.3.2 and the Conclusion;
- added a discussion regarding the representativeness of flux tower measurements and the need to disentangle site-specificities and regional signals to better calibrate model. It now reads:
    - l607-608: "Site-specific characteristics may also cause a disconnect between CO_2 flux variability and regional climate variability (Chu et al., 2017)."
    - l667-672: "The climate–NEE relationship is much noisier in both space and time in the observations than in the models, and it can vary substantially across different models. This indicates that local flux measurements may not reliably represent regional-scale dynamics, while models may exaggerate the influence of climate on CO_2 flux variability (despite underestimating its magnitude). Further work is needed to disentangle site-specific effects from broader-scale signals, a critical step toward improving the calibration of regional and global models that cannot resolve local heterogeneity".

Minor comments:

-Lines 71-78: These findings are questionable given the highly inconsistent patterns among products in Figure 9.

We don't fully agree with Referee #3 regarding the <u>high</u> inconsistency of the patterns. Our main point is that the influence of climate on CO_2 fluxes varies along the annual cycle. It now reads (l68-76): "At the interannual timescale, the climate does not show a significant influence on observed and modelled NEE when correlations are computed using monthly anomalies across all months combined. This apparent lack of relationship conceals meaningful seasonal patterns. In winter and fall, NEE tends to be positively correlated with temperature, soil moisture and VPD. In spring, NEE shows negative correlations with temperature and VPD, but a positive correlation with soil moisture. The summer pattern is reversed compared to the spring pattern. In the observations, these relationships are noisy in both time and space, suggesting strong site-specific effects. In contrast, the models exhibit more structured and spatially coherent patterns with strong correlations, which may reflect an exaggerated response to climate forcing despite underestimated magnitude in CO_2 flux interannual variability".

-Figure 1: The labels of 'CH-Dav' and 'IT-Ren' overlap.

Done.

-Line3 165-171: The definition of Northern, central, and Southern Europe appeared here. But it is unclear to tell from Figures 3-9 themselves. Adding this information to these figures would improve the readability.

Thank you for this remark. We added this information to all Figures.

-Line 298: Incident shortwave radiation is an essential driver of NEE and GPP. I would suggest keeping it.

The aim of this paper is not to identify the dominant climate driver of CO_2 flux variability, but rather to demonstrate that the relationship between climate and CO_2 fluxes varies substantially from month to month at the interannual timescale. The current analysis, which focuses on 2 m temperature, soil moisture, and vapour pressure deficit, is sufficient to support this conclusion. This parameter shows no clear seasonality in correlation patterns with NEE, suggesting that greater light availability generally enhances CO_2 sequestration. For this reason, we chose not to include this variable in the present study. This is now stated in Section 2.2.4.

-Figure 2: If authors still want to keep the model data, this figure is not necessarily important and could go to the supplement.

We decided to retain Figure 2 as a main figure because it highlights key challenges in comparing observations and models, while also illustrating the potential of models to overcome the limited temporal coverage of observational data.

-Figure 3: Add the abbreviation 'IV' into the caption.

Done.

-Line 371: Any significance test was done here?

No test was performed. We rephrased as follows (l354-355): "While the interannual variability of T_{2m} is the largest in winter regardless of the site, it increases markedly from south to north (Fig. 3d)".

-Lines 375-376: Why? Since it is not relevant, I would suggest removing it.

Removed as suggested.

-Lines 422-423: High correlation coefficient values could only indicate that the variance is captured well, but maybe not be true for the magnitude. Please rephrase it.

We rephrased as follows (l414-416): "The model skill in capturing the temporal phasing of the annual cycle and the magnitude in observed CO_2 fluxes is assessed in terms of correlation and mean bias, respectively (see section 2.3 for details). All models accurately capture the observed temporal phasing of GPP and RECO, with correlation values often above 0.8 (Fig. 5a)."

Note that correlation analyses are not performed for assessing magnitude errors, which is now clearly stated in section 2.3 (l311-313): "Two complementary metrics are used for model evaluation: the bias (model minus observation), which assesses errors in magnitude, and the Bravais-Pearson correlation coefficient (R), which evaluates temporal co-variability. These metrics capture distinct aspects of model performance and are not necessarily correlated".

-Lines 515-517: As mentioned above, this statement is questionable.

We removed this sentence, which was not necessary. However, we do not fully agree with Referee #3's point of view: despite the underestimated magnitude of interannual variability, the temporal co-variability can still be accurate (and vice versa).

-Lines 713-722: Same as above.

We maintained that the models are useful tools to complement observations for assessing the influence of climate on CO_2 flux interannual variability, but deleted speculative information. It now reads (l695-706): "Despite biased magnitude, the interannual variability of modelled fluxes correlates well with the observations. This supports the use of models to complement observations, whose limited temporal coverage and site specificities hinders the assessment of climate impacts on CO_2 interannual variability. We show that the influence of climate on CO_2 flux interannual variability is obscured when monthly anomalies are analyzed together. This apparent lack of relationship masks distinct seasonal patterns, which are concealed when considering all months together. Winter and fall CO_2 release increases under elevated temperature and VPD in northern and central Europe, while no clear signal emerges in southern Europe. The CO_2 sequestration increases under anomalously hot and dry conditions in spring and cold and wet conditions in summer in northern/central Europe. Anomalously cold and wet conditions also favor CO_2 sequestration in southern Europe from spring to summer. While these seasonal signals appear noisy in the observations, due to limited sample sizes and site-specific variability, they emerge more clearly in the models, albeit with some model-dependent differences".