Comment to revised manuscript

I am grateful to the authors for the great effort in reviewing and editing the manuscript. Its scientific correctness and significancy improved and I appreciate the effort conducted after the first round of reviews. I have several minor and technical comments that could enhance further the message and presentation of the manuscript; I would recommend after these, the manuscript could be accepted for publication. Comments refer to line numbers in the tracked-changes document the authors uploaded after the 1st round of reviews.

We are really grateful to Referee 1 for all her/his work done for the two rounds of review that significantly help us improving our work. We deeply took into account all her/his comments below as well those from Referee 2 doing further computations as requested and modifying in consequence the manuscript, we really hope it will now suit all referees and co-editor recommendations and be finally accepted for publication.

Minor and technical comments

1.- The main concern I have relates to clarity regarding the presentation of the story. Although no major changes are needed, regarding organization of sections or figures (except for some small comments that are detailed in what follows), I would recommend revising again the text. Try to reduce sentences and write them in a more reader-friendly style. Particularly, this would be beneficial for the discussion section.

We revised as much as possible our text again (in the discussion section as well) accordingly to the significant modifications in new computations, analysis, figures and text we did for this second round of review from Referees 1 and 2.

2.- Title: although I suggested the change in the title, it could be further refined by "Building a random forest machine learning model for annual net ecosystem exchange estimation in agricultural fields using discontinuous atmospheric Eddy Covariance measurements".

Done. The title now reads exactly as suggested: “Building a random forest machine learning model for annual Net Ecosystem Exchange estimation in agricultural fields using discontinuous atmospheric Eddy Covariance measurements.”

3.- Line 18: to estimate the annual net ecosystem exchange (NEE)

Done, implemented as suggested in the Abstract.

4.- Line 27: estimated from the sector-specific models indicated

Done

5.- Line 54: A suggestion: Change to "However, their overall role as a carbon sink or source depends largely also on agricultural management practices".

Done

6.- Line 69: Change "net carbon budget" to "net ecosystem carbon exchange" as suggested previously, as I think you don't specifically refer here to carbon budget.

Done

7.- Lines 73 and 179: Although I might contradict what other referee suggested, the word "atmospheric" is not necessary here.

Done

8.- Line 83: Baldocchi 2020 did not explicitly state that EC networks span the main climate and ecological regions; it said that the addition of new towers did not alter significantly the statistics of fluxes despite including new

regions. This does not mean that there is no bias in the global networks and global responses of fluxes, as there are many regions (for instance subtropical, dry regions, or mountainous forests) that are not well represented in continental or global networks and it could happen that many new towers are covering relatively similar biomes compared to the existing ones.

Done. The sentence now reads: "Baldocchi (2020) notes that adding new towers to these EC networks has not significantly changed the aggregate flux statistics despite extending coverage into new regions; this does not imply, however, that global EC networks are free of representativeness bias, as several biome types (e.g., subtropical and dry regions, mountainous forests, wetlands) remain under represented, and new towers may disproportionately cover already well-represented biomes."

9.- Lines 181-183: frequency, which occur in the surface boundary layer. The EC method

Done

10.- Line 207: You could remove "we nove use since several years" as that is not itself a justification; the justification is adequately provided later on.

Done

11.- Line 209: Maybe you could write "To justify the 10-min averaging period instead of the usual 30, comparisons were done between ..."

Done

12.- Please check throughout all text that 10 min is written without a "." after min. min is already the unit and does not need a dot denoting abbreviation.

Done in the main text

13.- Line 217: included

Done

14.- In line with a comment from previous round, Lines 239 and 242: detrending is based on the linear method; please avoid double writing the same. Also, generally for this section, as the averaging period was written above, it is not necessary to state several times that the 10-min was used (e.g. in line 244). Instead, you could write "the averaging period".

Done

15.- Line 244: contained

Done

16.- Please refer to my minor comment number 6 in the previous round; the section describing the filters that were applied to processed data still remains unclear, so I would encourage to re-organize it for clarity.

17.- Line 270, figure 3: Please increase labels and legend size of the figure

Done

18.- Line 282: remove ',' after identify.

Done

19.- Line 395: You could write "with the number of 10-min observations" and remove "recorded each 10 minutes".

Done

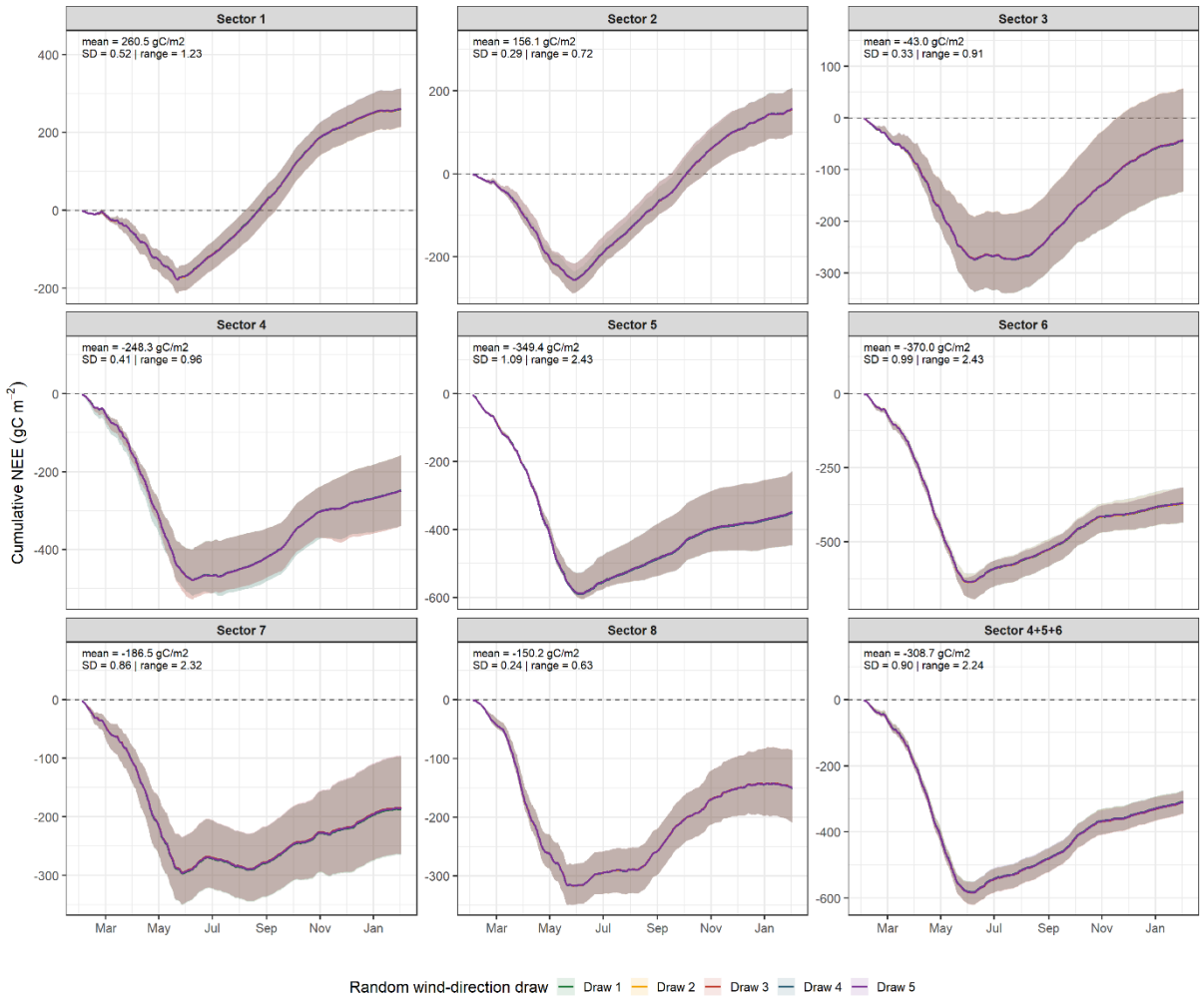
20.- Line 420: in this paragraph, state first what was done for the RF model, then say that the VSURF was not applied. It is generally better to start from the methods or strategies that were applied, instead of what was not.

Done

21.- Line 443: For this paragraph - what is the importance of the wind direction in predictions? If it is high, then choosing a random wind direction within the sector's angular range could be misleading.

We thank Referee 1 for this question, which was also raised by Referee 2. To answer it, we carried out a dedicated sensitivity test of the reconstructed NEE values to the random wind-direction assignment.

For each sector, the complete annual reconstruction was repeated five times, each time with an independent random draw of the assigned wind-direction values, within the sector's angular range, to the time steps at which the measured wind direction fell outside the sector. All other inputs and model settings were kept identical. The figure below shows the resulting cumulative NEE for the eight sectors and for the combined wheat sector (4+5+6).



Cumulative NEE reconstructed for each wind sector and for the combined wheat sector (4+5+6) with five independent random wind-direction draws (colored lines, with their uncertainty envelopes as shaded areas). The mean annual budget and the standard deviation and range across the five draws (gC m^{-2}) are given in each panel.

The five reconstructions are practically indistinguishable: in every sector, the five curves overlap throughout the year, and so do their uncertainty envelopes. Across the five draws, the annual cumulative NEE varies with a standard deviation of only 0.24 to 1.09 gC m^{-2} and a total range (maximum minus minimum) of 0.63 to 2.43 gC m^{-2}, depending on the considered sector, whereas the annual budgets themselves range from −370 to +260 gC

m^{-2}. In relative terms, the range between draws represents 0.4 to 1.2 % of the annual budget, and 2.1 % for Sector 3 only since its budget is the closest to zero (–43 gC m^{-2}). This spread is more than one order of magnitude smaller than the width of the uncertainty envelopes (shaded areas) and far smaller than the contrasts between sectors.

We can therefore conclude that the random assignment of wind-direction values has no significant effect on the predicted fluxes.

22.- Line 477: The number of trees was fixed at 300.

Done

23.- Line 478: Is mtry the absolute value of the squared root of p? Please rewrite the equation's characters for clarity. Also, if p=8, is mtry 2 because it is rounded to the lowest, close integer number?

Done. The equation is now written explicitly as mtry = max(1, ⌊√p⌋), with ⌊·⌋ defined in text as the floor function (rounding down, not to the nearest integer), and the worked example for p = 8 (⌊√8⌋ = 2, mtry = 2) is spelled out.

24.- Line 489: within the 300 trees

Done

25.- Line 544: Site conditions, measured NEE dynamics and …

Done

26.- Line 545: You could remove the first sentence, as it is clear you will show the data you measured over the period previously explained in methods.

Done

27.- Line 547: with the studied site

Done

28.- Line 552: harvest interventions / harvest events

Done

29.- Line 557: remove "consistent"

Done

30.- Line 559: Maybe it is not necessary to show the data prior to full filtering; this could be confusing for the later results. Besides, you already showed filter effect in Methods.

Indeed, Referee 1 is right, Fig. 5 has been modified showing solely CO_2 flux values after applying u_* and hard-limits filter. Filter effects are shown in Sect. 2.2.2 along Fig. 3 showing comparison between NEE values before and after filtering methods.

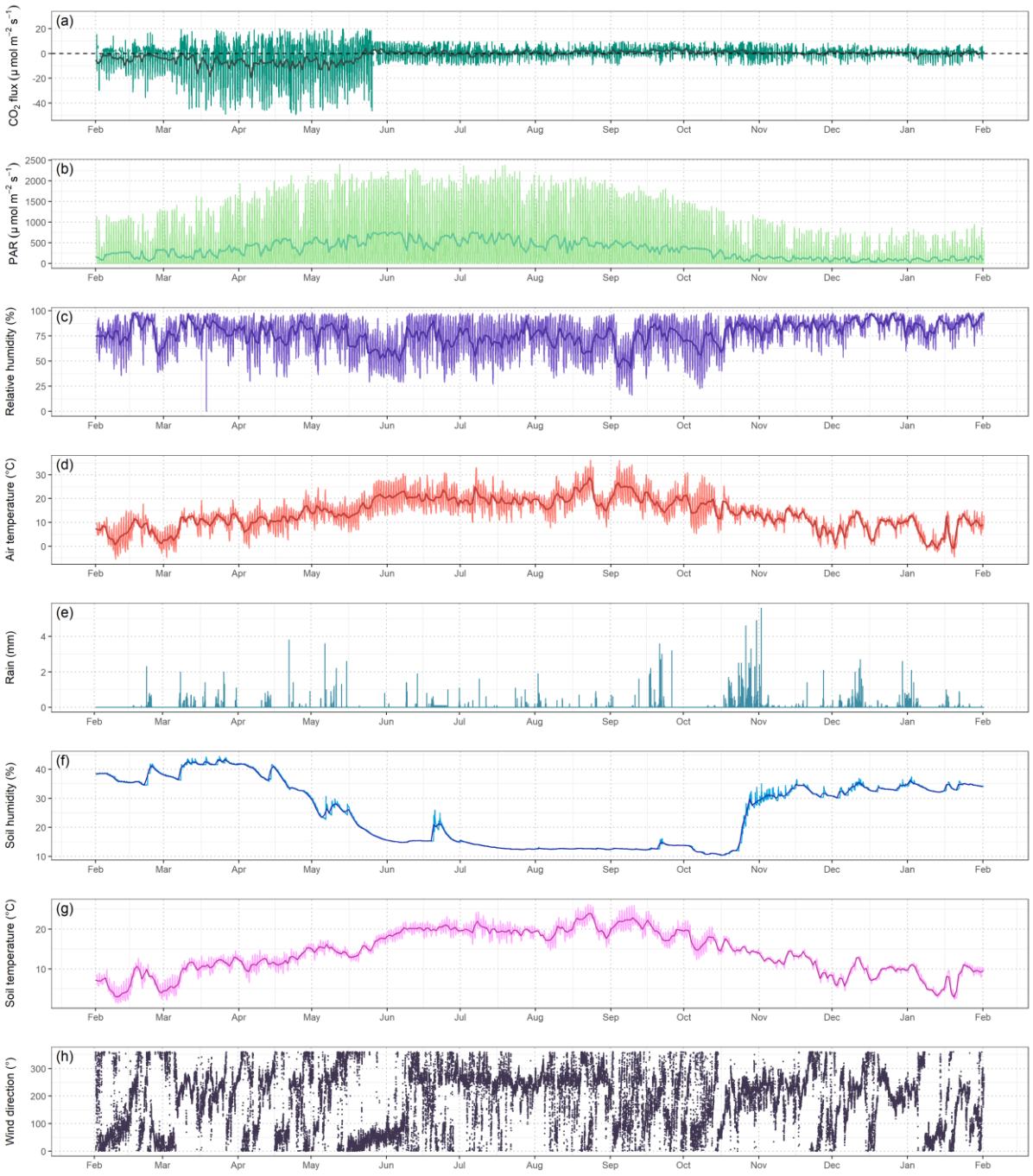


Figure 5. Atmospheric Eddy Covariance measured time series from 01 February 2023 to 31 January 2024 of (a.) atmospheric CO_2 flux (NEE), (b.) Photosynthetically Active Radiation (PAR), (c.) relative humidity, (d.) air temperature, (e.) rain, (f.) soil humidity, (g.) soil temperature and (h.) wind direction. For CO_2 fluxes, PAR, relative humidity, air temperature, soil humidity and soil temperature, the continuous line represents the daily average obtained from the 10 min values. CO_2 flux values correspond to filtered EC data after applying u_* and hard-limits filters as explained in Sect. 2.2.2 (see Fig. 3 too).

31.- Line 579: Please check if "MODIS" products need a citation.

Didan, K.: MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V061, NASA EOSDIS Land Processes Distributed Active Archive Center [data set], https://doi.org/10.5067/MODIS/MOD13Q1.061, 2021. Added to MS

32.- Line 610: "Nevertheless, the global model…"

Done

33.- Line 653: built from 100 resampled RF models

Done

34.- Figure 8 would benefit from adding the points on top of the boxplots as scatters, to better visualize the distributions. Also, the text is very small, please increase its size.

35.- Line 671: Please avoid repetition of methods; just jump directly to what we can see in the figure/table. The same applies to captions of Table 3 and Figure 9.

Done

36.- Line 709: This paragraph could be removed, as it is clear that in the next section you will discuss your results.

Done

37.- Line 714: is required for a heterogeneous, single EC-tower site?

Done

38.- Generally, the first part of discussion is difficult to read. Please try to be concise and go straight to the major discussion points.

Done

39.- Line 781: as commented previously, please avoid writing "Kowalski-partitioned" or in general, mentioning methods that were already explain in the corresponding section.

Done

40.- Line 803: in this paragraph and throughout discussion, please avoid writing results. If they were written before in section 3, remove them; otherwise, if these numbers are important, write them in the results section.

41.- Line 1032: remove "," after sampling density

Done

42.- Line 1151: the single-tower approach is typically not considered cost-effective, in the sense that conventional EC is costly. It is right that the approach saves money in comparison to several conventional EC systems, but I would suggest to remove this term here.

Done

43.- Table B1: Please unify letter size across the table.

Done

Building a random forest machine learning model for Net Ecosystem Exchange estimation in agricultural fields using discontinuous atmospheric Eddy Covariance measurements by Camille Pery et al.

This is my second review of the manuscript following revision. The authors have clearly put substantial effort into addressing the previous comments, and the manuscript has improved in terms of structure, clarity, and presentation. The study addresses an interesting and relevant question, and the combination of eddy covariance measurements, wind-sector partitioning, and machine-learning-based reconstruction has the potential to provide a useful framework for investigating carbon dynamics in heterogeneous agricultural landscapes. However, several of my previous concerns remain insufficiently addressed, particularly those related to uncertainty quantification. Throughout the manuscript, uncertainty estimates are derived from bootstrap RF ensembles, yet many conclusions are subsequently framed in terms of confidence, robustness, precision, and well-constrained annual carbon budgets. I remain unconvinced that the reported uncertainty estimates support these stronger claims. Given the extensive reliance on gap-filled data and the large differences in observational support among sectors, it remains difficult to assess the confidence that can reasonably be placed in the reported NEE, GPP and Reco estimates and in the ecological interpretations derived from them.

Major comments

The manuscript presents uncertainty estimates derived from an ensemble of 100 bootstrap Random Forest models and refers to the resulting intervals as confidence or prediction intervals. However, these intervals appear to represent only the variability among RF model realizations and therefore capture only one component of the overall uncertainty associated with the reconstructed NEE time series. Several potentially important sources of uncertainty do not appear to be included, including: uncertainty in the EC flux measurements themselves, including u* filtering sensitivity, uncertainty associated with footprint attribution and sector delineation, uncertainty arising from assigning observations to specific wind sectors, and uncertainty associated with the subsequent NEE partitioning procedure.

This issue is particularly important because many sector-period combinations rely heavily on reconstructed rather than directly observed fluxes (Table 1), with gap-filled proportions frequently exceeding 90% and reaching up to 98%. Consequently, the reported annual NEE budgets are driven largely by model predictions, making a comprehensive treatment of uncertainty essential for assessing the robustness of the conclusions. More generally, I am left wondering to what extent the annual sector-specific NEE estimates can be interpreted as robust ecological quantities given the very large differences in sample size among sectors and the extremely high proportion of reconstructed data in several cases. The manuscript places considerable emphasis on differences in annual NEE among crops and sectors; however, it is not yet clear whether these differences are sufficiently constrained by observations or are primarily a product of the modelling framework. Given the extremely high proportion of reconstructed data for several sectors, please provide evidence that the resulting annual NEE estimates are robust to the modelling choices employed. While the modelling framework is carefully developed, it remains unclear how robust annual cumulative NEE estimates can be when they are derived from such limited observational support. The authors should discuss this limitation more explicitly and justify why annual sector-specific carbon budgets remain reliable despite the very high degree of extrapolation required.

In addition, GPP and Reco are derived from RF-reconstructed NEE, yet no uncertainty estimates are reported for these quantities. It is unclear how uncertainty in the NEE reconstruction propagates through the light-response partitioning approach. A possible solution would be to apply the partitioning procedure to each bootstrap RF realization separately and derive confidence intervals for annual GPP and Reco from the resulting distributions. I therefore encourage the authors to clearly define what is and is not captured by the reported uncertainty intervals, discuss the uncertainty sources that are not represented, and, where possible, propagate RF reconstruction uncertainty through to the final NEE, GPP and Reco estimates. Without this information, it is difficult to assess the confidence that can be placed in the reported differences among sectors and land-use types.

I am also concerned about the treatment of wind direction during reconstruction. Wind direction is included as a predictor and justified as accounting for directional heterogeneity in source areas. However, when reconstructing sector-specific time series, wind direction is randomly assigned from within the sector boundaries whenever the actual wind direction falls outside the sector. Please report the variable importance of wind direction and evaluate the sensitivity of reconstructed NEE to this random assignment. If predictions are insensitive to the selected wind-direction value, this should be demonstrated quantitatively. If they are sensitive, the rationale for assigning random wind-direction values requires further justification.

The explanation provided in Sect. 4.1 for the differing behaviour of the global and sector-specific models relies heavily on the assumption that wind direction is only weakly captured by the global RF model because it competes

with other predictors during tree construction. However, this interpretation appears speculative and is not directly supported by the analyses presented. I encourage the authors to provide additional evidence supporting this claim, for example through the variable importance of wind direction in the global model, partial dependence analyses, or other diagnostics illustrating the extent to which the global model is able to recover sector-specific structure from the wind-direction predictor.

We thank the Referee 2 for the careful and constructive second review of our manuscript. First, we would like to recall that the issue of uncertainty and the limitations associated with the reconstruction of discontinuous EC measurements were already explicitly addressed during the first revision, following Referee 2's initial comments. In response to the first round of review, we substantially revised the manuscript and implemented several additional analyses. In particular, we:

- introduced a bootstrap-based uncertainty analysis using 100 RF ensembles to quantify the uncertainty associated with the RF reconstruction;
- explicitly acknowledged that this uncertainty estimate does not represent the total uncertainty of the reconstructed carbon budget, and revised the Discussion section accordingly;
- performed the partitioning of reconstructed NEE into GPP and Reco for all eight sectors and for the complete wheat sector;
- expanded the discussion of the uneven distribution of observations among sectors and of the implications of reconstructing NEE from discontinuous EC measurements;
- substantially revised the interpretation of the global versus sector-specific RF models, avoiding any inference of distinct mechanistic controls from differences in model performance;
- and revised the terminology throughout the manuscript to avoid overstating the degree of "robustness", "precision" or "confidence" that can be attributed to the reconstructed annual budgets.

Thus, the revised manuscript already explicitly distinguishes between the uncertainty associated with the RF reconstruction and the broader sources of uncertainty affecting the reconstructed NEE, GPP and Reco budgets. The bootstrap intervals are not presented as a comprehensive estimate of total uncertainty, but as a quantification of the variability associated with the RF reconstruction conditional on the available observations and predictors.
We also emphasized that the objective of the reconstruction in the present study was not to provide an error-free estimate of the annual carbon budget, but to use the available discontinuous EC observations to reconstruct the temporal dynamics of CO_2 exchange and investigate their spatial and temporal variability across contrasting land-cover and management conditions. Accordingly, the ecological interpretation in the revised manuscript was based primarily on the temporal patterns of the reconstructed fluxes and their consistency with the available observations and environmental and management drivers, rather than on the bootstrap intervals alone.

Nevertheless, for this second round of review, we appreciate Referee2's request to further examine how the reconstructed annual budgets depend on the amount of observational support. Therefore, we have complemented the analyses already introduced during the first revision with the additional sensitivity analysis described below, specifically designed to address this remaining concern. Referee 2's remaining concern can be formulated as a single, well-defined question: **given the high proportions of gap-filled data reported in Table 1 and the large differences in sample size among sectors, to what extent can we be confident that the reconstructed annual NEE budgets, and the contrasts drawn between sectors, are robust rather than artefacts of the modelling framework?**
We take this concern seriously and have re-revised the manuscript accordingly for this second round of revision. We have used it as an opportunity to (i) show, with a new figure integrated into the manuscript, that the "scarcity" suggested by the raw percentages in Table 1 does not necessarily translate into poor temporal coverage of the annual cycle; (ii) quantitatively assess how much, or how little, our reconstructions and annual budgets depend on the amount of training data available; and (iii) revise the manuscript's wording throughout so that statements referring to "robustness", "precision", or "confidence" are explicitly linked to the specific evidence supporting them, rather than implying a level of certainty that the bootstrap intervals alone cannot provide.

1. Temporal coverage underlying the gap-filled percentages (new Figure 7)

Table 1 reports, for each sector and period, the percentage of the annual time series that had to be reconstructed rather than directly observed. Considered in isolation, these values understandably raise the question highlighted by Referee 2. To place these percentages in their appropriate temporal context, we have added a new figure and

associated text (Figure 7 and l. 561-576 in the revised manuscript, inserted in Section 3.2) showing, for each of the eight wind sectors, the complete annual time series of directly observed 10-min NEE values together with the corresponding RF reconstruction, from 1 February 2023 to 31 January 2024.

This figure shows that, although the total number of retained observations differs substantially among sectors, the direct measurements are temporally distributed throughout the annual cycle in all sectors. In particular, key periods including peak growth, senescence, harvest, and the post-harvest respiration regime are represented by direct observations. Thus, a sector may contain a comparatively small number of observations in Table 1 while still having observations distributed across the main phases of the annual phenological cycle. We consider this temporal distribution to be more informative for constraining the RF reconstruction than the overall proportion of gap-filled data considered alone.

2. How strongly does the reconstruction depend on sample size?

To move beyond this qualitative assessment, we designed a direct sensitivity test addressing Referee 2 concern regarding the influence of sample size. We selected our best-sampled sector (Sector 6, wheat; n = 12,871 retained observations) as the "Full" dataset, and generated two additional training datasets by randomly subsampling these observations: a "Medium" dataset (n = 6,254) and a "Small" dataset (n = 1,207), the latter being closer to the sample sizes available for some of the more sparsely sampled sectors.

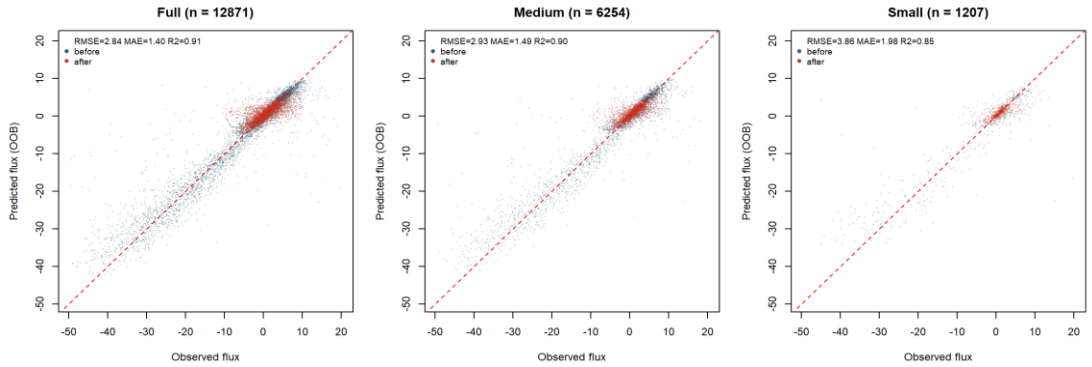
We then independently re-ran the complete RF pipeline for each of the three training datasets, using the same predictors and the same bootstrap procedure, and compared (a) their out-of-bag predictive performance and (b) the resulting annual cumulative NEE values for Sector 6.

The results are, in our view, conclusive:

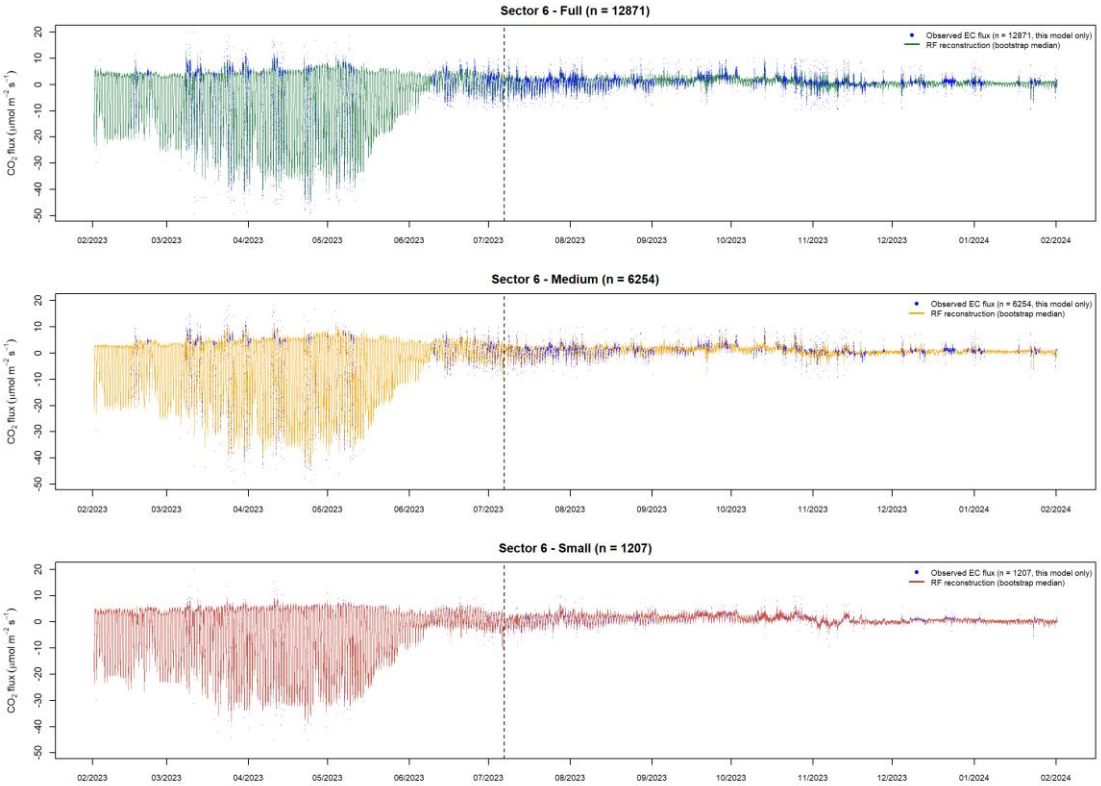
Training set	n	OOB RMSE (µmol m^{-2} s^{-1})	OOB R^2	Cumulative annual NEE (gC m^{-2} yr^{-1})
Full	12871	2.84	0.91	−368
Medium	6254	2.93	0.90	−391
Small	1207	3.86	0.85	−377

Reducing the training sample by approximately one order of magnitude (Full → Small) resulted in only a moderate decrease in point-prediction performance (R^2 from 0.91 to 0.85, with a 36% increase in RMSE), while the reconstructed annual cumulative NEE values differed by only approximately 2-6%. Interestingly, the Full and Small reconstructions were in fact closer to each other in terms of cumulative annual NEE values than either was to the Medium reconstruction.

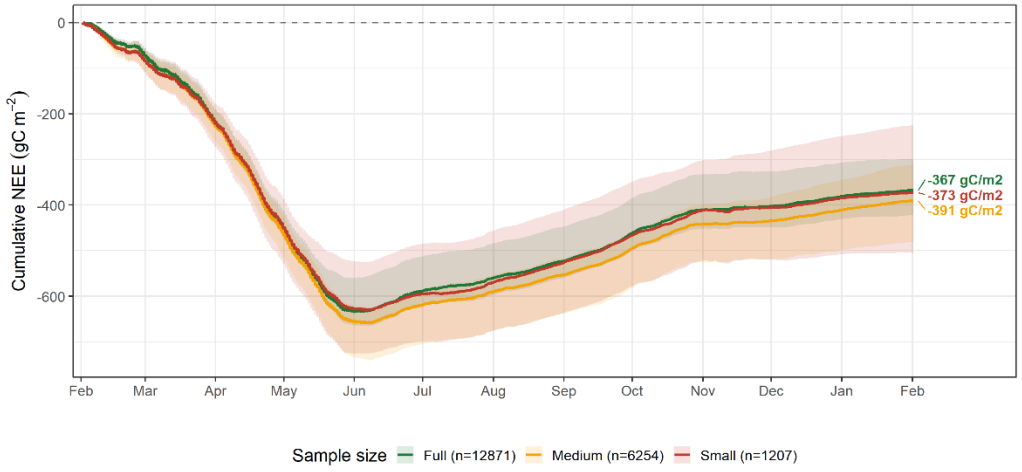
These results indicate that the absolute number of observations, and therefore the percentage of gap-filled data reported in Table 1, is not by itself a reliable indicator of the degree to which the annual NEE is constrained. Rather, the temporal and phenological representativeness of the retained observations appears to be an important factor. This is illustrated by the new Figure 7, which shows the distribution of direct observations throughout the annual cycle for all sectors, including those with comparatively fewer observations (Sectors 2, 3, 4, 5 and 7).



Out-of-bag observed vs. predicted flux for the Full, Medium and Small training sets (Sector 6).



Full annual reconstruction of Sector 6 NEE for the Full, Medium and Small training sets.



Cumulative annual carbon budget for Sector 6, by training sample size.

3. Implications for the interpretation of the reconstructed annual NEE values

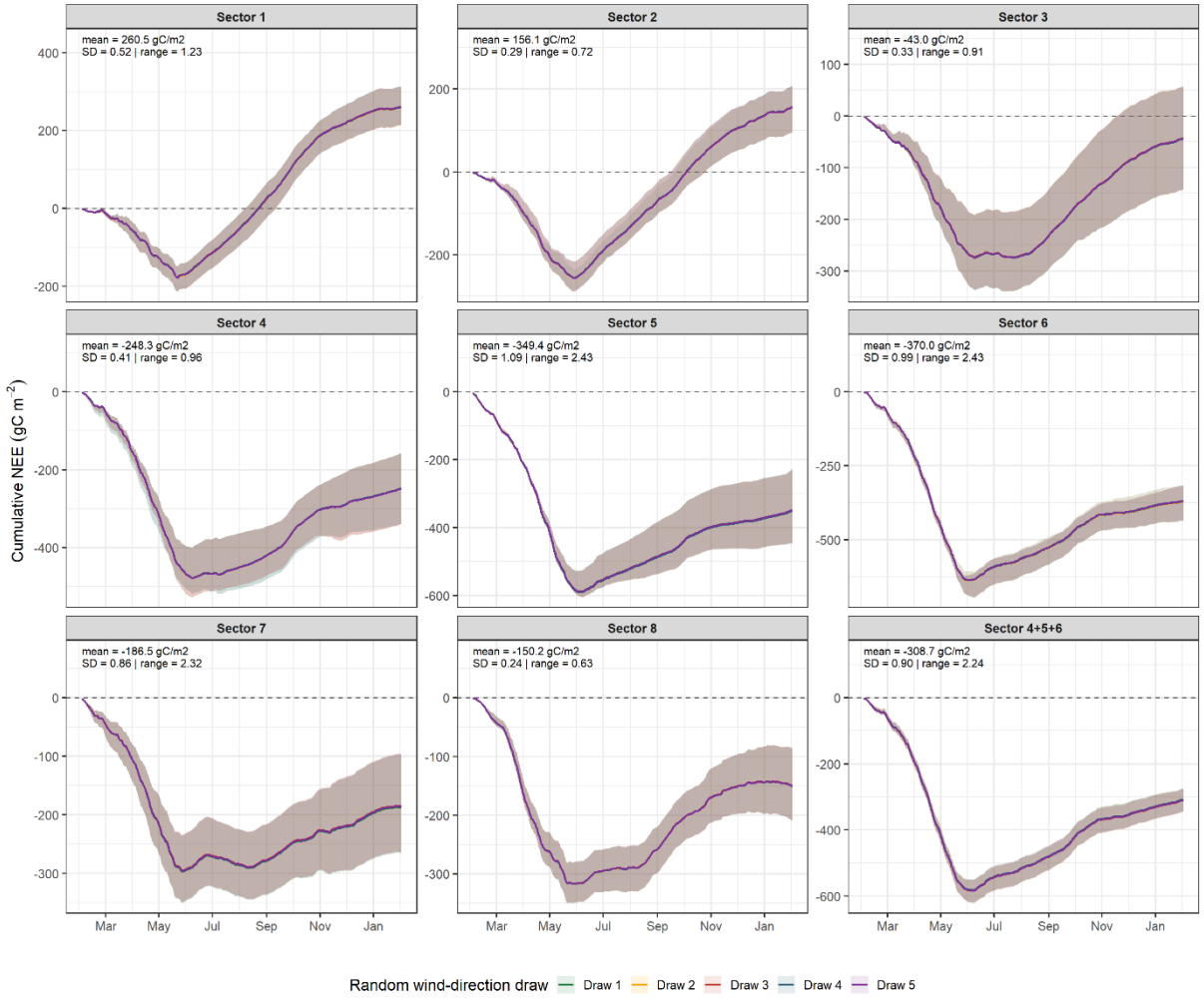
Taken together, the new Figure 7 and the sample-size sensitivity analysis support a more nuanced conclusion that we now state explicitly in the revised manuscript. The qualitative and semi-quantitative patterns identified in our study, including which sectors behave predominantly as net sources or net sinks, the broad phenological contrasts among wheat, mixed-grain and grassland, and the consistency observed among the sector-specific subdivisions of the wheat field, are not strongly dependent on the amount of training data available.

Importantly, we do not interpret this analysis as demonstrating that the annual sector-specific carbon NEE values are free from uncertainty, nor that the bootstrap intervals provide a complete assessment of their uncertainty. Rather, the analysis specifically addresses Referee 2's concern regarding the potential dependence of the reconstructed annual NEE values on the amount of observational support available for model training. Within this context, the relatively limited sensitivity of the reconstructed annual NEE values to a substantial reduction in training sample size provides evidence that the main patterns identified in the study are not simply a consequence of the different numbers of observations retained in each sector.

We have therefore revised the wording throughout the manuscript to ensure that the strength of our conclusions remains consistent with the level of observational support and with the scope of the uncertainty analysis. In particular, sector-specific annual values are now presented as model-based reconstructions constrained by the available EC observations, rather than as direct observations of annual NEE values. We also explicitly acknowledge that sectors with limited observational support remain more strongly dependent on model extrapolation and should therefore be interpreted with greater caution when considering absolute annual NEE values.

4. Sensitivity of the reconstruction to the random wind-direction assignment

Referee 2 (as Referee 1 too) asks us to evaluate the sensitivity of the reconstructed NEE to the random assignment of wind-direction values and, if the predictions are insensitive to it, to demonstrate this quantitatively. We carried out a dedicated sensitivity test for this purpose.
For each sector, the complete annual reconstruction was repeated five times, each time with an independent random draw of the wind-direction values assigned, within the sector's angular range, to the time steps at which the measured wind direction fell outside the sector. All other inputs and model settings were kept identical. The figure below shows the resulting cumulative NEE values for the eight sectors and for the combined wheat sector (4+5+6).



Cumulative NEE values reconstructed for each wind sector and for the combined wheat sector (4+5+6) with five independent random wind-direction draws (colored lines, with their uncertainty envelopes as shaded areas). The mean annual NEE values and the standard deviation and range across the five draws (gC m^{-2}) are given in each panel.

The five reconstructions are practically indistinguishable: in every sector the five curves overlap throughout the year, and so do their uncertainty envelopes. Across the five draws, the annual cumulative NEE varies with a standard deviation of only 0.24 to 1.09 gC m^{-2} and a total range (maximum minus minimum) of 0.63 to 2.43 gC m^{-2}, depending on the sector, whereas the annual budgets themselves range from –370 to +260 gC m^{-2}. In relative terms, the range between draws represents 0.4 to 1.2 % of the annual budget, and 2.1 % for Sector 3 only because its budget is the closest to zero (–43 gC m^{-2}). This spread is more than one order of magnitude smaller than the width of the uncertainty envelopes (shaded areas) and far smaller than the contrasts between sectors.
We can therefore conclude that the random assignment of wind-direction values has no significant effect on the predicted fluxes.

Additional comments

• Was soil moisture and temperature recorded separately for the three different crop fields? If not, how can their inclusion as sector-specific predictors be justified? The manuscript argues that NDVI could not be used because the fluxes represent footprint-integrated measurements rather than pure-field observations. However, a similar argument could potentially be made for soil temperature and soil moisture if these variables were measured only at a single location.
We already modified the revised MS with regards to the community referee comment (see online comment's response and the re-revised MS l.392-397).

• Why use a day/night indicator when PAR is already included as a predictor?

We already answered this question in our Referees 1 and 3 comments' responses (see associated online files).

• Was mtry tuned or sensitivity-tested? If not, why was the default heuristic adopted rather than selecting mtry based on predictive performance?
mtry was not tuned or sensitivity-tested. It was fixed once, for all sectoral and global models, using the heuristic mtry = max(1, floor(sqrt(p))) (p = number of predictors, giving mtry = 2), applied identically across all bootstrap replicates. We did not tune it, mainly for computational-cost reasons given the number of models (13 sectoral/period models + global, each with up to 100 bootstrap replicates).

• The description of the OOB validation procedure is unclear. Because the analysis includes both the internal RF bootstrap used to generate OOB predictions and an additional layer of B = 100 bootstrap-resampled RF models, please clarify how RMSE, MAE and R^2 were calculated. Were performance metrics computed separately for each bootstrap model and then averaged, or were predictions aggregated across bootstrap models prior to calculating the performance metrics?
Predictions were aggregated across bootstrap models before computing performance metrics, not computed per model and then averaged. Specifically, for each observation, we collected its OOB prediction from every bootstrap replicate for which it was out-of-bag, took the median of these predictions as a single aggregated OOB prediction per observation, and then computed RMSE, MAE and R^2 once, across the whole dataset, by comparing observed values to these aggregated OOB predictions. No per-replicate metric was computed or averaged. We have clarified this explicitly in Sect. 2.3.4 (see modified text, l.437-446)

• The uncertainty intervals seem to characterize bootstrap model variability, yet differences among modelling choices produce much larger changes in annual NEE estimates. Please clarify what uncertainty is and is not captured by the reported intervals.
We agree, and have added an explicit clarification at the start of Sect. 4.2 (l. 736) stating what the reported bootstrap 95% confidence intervals (IC95) do and do not represent.

• Figure 8 presents uncertainty at the sector level only. Since the main conclusions of the manuscript are ultimately discussed at the crop level (wheat, mixed grain and grassland), I suggest including a corresponding summary figure showing the reconstructed NEE and associated uncertainty for the aggregated crop categories, particularly for the combined wheat sector (4+5+6). This would help readers assess the robustness of the crop-level conclusions and better relate the reported uncertainties to the final annual carbon budgets discussed in the manuscript. This also applies to the discussion, which is hard to follow as references are made towards sectors rather than land cover.
We thank the Reviewer for this suggestion. Throughout the manuscript, the analytical unit is the wind sector rather than the crop parcel. The sector is the unit at which our observations are collected and our models are built, and sector footprints do not exactly match parcel boundaries. Only the wheat sectors lie mostly within a single parcel, while the mixed-grain and grassland sectors integrate contributions from neighbouring surfaces (Sect. 4.4). Presenting results by crop category would therefore suggest pure parcel estimates that our data do not support, so we kept the sector as the reference unit.
We agree, however, that the Discussion was hard to follow, and we revised it. We also state explicitly which sectors represent a single crop and which integrate several surfaces.

• Paragraph starting on line 623: It is unclear what is meant by "precise" in this context. The manuscript does not quantify uncertainty in annual cumulative NEE estimates, but only RF bootstrap uncertainty. Therefore the statement that some annual NEE estimates are more or less precise does not appear to be demonstrated.
Please see our main response above with regard to this comment we particularly addressed and hope it will help Referee 2 to better understand the main objective of our study.

• I am struggling to understand the purpose of Sect. 4.3. The Kowalski partitioning is applied to the RF-reconstructed NEE, therefore the comparison between partitioning-derived NEE and RF-derived NEE does not appear to be a comparison of independent estimates. Please clarify what information is gained from this comparison and in what sense the two methods can be considered independent.
We agree that the two estimates are not independent, since the partitioning is applied to the RF-reconstructed NEE. The comparison is an internal consistency check of the GPP/Reco decomposition, not a validation of the NEE reconstruction. Sect. 4.3 has been revised accordingly, and the term "independent methods" has been removed.

• Section 4.4 starting on line 709: I find this section overly speculative. The mixed-grain and hybrid estimates are repeatedly described as footprint-contaminated and not fully representative of the target parcels, yet the discussion proceeds to attribute the observed patterns to specific mechanisms (land-use history, productivity, microbial processes, respiration dynamics, etc.). These interpretations appear stronger than the evidence presented.
Please see our main response above with regard to this comment we particularly addressed and hope it will help Referee 2 to better understand the main objective of our study.

Minor comments

There are several minor grammatical and stylistic issues throughout the manuscript (e.g. line 224, line 251 and several others).

Line 55: the chain of reasoning would benefit from additional references.
References added to support this reasoning.

Line 78: VPD has already been defined previously.
The redundant definition of VPD has been removed.

Lines 119-120: GPP and Reco have already been defined.
The redundant definition of GPP and Reco has been removed.

Line 138: Please specify which species comprised the mixed-grain crop.

Cereals and legumes generally comprised mixed-grain crop in Charente-Maritime region, it is now specified in the revised MS.

Line 140: Did the wheat field also remain bare after harvest, similar to the mixed-grain field?
Clarified: the wheat field remained as bare/senescent stubble for the remainder of the measurement period following its harvest (7 July 2023). This differs from the mixed-grain field, which remained bare only from its first harvest (24th May 2023) until it was re-sown with mixed grain in autumn 2023.

Lines 151-153: EC and NEE have already been defined.
Redundant definitions of EC and NEE removed.

Lines 179-187: This addition is useful, but I recommend rewriting the paragraph slightly to improve readability.
This paragraph has been reworded for clarity and readability.

Line 189: Please correct the manufacturer information in parentheses.
Corrected: "Campbellsci" → "Campbell Scientific".

Line 220: The sentence beginning with "This parameter" should be rewritten.
This sentence has been rewritten for clarity.

Line 226: By "strong meteorological disturbances", are the authors referring to instrumental artefacts or physiologically plausible flux responses?
Clarified: these disturbances refer to instrumental artefacts (flow distortion, non-stationarity), not physiologically plausible ecosystem responses.

Figure 3: I recommend increasing the font size for readability. The authors may also consider moving this figure to the Supplementary Information.
Font and legend sizes have been increased in the revised figure version we prefer keeping in the MS instead of Supp. Infor. for a better understanding of our methodology procedure as requested by Referees.

Line 293: The sentence following the reference should be rewritten.

This sentence has been rewritten for clarity.

Line 300: The wording “pure agricultural parcels” is unclear. I assume the authors refer to the three crop parcels described above, but the wording could be improved.
Wording clarified.

Line 389: NEE, GPP and Reco are redefined although they were introduced previously.
Redundant redefinition of NEE, GPP and Reco removed.

Line 415 (and Section 3.4): Soil respiration was not investigated in this manuscript and should be excluded from the results section.
Corrected.

Line 446: This sounds like a repetition of the methods.
It has been modified accordingly and shortened to avoid repetition in the Methods wit cross-references to Sect. 2.3.4 as well.

Line 502: From this point onward it would help to provide uncertainty estimates for GPP and Reco.
Acknowledged as an explicit limitation: no bootstrap uncertainty is currently propagated through the GPP/Reco partitioning.

Line 544: The title is not formulated as a question and should be reworded.
Section title reworded from a question to a statement: "Sector-specific modelling is required at a heterogeneous single-EC-tower site."

Line 603: I do not agree with the phrasing that the cumulative annual NEE value for the wheat parcel represents a “genuinely well-constrained” value, given the uncertainty quantification issues discussed above.
Phrasing softened.

Line 607: Adding the references to Kljun et al. (2015) and Kumari et al. (2020) here does not seem appropriate when discussing the authors' own data.
The Kljun et al. (2015) and Kumari et al. (2020) references have been removed from this sentence.

Line 642: The manuscript describes the RF reconstruction and the light-response partitioning as two "independent methods", but the partitioning is applied directly to the RF-reconstructed NEE time series. The agreement between cumulative NEE estimates therefore represents an internal consistency check rather than an independent validation of the GPP/Reco decomposition.
Indeed, it is true as rightly noticed by Referee 2 and revised MS has been modified accordingly (l.xx, p.xx).

Line 672: I do not fully understand this statement. The manuscript appears to compare cumulative NEE derived from the Kowalski partitioning with cumulative NEE from the RF reconstruction, yet the partitioning is applied to the RF-reconstructed NEE time series. Could the authors clarify exactly which datasets are being compared and how the agreement is evaluated?

Line 674: Throughout the discussion, the authors argue that the annual NEE estimates differ in precision, confidence and robustness among sectors. If so, discussing and comparing the absolute NEE values among cropping systems is concerning, particularly when some estimates are explicitly described as less reliable because of footprint mixing.
Please see our main response above with regard to this comment we particularly addressed and hope it will help Referee 2 to better understand the main objective of our study.

Line 697: The meaning of "some residual caution" is unclear.
It has been reworded to state the specific source of uncertainty explicitly.

Line 704: It is unclear what the authors mean by interpreting the mixed-grain estimate "qualitatively rather than quantitatively". If the estimate is not considered an acceptable field-specific quantitative carbon budget, it is

unclear how conclusions regarding its underlying controls (e.g. productivity, soil respiration, GPP/Reco balance) can be drawn.

It is now clarified what is and is not retained from the mixed-grain sector estimate.