Author response to RC1

We thank the reviewer for their thoughtful and detailed comments and considerations. We indeed think that addressing them all will make the paper much better readable and more to the point. In the following we reproduce the RC1 comments (in italic) and provide a discussion of our planned reaction.

General comment

This study attempts to answer the question of how sensitive the results of VPRM (Vegetation Photosynthesis Respiration Model) simulations of carbon dioxide fluxes are to changes in model resolution and the selection of VPRM parameter settings. Building upon a previous synthetic study, the authors argue that uncertainties in CO_2 fluxes over mountainous regions, stemming from variable resolution, could impact future climate predictions. In the current study, they aim to expand the analysis to a real-world case. To obtain the accurate CO_2 fluxes required for their investigation, the authors evaluate and optimise the VPRM model parameters using observational data from FLUXNET and previously described methodologies to arrive at improved gross ecosystem exchange (GEE) and respiration (RECO) fluxes.

Throughout the text, the authors demonstrate a good grasp of the topic and a fluency in the technical aspects of the numerical model and data analysis. The design of the experiment is sound overall, and the methods and tools employed are appropriate. The language is of high quality, figures are well prepared (some might use larger fonts).

We thank the reviewer for the thorough assessment, in particular for recognizing that the experimental design is sound and that the methods and tools are appropriate. We will enlarge the fonts in some figures as suggested.

Nevertheless, the paper suffers from some major deficiencies. The text's structure is chaotic, with a long and rambling introduction and an undisciplined approach to separating methods, results, and discussion. This, together with repeated information and five appendices, makes it difficult to follow. Please see below for some suggestions about restructuring, but note that these are merely indicative, as some sections require a complete overhaul (e.g. the introduction).

We acknowledge that the text structure and wording needs an overhaul. We will shorten the introduction as described in more detail below under the specific major comments. Furthermore the Methods, Results and Discussion will be separated more strictly following the suggestions by RC1. Aspects of interpretation will be moved from the Results to the Discussion while the equations explaining the details of the analysis will be moved to Methods. The detailed VPRM formulation will be moved to an appendix. Repeated passages will be removed, long sentences shortened, and acronyms pruned to those actually reused.

Most importantly, I believe that the authors are attempting to draw significant conclusions from a dataset that is too limited. With only 24-days of simulations selected to represent cloudy and clear conditions in equal measure, the authors are left with a very limited dataset on which to base conclusions about a highly dynamic process such as ecosystem CO_2 exchange. This makes the robustness of the presented results doubtful at best, a fact of which the authors are somewhat aware (L395–406). In the text, the authors suggest a multi-year analysis (L400–404); however, this might not be necessary. If a climatologically regular year is chosen, a single 365-day period should represent seasonal changes adequately. If this is computationally too expensive, perhaps a comparison between winter and summer would be sufficient, although with shorter periods it might be harder to achieve a balance between cloudy and clear conditions.

In its current form, I suggest rejecting the paper and requesting another complete review round after all the listed points have been addressed.

We thank the reviewer for their open and clear comment. We certainly agree that wording like "approximate yearly NEE" is misleading and not justified by using only 24-days of a year. We will replace the wording "approximate yearly NEE" by a more appropriate description, e.g. the average NEE over the 24 sample days. We clearly state the results have several limitations and might not be representative (L395–406). We will make this more transparent in the text.

The main reason for choosing this 24-day sample is indeed limited computational resources. To calculate these 24 days at 1 km resolution, using 256 cores, took 8 days of pure computing time on the LEO5 HPC cluster of the University of Innsbruck, with additional waiting time in the queue it took around 2 months to run the simulations. To run a full year, this would then translate to 120 days of pure computation time and possibly a year or more to run the simulations. First test simulations revealed that the 1-km resolution was necessary since topography and, in particular, convection is not well represented in WRF on the 3 km grid. Also, they revealed that the largest difference in NEE occurred between clear-sky and cloudy/rainy days. From these first results our approach emerged to approximate the seasonal cycle of NEE by simulating a clear sky and a cloudy/rainy day for each month. In the revised version we will i) extend the discussion on the impact of weather conditions on NEE in the main text, i.e. better motivate the chosen simulation strategy, ii) add a new table to appendix D in which resolution and parameter effects are presented on a seasonal (DJF/MAM/JJA/SON) basis, iii) replace Figs. E6-E15 (appendix E) with averages of three clear-sky and three cloudy days per season and iv) avoid the wording 'approximate yearly average' (as already mentioned before). These changes will be accompanied by re-formulations in the respective parts in the text. This will strengthen the reasoning (i.e., why do we choose clear sky vs. cloudy/ rainy days?), the interpretation of the results (rough first guess for a seasonal cycle) and will not trigger false expectations like those for an "approximate yearly NEE".

Specific major comments

> *The introduction shows that the authors have dedicated time to studying a broad range of topics, but the themes described change too radically. The paragraphs describe, in order, the terrestrial carbon sink, observation data, night-time biases in mountainous terrain, CAMS CO_2 fluxes, the influence of complex topography on atmospheric dynamics and various ecosystem models (three paragraphs), and the influence of topography on CO_2 fluxes. Afterwards, the description of the experiment is slightly more structured. Please shorten and rearrange the introduction, focusing only on information relevant to the study.*

Thank you for this comment, which will be treated with high priority in the revised version (as it also overlaps with RC2). Many details in the introduction are not relevant to the research question. We will cut down and reorder the introduction substantially. 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 temperature 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.

> *There are also some important literature references which the authors should not omit, given the topic. Ahmadov et al. (2007) describe the first implementation of the WRF model with the VPRM, which later became WRF-GHG (see the remark below regarding L91). Furthermore, Pillai et al. (2011) have studied CO_2 simulations with WRF-VPRM over the complex terrain around the tall tower on Ochsenkopf mountain. While not alpine, the authors might also note that the analysis of model performance at varying resolutions was discussed there too, albeit the fluxes themselves were not the primary focus.*

We discuss our reaction to this major comment with our reply to L91 below. Additionally, we thank the reviewer for bringing up the Pillai et al. (2011) reference. We will add to the introduction that Pillai et al. (2011) simulated atmospheric CO_2 with WRF-VPRM over the Ochsenkopf tall tower in complex terrain, and that their resolution-sensitivity test (2 km vs. 6, 12 and 24 km) showed the simulated CO_2 to degrade markedly at 24 km.

> *I am also uncertain of the usefuleness of CAMS products in the study. While the authors reasonably state that they are using it for comparison purposes, they also refer to a study by Custodio, Borrego and Relvas (2022), which focuses on the atmospheric CO_2 concentrations of the CAMS EGG4 global reanalysis product. However, in other parts of the text, global forecasts are mentioned (L53 and L62 refer to 9–10 km resolution, whereas EGG4 is much lower at 0.75° x 0.75°), which are a different product, for which the results of the Custodio et al. study might not be relevant. The authors do not seem to be aware of the differences between them and simply refer to the data as CAMS. The data product used must be clearly identified. Also, while the results may be of some interest, considering the minor role of CAMS in the discussion, the authors should seriously consider moving most of these results to the supplement.*

We thank the reviewer for this thoughtful consideration. In the revised version, we will identify the product unambiguously as the CAMS EGG4 global reanalysis ($\sim$0.75°), introduced once as "CAMS (EGG4)" and thereafter abbreviated "CAMS". We will remove the misleading text on the 9–10 km operational IFS product in the introduction and update CAMS to "CAMS (EGG4)" in the code and data-availability section.

Reviewer 2 even suggests to omit CAMS completely from the manuscript. We argue to retain it 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 will, however, 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.

Other comments

L09

> *DEF would be a more natural acronym since length is not an issue to other sets (ALPS, SITE)*

Accepted. We agree that the acronym should be more natural and will rename the default set throughout the manuscript. We prefer the unambiguous "DEFAULT" over "DEF", which could also be read as "definition"; as the reviewer notes, length is not an issue here.

L22

> *"resolution-induced errors govern NEE uncertainty" – this suggests like resolution is the main driver of the overall uncertainty. This is too general, there are other significant sources of error for annual NEE. Please reword that this relates only to results of this numerical experiment.*

Accepted. We acknowledge that making such a claim about annual NEE uncertainty in the abstract is misleading. As mentioned above we will not estimate on the "approximate yearly NEE" in general any more but describe it as the 24-day average it actually is.

L25

> *which temperature is meant here? Please be precise.*

Accepted. We will specify the 2 m air temperature ($T_{2\mathrm{m}}$) in L25.

L68–69, and in other places

> *Throughout the text authors conflate "plant functional types" (PFT) with "vegetation classes", but the distinction is not made clear. In all of the past studies describing VPRM and WRF-VPRM, "vegetation class" is used consistently. PFT is meant to distinguish groups of vegetation meant to behave same way in terms of process, whereas vegetation classes typically describe dominant ecosystem type over a certain area – related to what eddy covariance tehcnique is able to measure. I suggest authors to refrain from using PFT and use the more appropriate term "vegetation class" – consistent with previous literature.*

We will use "vegetation class" consistently in the revised manuscript, in line with the WRF-VPRM literature.

L83–84

> *In the present study... – this fits better in "Discussion"*

Accepted. This passage will be moved to the Discussion.

L91

> *"...CO2, CH4, and CO as passive tracers". Historically, WRF-VPRM was developed only for CO2 (Ahmadov et el. 2007). Augmented version with three tracers is more appropriately referred to as WRF-GHG. See section 2.2 in Beck et al., 2013. In the context of this current study, perhaps WRF-GHG would also be more appropriate, but WRF-VPRM is acceptable.*

Accepted. We will cite Ahmadov et al. (2007) as the first WRF–VPRM coupling and add a more complete description on the history on the development of WRF–VPRM to WRF–GHG, as documented in the publications of Beck et al. (2011) and Beck et al. (2013). As the acronym WRF–VPRM prevailed in most publications and we do not use the features for methane and carbon monoxide, which were the major improvements of WRF–GHG, we prefer to keep using WRF–VPRM as acronym.

L111–112

> *50% of the global land surface seems excessive in the context of CO2 fluxes. Does it mean that over 50% of the land area exhibit significant biases of CO2 fluxes in the climate models? Consider dropping the reference.*

Partially accepted. The sentence is misleading and will be reformulated. In Rotach et al. (2014) the ">50% of the global land surface" figure refers to the areal fraction of land classified as complex terrain (defined there by a standard deviation in terrain height exceeding 50 m on a 20 km × 20 km grid), not to the fraction of the land surface where modelled CO_2 fluxes are significantly biased.

We would therefore keep the reference, because it is the source of the hypothesis that motivates this study. Additionally we will cite Kapos et al. (2000) who define 24% of the landmass to be mountains. Rotach et al. (2014) explicitly state that numerical models potentially can have errors due to complex terrain,¸ but the magnitude of this effect is "largely unknown"; we will say in the revised manuscript that the resulting flux error has not been quantified, but that a systematic, one-sided bias could accumulate in long-term carbon-budget and climate simulations. Quantifying this error for real Alpine terrain is precisely the aim of the present study. The revised sentence will state that, according to the definition of Rotach et al. (2014), more than half of the global land surface is complex terrain for numerical models, and that neglecting subgrid-scale contributions to turbulent fluxes over such terrain may lead to a systematic underestimation of the modelled net CO_2 exchange.

L121

> *Before this paragraphs the text feels chaotic and disjointed. Intro needs overhaul. See major comments.*

Accepted. Please refer to our reply on the overhaul of the introduction above.

L145

> *Single days (the classic case study) – Too general. I am not aware of single days being "classic case study". Please rephrase.*

Accepted. We will remove the wording of "the classic case study" and the "extended case study" which relates to it. We will rephrase L144-149 accordingly.

L147

> *How do the authors assure that the days selected are unbiased and representative? Conditions on selected days could have been easily compared against climatology. One could argue that 12 days is a sufficient sample size to randomize the selection and avoid bias, but the issue is that CO2 fluxes are seasonally varying with large variability, which effectively reduces the sampling to 3 per season. The results could be unbiased, but quite robustness of the results will remain in question with such a limited sample.*

Partially accepted. Our main focus is to understand the processes and separate the underlying contributions. As mentioned above the strongest deviations occurred in distinct weather conditions, clear sky and cloudy/rainy. We will remove the claim to have chosen a representative sample also in L147. We will add a seasonal analysis of the clear sky and cloudy/rainy weather conditions to show the robustness of the results.

L154

> *Factor 6x between d01 and d02 is not a typical choice for WRF. The model authors recommend "using odd numbers, and typically only advise to use either a 5:1 or 3:1 ratio".. Did the authors evaluate the consequences of using the even, hight ratio between domains? What was the spacing between domain boundaries? Distance between domains is recommended to diminish the boundary distortions. Authors only write that smaller domains were "same size". Please be precise.*

Accepted. We will be more precise. There is no 6:1 nest: the 54 km and 9 km fields are independent single-domain runs, each driven directly by ERA5 boundary conditions. The only nest is a single 3:1 step from 3 km to 1 km. The intermediate 3 km domain is driven by ERA5 and its output is discarded. As described in L176-178 the analysis is done only in the area of the 1 km nest without the 10 outermost gridpoints.

L158

1. Was pressure data used, or model-level data? ERA5 comes in both, selection determines the vertical resolution of the input data. 2. What was the thickness of the lowest layer? 3. How many levels were below 3km altitude on average?

Accepted. We will specify first that ERA5 is used on model levels (`num_metgrid_levels` = 33), interpolated onto 81 WRF eta-levels. Second, the lowest model layer is approximately 20 m thick (lowest scalar level about 10 m above ground) and third, roughly 40 of the 81 eta-levels lie below 3 km.

L161

"Each run started..." – What was the rationale behind this? Please note that this is the same setting as Ahmadov et al. proposed in 2007 (see par. 25, page 5 in their study), I suggest adding the reference here.

Accepted. We will state the rationale for the run setup, that each run starts at 18:00 UTC, including a 6 h spin-up, so that the boundary layer and transport are established before the analyzed day. We will add the reference to Ahmadov et al. (2007), who proposed the same strategy of discarding an initial spin-up period before the analysed day.

L167

"are mapped to plant functional types (PFT) of VPRM" – 1. See comment to L68-69 above. 2. This acronym is already explained earlier in L68. In other places, however, the authors skip explaining acronyms and refer to the Table D8 (e.g. NSSE). Please make it consistent. Also, I suggest to move the acronym table as the very first appendix.

Accepted. We will make the terminology consistent, remove redundant acronym expansions, and move the abbreviations table to the first appendix.

Figure 2, caption

"Areas with a standard deviation of subgrid-scale topography of STDTOPO < 200 m" – Consider calling that "area of interest" or "area of analysis" for easier reference in text.

Accepted. We will introduce the areas with a standard deviation of subgrid-scale topography of $STD_{TOPO} < 200$ m as the "area of analysis" and use that term in the text.

L205

This detailed description shows the authors carefully applied the formulas for VPRM, which is commendable. Scientifically, however it doesn't bring new knowledge. For paper brevity, I would recommend moving most of this section to the supplement, leaving only a short description of the VPRM model and underline significant differences to previous applications – this is done in the next subsection, so both can then be merged.

Accepted. We will move the detailed VPRM equations to an appendix (also Sec. 2.4.1 and 2.4.2), leaving a concise description and the differences from previous applications in the main Methods, accordingly the novel T_{opt} derivation (Section 2.4.3) will be kept in the Methods.

L219

Per my understanding this paragraph is about a minor unit change with linear scaling, and results should be 1:1 identical. If I'm not mistaken – again, for brevity, consider moving to supplement, with reference here (unless the whole section is moved, see above).

Partially accepted. We will condense the unit-conversion passage and move the equations to the appendix. We would, however, like to retain the essential information in the main text, because the change is not a purely linear rescaling that leaves the results unchanged.

The conversion only cancels if $S\downarrow$ together with $PAR_{0,v}$ and λ are converted consistently. In Eq. (3) $PAR_{0,v}$ is the half-saturation value, so switching between $PAR = S\downarrow/0.505$ and $PAR = S\downarrow$ without converting $PAR_{0,v}$ changes the saturation term by approximately a factor of two. This matters here because the default parameters in WRF-VPRM are not neutral with respect to this choice. They were tuned under the $PAR = S\downarrow$ assumption.

During model development, already about 2 years ago, we initially applied the approximation $PAR = S\downarrow/0.505$ in the parameter tuning (direct PAR measurements were not available for all site-years) and only later found that $S\downarrow$ (in WRF "swdown") is passed to the VPRM subroutine not PAR (in the VPRM subroutine in WRF/chem/module_ghg_fluxes.F it is renamed to "RAD" but is essentially equal to swdown), so that the standard WRF implementation effectively assumes $PAR = S\downarrow$ (so, we had wrong values of $PAR_{0,v}$ as we did not convert them back by 0.505). Two consistent options therefore exist: either tune $PAR_{0,v}$ against PAR from FLUXNET and convert the result to $PAR_{0,v}(S\downarrow) = 0.505\, PAR_{0,v}$ (with the corresponding $\lambda_v/0.505$) before entering it in the WRF-VPRM parameter table, or tune $PAR_{0,v}$ and λ directly using $PAR = S\downarrow$.

Studying the WRF-VPRM literature on this point and personal communication with several WRF-VPRM developers, we confirmed that the current WRF default parameters are tuned using $PAR = S\downarrow$. We then tested the conversion by repeating the optimization with $PAR = S\downarrow$ (so using $S\downarrow$ from FLUXNET data) and comparing the resulting parameters to those obtained by $S\downarrow/0.505$. The recovered factors deviated from the theoretical value of 0.505: e.g. for grassland at Monte Bondone (IT-Mbo) in 2012 we obtained $PAR_0 \times 0.513$ and $\lambda/0.511$, but averaged over all 19 Alpine sites and approximately 180 site-years the factors were close to 0.6. We attribute this to the optimization procedure, in which R_{eco} and GPP are fitted separately but constrained by the NEE quality control, and to our use of median values of T_{opt} across site-years per vegetation class. The theoretical conversion is therefore not exactly recoverable in our specific tuning method, which is why we optimized under the same convention as the original WRF implementation of Beck et al. (2011), so that SITE and ALPS remain directly comparable to DEFAULT.

We would like to keep a condensed statement of the two conventions and their consequence for the parameters in the main text, and move the supporting detail to the Appendix, where the optimization procedure is already described.

L278

> *"There is no dedicated publication on the VPRM parameters that are currently implemented in WRF-VPRM." – not entirely precise; to the best of my knowledge, the parameters for US are still the original ones from Mahadevan et al., so for default values the study of Mahadevan et al. can be used as ref; for Europe these were implemented in Beck et al. 2011, however the authors are correct that there is no dedicated description on how the European callibration was made.*

Accepted. We will be more precise saying that there exists no dedicated description of the European parameter calibration.

L290

> *"... are minimized with a cost function": Please provide the formula*

Accepted. We will provide the explicit cost-function formula in the appendix:

$$\text{cost} = 0.5\,\frac{\Delta h}{\Delta h_{\max}} + 0.1\,\frac{\Delta\text{slope}}{90^\circ} + 0.1\,\frac{\Delta\text{aspect}}{180^\circ} + 0.6\,(1-\text{veg}_{\text{frac}})^2, \tag{1}$$

where Δh, Δslope and Δaspect are the absolute differences in terrain height, slope and aspect between the candidate grid-point and the site, and $\Delta h_{\max}$ is the maximum of Δh over the terrain field within a radius, used to normalize the elevation term. veg_{frac} is the fractional cover of the site's vegetation class at the candidate grid-point. All terms are normalized to the range $[0, 1]$, and the grid-point with the minimum cost within a 30 km radius is selected.

L300–301

"To partly account for this, a PFT-specific Topt was determined for the Alps. T2m and NEE data from nineteen FLUXNET sites. . . " – note that Fig 2 shows 5 sites. Table D2 shows it's 15 sites in the Alps. Please clarify and give reference to D2

Accepted. We will clarify that the optimization uses nineteen FLUXNET sites (fifteen in the Alps plus four additional European sites), that Fig. 2 marks only the sites falling within the "area of analysis", and we will reference the corresponding table.

L323

Method of interpolation?

Accepted. The method of interpolation is the nearest-neighbour for the vegetation-class fields, it will be named likewise where it is used.

L333, all sections

This, while interesting, is less important for the scientific part. Consider moving to the supplement for brevity.

Accepted. We will move "Changes to WRF-VPRM code" to the appendix.

L367–371, Figure 3

The authors use it as an example plot, but it depends on the T2m simulated in the model, so rather fits into a "Result" section. My suggestion would be to move it there.

Accepted. We will move Fig. 3 to the Results section.

L375

"IT-MBo was selected because it exhibited the best overall model-data agreement" – I do not think this is methodologically appropriate. Rephrase. Also: consider bringing all results (E2-E5) into the main text, and Tables D5 and D7. The results form the basis of evaluation and are critically important considering the study title.

Accepted. We agree that the description "best overall model-data agreement" is methodologically inappropriate. We will rephrase this focusing the description on other criteria that led to the decision to use IT-MBo: it suits well to the research question, its representative grid-point contains a 91% fraction of grassland and differs by only 90 m in elevation from the tower (see Table D5), and the site lies in complex terrain at 1550 m AMSL (see Table D2).

Regarding the placement of the validation results, we agree they are central to the study. This connects to RC2 on the optimum temperature: as part of the revision we will derive T_{opt} from GPP rather than NEE and add a sensitivity analysis across its range per vegetation class. We therefore plan to promote Fig. E1, which documents the parameter optimization, into the main text in an extended form: panel (a) will show the fitted T_{opt} across its range over the years rather than a single value, panel (b) will show the per-vegetation-class distribution as box plots (similar to Fig. E17a, it could also be placed in E17a per vegetation class and E17b per site), and panel (c) will show the actual spread of T_{opt} instead of only the class-mean values. We would prefer, however, to keep the evaluation tables E2-E5 in the appendix, as they would further extend the main text with, from our point of view, little additional insights. Moreover IT-MBo is currently used as the running example for the parameter optimization in Fig. E1a and E16, but not for the other 4 validation sites. To extend the focus in validation we may add likewise figures to the appendix as well for the remaining sites (AT-Neu, CH-Dav, IT-Lav and IT-Ren).

L388

> *"Detailed MB and MAE values are listed in Appendix D (Table D5)." – repeated info, already given in previous paragraph*

Accepted. We will delete the duplicated MB/MAE sentence.

L395–399

> *This goes directly towards my major comment about robustness. I'm sorry but in my mind this undermines the whole analysis.*

Accepted. We refer to the 24-day response in detail above. The annual claims will be softened and robustness evidence added.

L400–406

> *Also see my major comment. Unrelated: this fits more into the Discussion section.*

Accepted. The multi-year outlook will be moved to the Discussion.

L416–419

> *This fits into Methods section.*

Accepted. L416-419 will be moved to the Methods.

L420

> *"warmer T2m" – should be "higher T2m".*

Accepted. We will change "warmer T_{2m}" to "higher T_{2m}".

L432

> *"resulting from CAMS being higher than" – I assume it should be "CAMS CO2 fluxes being higher. . . "*

Accepted. We will rephrase to ". . . resulting from CAMS CO_2 fluxes being higher than in all WRF runs in winter and lower in summer."

L433

> *"Figure 6 shows the spatial distribution of S↓ and RAD at 54 km and 1 km grid spacing for a morning hour when differences are most pronounced." – again, this might seem like cherry-picking.*

Accepted. We will justify the choice of the morning hour more clearly, a systematic effect is shown here: The resolution-induced radiation difference is largest shortly after sunrise, when the low solar elevation maximizes the influence of slope orientation and terrain shading.

L440

> *"Under clear-sky conditions at 1 km, DF reaches 64% and 54% of ALPS GPP and Reco, respectively, ..." – usually the older value, even if imperfect, is treated as reference. Consider describing ALPS as percentage of DF.*

Accepted. We will swap the numerator and denominator for the percentage values to get the comparison with the default (DEFAULT) set as the reference.

L441

> *"This difference arises from..." – move to Discussion please.*

Accepted. This will be moved to the Discussion.

L444

> *"This apparent improvement..." – same.*

Accepted. This will be moved to the Discussion.

L472

> *"While CAMS employs a distinct biosphere model..." – move to Methods (or Discussion)*

Accepted. This will be shortened and moved to the Discussion.

L477

> *Overall – same.*

Accepted. This will be moved to the Discussion.

L482–489

> *This is Intro.*

Accepted. This will be moved to the Introduction.

L603

> *"A feasible approach to reduce structural resolution effects is a subgrid-scale flux..." – The authors write with a high degree of certainty, however it is speculative whether such a parameterization would in fact fix the lack of resolution. Similar issues faced by climate models related to convection and other scale-sensitive phenomena have shown that deficiencies are to be expected. So until demonstrated, I gently suggest to soften the language somewhat, until method is demonstrated.*

Accepted. We will soften the language: Such a parameterization could reduce structural resolution effects, but whether it can in practice recover the resolved-scale fluxes remains to be demonstrated, as residual deficiencies are to be expected.

L612

"bias CO2 update, with a direction" – change to "sign"

Accepted. We will change "direction" to "sign".

L617

"(Q2) At fine horizontal grid spacing (O(10 km)), parameter choice is the" – consider expanding what parameter is meant.

Accepted. We will name the parameters explicitly (λ, PAR_0, α, β and T_{opt}).

L622

"shows that site-specific parameters" – "shows that site-specific (i.e. adjusted based individually for each site) parameters"

Accepted. We will expand the phrase to "site-specific (i.e. adjusted individually for each site) parameters".

References

Ahmadov, R., Gerbig, C., Kretschmer, R., Koerner, S., Neininger, B., Dolmann, A. J., and Sarat, C.: Mesoscale Covariance of Transport and CO_2 Fluxes: Evidence from Observations and Simulations using the WRF-VPRM coupled Atmosphere-Biosphere Model, J. Geophys. Res., 112, D22 107, 2007.

Beck, V., Koch, T., Kretschmer, R., Marshall, J., Ahmadov, R., Gerbig, C., Pillai, D., and Heimann, M.: The WRF Greenhouse Gas Model (WRF-GHG). Technical Report No. 25, Tech. rep., Max Planck Institute for Biogeochemistry, Jena, Germany, 2011.

Beck, V., Gerbig, C., Koch, T., Bela, M. M., Longo, K. M., Freitas, S. R., Kaplan, J. O., Prigent, C., Bergamaschi, P., and Heimann, M.: WRF-Chem simulations in the Amazon region during wet and dry season transitions: evaluation of methane models and wetland inundation maps, Atmos. Chem. Phys., 13, 7961–7982, doi:10.5194/acp-13-7961-2013, 2013.

Kapos, V., Rhind, J., Edwards, M., Price, M., Ravilious, C., and Butt, N.: Developing a map of the world's mountain forests., Forests in sustainable mountain development: a state of knowledge report for 2000, Task Force For. Sustain. Mt. Dev., pp. 4–19, 2000.

Pillai, D., Gerbig, C., Ahmadov, R., Rödenbeck, C., Kretschmer, R., Koch, T., Thompson, R., Neininger, B., and Lavrié, J.: High-Resolution Simulations of Atmospheric CO_2 over Complex Terrain – Representing the Ochsenkopf Mountain Tall Tower, Atmospheric Chemistry and Physics, 11, 7445–7464, 2011.

Rotach, M., Wohlfahrt, G., Hansel, A., Reif, M., Wagner, J., and Gohm, A.: The World is Not Flat: Implications for the Global Carbon Balance, Bulletin of the American Meteorological Society, 95, 1021–1028, 2014.