Author response to RC2

We thank the reviewer for the constructive and clearly structured comments. We agree that the manuscript will gain considerably from the requested shortening and sharpening. In the following we reproduce the RC2 comments (in italic) and provide a discussion of our planned reaction.

Overall assessment

The Vegetation Photosynthesis and Respiration Model (VPRM) is a light-use efficiency model that estimates carbon dioxide exchange between the atmosphere and the terrestrial biosphere using satellite imagery and meteorological reanalysis data. It is frequently coupled with atmospheric transport models such as WRF. In this study, the authors investigate how VPRM results are influenced by model resolution and parameter choice in a topographically challenging environment, namely the European Alps. This is a relevant research question, as VPRM is widely applied in carbon cycle studies across Europe.

The authors conclude that, at high spatial resolution, parameter optimization has a larger impact on model performance than spatial resolution itself, whereas at 50 km resolution unresolved topographic effects dominate. They further show that differences between resolutions can largely be attributed to temperature. This result is plausible, as VPRM is primarily driven by incoming shortwave radiation and surface temperature, the latter of which is strongly influenced by topography.

We thank the reviewer for the positive overall assessment of our work. The reviewer summarizes our key finding correctly, and we appreciate that the temperature attribution is considered plausible.

Overall, the analysis is methodologically convincing. However, the manuscript is written in a way that makes it difficult to follow the main arguments and identify the key findings. I believe this issue can be addressed, but it will require substantial revision.

Readability could already be improved considerably by shortening sentences and removing redundancies. In addition, very long citation lists (e.g., Segura-Barrero et al., 2025; Jose et al., 2025; Zhao et al., 2023; Huggannavar and Indu, 2023; Raju et al., 2023; Parazoo et al., 2022; Callewaert et al., 2022; Gourdji et al., 2022) and the excessive use of acronyms should be avoided.

We acknowledge that the manuscript requires a serious overhaul in its structure and readability, in order to better make our points. We thank the reviewer for constructive suggestions for the latter, which will be adopted in a revised version. As suggested by reviewer 1 and your structural comment below, we will also completely re-write the introduction and move certain discussions to other sections (or sub-sections) in order to improve the 'flow' of the discussion and the argumentation.

Structural comments

The introduction could easily be shortened by about one-third without losing relevant information. At times it reads more like a review article than the introduction to a research paper. For example, the P-model is described in detail, despite the authors explicitly stating that it is not used in this study. This is followed by an extensive description of VPRM, including information that is not directly relevant to the research question. Similarly, the paragraph on CAMS introduces numerous acronyms and model names that many readers are unlikely to be familiar with. I encourage the authors to carefully review this section and remove material that is not essential for understanding the study or its motivation.

Thank you for this comment, which will be treated with high priority in the revised version (as it also overlaps with the comments in RC1). Many details in the introduction are not relevant to the research question, so we will cut down the introduction accordingly. The P-model will be removed from the introduction, where its connection to the rest of the study is indeed not apparent, and instead introduced briefly in the Discussion, at the point where acclimation is actually taken up. The two CAMS paragraphs will be shortened and its description generalized, to get rid of surplus acronyms. The VPRM description will be trimmed with a focus on the research questions.

> *The distinction between methodology, results, and discussion is often unclear. Equations (13)–(15), for example, belong in the Methods section and should be introduced and explained there. After reading the methodology, the reader should understand what analyses will be performed and why they are necessary.*
>
> *The Results section would also benefit from a stronger focus on the findings that are most relevant for the discussion and conclusions. At present, it is easy for the reader to lose track among the many numerical values and abbreviations. The authors should identify the key results that support their conclusions and emphasize those more clearly.*

This comment coincides with the suggestions of reviewer 1 on the same point, and we will act on both together. We will do a major restructuring to clearly separate statements into Methods, Results and Discussion. Interpretations will be moved from the Results to the Discussion and the analysis-defining Eqs (13)–(15) to the Methods, in conjunction with the related definitions. The Methods will close by stating which analyses follow and why. In the Results, each subsection will be introduced by stating its key topics and redundant numbers and acronyms will be trimmed. Finally, the conclusions will focus more on the main results, such that the main outcomes stand out.

Methodological comments

> *Temperature optimum calculation: The optimum temperature is indeed an important parameter in biogenic flux models and is often not treated with sufficient care. I therefore appreciate that the authors investigate this aspect. However, the proposed approach has important limitations. Fitting average NEE as a function of temperature does not necessarily recover the optimal temperature for GPP for several reasons. First, NEE depends on both GPP and ecosystem respiration (Reco), each of which has a strong temperature dependence. Second, this approach neglects the influence of other important environmental drivers, particularly vapor pressure deficit (VPD). Since the manuscript also investigates the sensitivity of the results to temperature, I wonder to what extent these findings depend on the estimated optimum temperature, as this parameter directly affects the shape of the temperature response function in the VPRM GPP formulation.*

We thank the reviewer for bringing up this important point, which we consider the most substantial methodological improvement of the revision and we ask ourselves why we did not use GPP to determine the optimum temperature (T_{opt}) in the first place. As respiration increases with temperature, the NEE-based T_{opt} is shifted towards lower values. From the VPRM input and WRF output we will post-process the VPRM fluxes using the GPP derived T_{opt}. To address sensitivity to T_{opt} we will add an analysis where we tune parameter sets for several percentiles (p10/p25/p50/p75/p90) of the various T_{opt} per site and per vegetation class, for the parameter sets of SITE and ALPS, respectively. Furthermore we agree that VPD deserves a more detailed discussion in the Discussion. Currently it is only implicitly described in L598 that "A more advanced biospheric model should be considered for future work, with explicit stomatal conductance responding to VPD and CO2".

DF VPRM parameters: In line 280, the authors note that the standard WRF parameterization suffers from shortcomings due to the one-step fitting procedure. However, Glauch et al. (2025), which is cited in support of this statement, also proposed an updated parameterization for Europe. Why did the authors retain the older parameter set instead of adopting the revised version?

The DF (will be renamed to DEFAULT according to RC1) parameter set was chosen because it is the set that WRF-VPRM has as default values for Europe. Quantifying its difference against the optimized ALPS set is only one of the results. We were able to show that the effects on resolution difference were consistent for both, DEFAULT and ALPS. In retrospect, the updated European parameters of Glauch et al. (2025) (their Table 3) would have been a valuable additional reference. They appeared, however, only after our own parameter optimization and the WRF simulations had been completed. At that point pyVPRM v3.0, a pre-publication version, had been used here solely to preprocess the satellite indices. The Glauch et al. (2025) parameters are moreover given for PAR $\approx$ S$\downarrow$/0.505, whereas the WRF-VPRM default is tuned for PAR = S$\downarrow$ (see our reply to RC1 on L219). Since this conversion is not exactly recoverable with our method of tuning, mixing the two conventions would introduce an additional error and leave DEFAULT, ALPS and SITE no longer directly comparable.

We will extend the future-work discussion and recommend a comparison against the updated European parameters of Glauch et al. (2025) (their Table 3) and Gerbig (2024) (which Glauch et al. (2025) use as a reference), in particular against a parameter set tuned with the current pyVPRM version for the Alpine domain using the Sentinel-2 and VIIRS satellite data. The methodological differences between the optimization of Glauch et al. (2025) and that of Ranit et al. (2025), together with the PAR $\approx$ S$\downarrow$/0.505 conversion, deserve a closer investigation. The T_{opt} sensitivity analysis will be added in this revision (a range of T_{opt} per site and per vegetation class) provides a further method to be tested and evaluated with the Sentinel-2 and VIIRS satellite data supported by pyVPRM.

SITE parameters: I am not convinced that this parameter set adds much to the paper. As I understand it, the parameters are calibrated and evaluated using the same site, meaning the results primarily demonstrate goodness of fit rather than predictive skill. If this interpretation is correct, I would place less emphasis on this comparison.

We agree that SITE should not be over-interpreted and will de-emphasize it in the Results, presenting it explicitly as a best-case benchmark; the reading as a pure goodness-of-fit test, however, needs one clarification. The SITE parameters are calibrated offline, whereas the validation reported here evaluates the online WRF-VPRM, i.e. VPRM driven by WRF-simulated meteorology at the nearest representative grid-point. The parameters are therefore not evaluated on the data to which they were fitted, the meteorological forcing and the spatial representation both differ between calibration and validation, so the SITE evaluation is not a pure goodness-of-fit test. With the biology optimally tuned, the residual SITE-to-observation gap mainly reflects the WRF meteorological and representativeness error rather than parameter overfitting. This is precisely the role we intend for SITE: an upper bound on achievable skill that isolates the forcing contribution, against which the domain-wide ALPS and DEFAULT sets are measured. We will make this calibration/validation distinction explicit in the revised version and, correspondingly, put less emphasis on SITE's overall ranking.

Sample size: I assume that the decision to analyze only 12 sunny and rainy days was motivated by computational constraints. Under those circumstances, this approach may be justified. However, I wonder how much uncertainty is introduced by this selection. Have the authors evaluated the sensitivity of their conclusions to the number of sampled days, for example by repeating the analysis with only 8 or 10 days?

This is indeed an important point that the reviewer is raising here. And yes, the reason for simulating only 24 days are limited computational resources, a full year at 1 km grid spacing is computationally infeasible. See also our response to the third part of the general comment of reviewer 1. First of all and based on both reviewers' comments, we will use the 24 days (a clear sky and a cloudy/rainy day per month) no longer as an estimate for the "approximate yearly NEE". Rather, the average NEE over the 24 sample days will be used as a rough estimate in the revised manuscript. As robustness estimate we will add a new table on a four-season (DJF/MAM/JJA/SON) breakdown of the resolution and parameter effects to an appendix. Also, we plan to replace Figs. E6-E15 with averages on three clear-sky and three cloudy days per season. This will demonstrate that the variability among the days per season is smaller than the differences between the seasons and hence support the robustness of the sign and magnitude of the effects across smaller subsamples. A growing season 9-day subsample (MAM/JJA/SON) can also be computed from the existing output.

> *CAMS comparison: I am not convinced that the comparison with CAMS substantially strengthens the manuscript. It is not mentioned in the abstract, and in the discussion the only conclusion drawn is that: "Comparison with CAMS confirms that WRF-VPRM results fall within a realistic range for the annual carbon budget, despite differences in seasonal cycle and meteorological drivers." This finding is not particularly surprising, given that the model already reproduces the station observations with reasonable accuracy. In my opinion, this comparison could be omitted, as it increases the length and complexity of the manuscript without adding substantial insight.*

We take this point seriously and will substantially reduce the role of CAMS. It will be retained only as an independent, domain-averaged external benchmark and it will be removed from the site-scale validation, where its coarse grid cannot resolve the tower footprint.

We already made a suggestion to reviewer 1: We will point out more clearly what we consider the one relevant insight that can be obtained from the comparison to CAMS. As visible in Fig. E8 and E13, CAMS does not reproduce a realistic seasonal cycle of GPP over the Alpine domain. The close agreement in the annual mean between CAMS and WRF-VPRM is coincidental, it results from offsetting seasonal biases (CAMS GPP is higher than in all WRF-VPRM runs in winter and lower in summer) rather than from an accurate representation of the fluxes. In the revised manuscript we will discuss this shortcoming more critically in the Discussion. CAMS is a widely used, assimilation-constrained product in carbon-cycle science, and documenting that it nonetheless misrepresents the Alpine seasonal cycle is, in our view, a valuable result - thus avoiding the tendency for such 'negative results' to go unreported.

Closing

> *Overall, I believe the study addresses an important and relevant research question, and the analysis is generally sound. If the authors adequately address the comments above I'm happy to see the (revised) manuscript being published.*

References

Gerbig, C.: Parameters for the Vegetation Photosynthesis and Respiration Model VPRM, ICOS-ERIC - Carbon Portal [data set], doi:10.18160/HF4C-G0KK, 2024.

Glauch, T., Marshall, J., Gerbig, C., Botía, S., Gałkowski, M., Vardag, S. N., and Butz, A.: pyVPRM: A next-generation Vegetation Photosynthesis and Respiration Model for the post-MODIS era, Geoscientific Model Development, 18, 4713–4742, 2025.

Ranit, Bao, S., Koirala, S., Brenning, A., Reichstein, M., Tagesson, T., Liddell, M., Ibrom, A., Wolf, S., Šigut, L., et al.: Addressing challenges in simulating inter–annual variability of gross primary production, Journal of Advances in Modeling Earth Systems, 17, e2024MS004 697, 2025.