RC2: 'Comment on egusphere-2026-524, Anonymous Referee #2, 17 Apr 2026

The manuscript “Building a random forest machine learning model for carbon budget estimation in agricultural fields using discontinuous atmospheric Eddy Covariance measurements” by Pery et al. applies Random Forest models to eddy covariance CO_2 flux data from a heterogeneous agricultural landscape over a full year, comparing a single global model with sector and crop specific models for different crop types. Both approaches use the same meteorological and soil predictors and are evaluated over one year using standard performance metrics. Results show that both strategies capture broad seasonal NEE patterns with generally similar predictive performance, while the crop specific model outperforms the global model when representing phenological stages and harvest events. A major strength of the study is the very thorough and careful analysis of wind sectors and turbulence characteristics, which provides a solid and transparent foundation for footprint attribution and sector based modelling. At the same time, a major limitation of the study is that several key conclusions are driven by differences in model structure, data availability, and temporal segmentation rather than independent evidence of process representation, while incomplete carbon accounting, sparse data coverage for many vegetation types, and the absence of uncertainty quantification limit the strength and generality of the interpretations.

We are grateful to Referee 2 for her/his positive comments on our submitted manuscript in particular on the attention we done with regard to the wind sectors and turbulence characteristics analysis. At the same time and as noticed by Referee 2, we recognize the main limitation of our study approach based in majority on modelled flux computations that definitely need further uncertainties computations/quantifications (as generally noticed by Referees) to avoid too speculative discussion and conclusion and to the contrary to give the real strength of our data interpretations deserved by the approach we propose. Owing to the significant work done by Referee 2 and the others, we deeply address all Referee comments through further computations and deep modifications in the associated coming revised manuscript according to the Co-Editor decision.

The document has (very) few very small grammatical or spelling errors (i.e., line 58 Gaz, instead of Gas, Campbellsci, instead of Campbell Scientific, Logan Utah, etc.). Furthermore, several sections reintroduce acronyms such as NEE or GPP, which only need to be defined at their first mention.

We took into account Referee comment above and modified in the submitted manuscript in this way.

Major concerns

The comparison between the global and sector specific Random Forest models is difficult to interpret. The sector specific models are trained on more homogeneous data and are independently optimized, which inherently increases their flexibility. Their improved performance is therefore largely expected and does not provide evidence for distinct mechanistic controls on NEE. Rather, the comparison demonstrates spatial and management driven heterogeneity within the flux footprint and should be interpreted as descriptive rather than explanatory. This issue is compounded by the strong imbalance in data coverage across vegetation types. Vegetation classes with sparse observations, representing only 2 to 4 percent of the dataset and showing limited variability, exhibit the highest R^2 values, whereas wheat, which represents the largest and most dynamic subset, shows substantially lower performance, particularly post harvest. High R^2 values in sparsely sampled systems are therefore more likely to reflect limited signal complexity than superior predictive skill, and cross vegetation comparisons should be interpreted with caution.

The comparison between the global and sector-specific RF models will be substantially revised in the revised version of the manuscript to avoid any interpretation in terms of distinct mechanistic controls on NEE. The objective of the sectorisation is primarily to characterize the spatial heterogeneity of the flux footprint and the effects of contrasting management practices within the study area.

In addition, the revised analysis we carried out includes several methodological improvements as described above, namely the application of hard limits to remove spurious flux values and the inclusion of the three wheat subsectors (sectors 4, 5 and 6). These modifications significantly improved the robustness of the RF models and provide a more balanced basis for evaluating the impact of footprint heterogeneity on model performance.

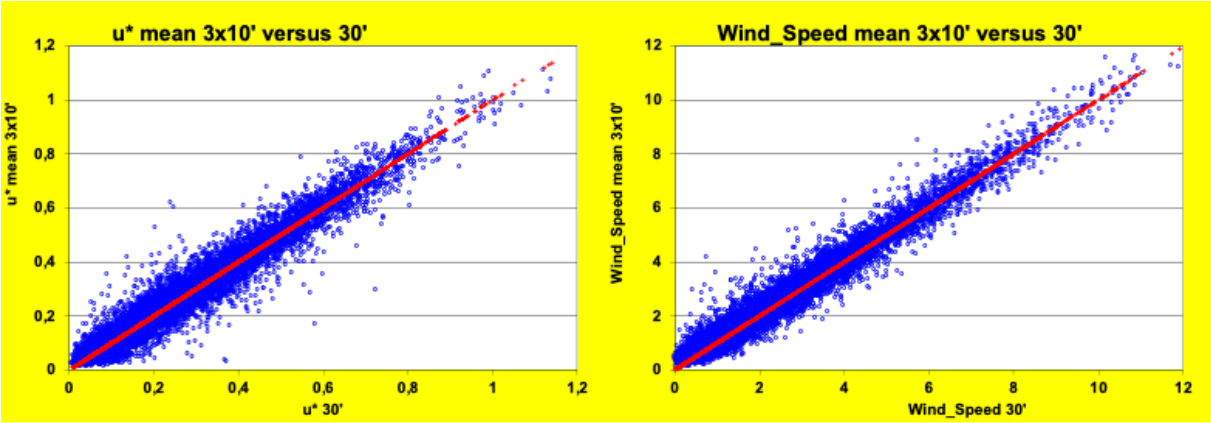
These results suggest the lower performances reported in the original manuscript were largely driven by a limited number of extreme flux values and by methodological choices related to footprint partitioning rather than by

intrinsic differences in ecosystem predictability. Consequently, the revised manuscript will emphasize that sector-specific models should be viewed as a descriptive framework for exploring footprint heterogeneity and management effects, while comparisons of model skill across vegetation types must be interpreted with caution because of differences in sample size, variability, and phenological dynamics.

The use of 10-minute averaging intervals instead of the standard 30 minute eddy covariance averaging period requires stronger justification. Shorter averaging windows may exclude low frequency flux contributions and bias NEE estimates, particularly under stable or disturbed conditions. The manuscript does not provide supporting diagnostics such as ogives or spectral analyses to demonstrate that flux carrying frequencies are adequately retained within the 10-minute interval. In the absence of such evidence, the validity of the calculated fluxes and their subsequent interpretation remains uncertain.

We thank Referee 2 for her/his first comment also mentioned by Referee 3 and Professor Sun. We understand this comment with regard to the 10 min. average flux window we now use since several years, as adapted for more coastal and dynamic systems such as wetlands (present study), salt marshes (i.e. Mayen et al., 2024, bg-21-993-2024) and tidal flats (i.e. Polsenaere et al., 2012, bg-9-249-2012), instead of the commonly used 30 min. averaged flux window for EC data processing over terrestrial ecosystems. The comment is fair since, our study site though being a wetland system, is surrounded by different crop and grassland fields and in any cases, we need, as for all atmospheric EC studies to be sure to measure and catch all eddy that contribute to the fluxes and be careful with the corresponding EC set-up in terms of measurement height and time averaging window for the chosen specific studied site.

To address this point and more strongly justify the 10 min. averaging time period chosen in the present study as rightly noticed, we did further computations comparing the friction velocity, wind speed and flux data between the mean of the three 10 min. values and the corresponding mean values obtained by reprocessing all data (February-September 2023 period) through EddyPro with an averaging time period of 30 min. (see attached graphs below). Although we can see some dispersion between both 10 min. and 30 min. mean values (u*, wind speeds, mean fluxes), we also clearly notice that we did not have any low frequency losses on the 10 min. flux data.



Furthermore, we decided to choose a 10 min. averaging time period instead of 30 min. to lose less data due to non-stationary conditions and/or electronical noises leading to potential and more data suppression (with the 30 min. choice) according to our data quality check procedure adopted in the present study. The new computations and comparisons done between 10 min. and 30 min. data showed that, according to the land parcel, we would lose between 21% (wheat, sector 213°-283° with dominant wind directions) and 49% (wheat, sector 133°-173°) of data with a choice of 30 min. mean flux calculation instead of the 10 min. mean flux procedure used here.

All these considerations and new computations will be added to the revised version of the manuscript to specify and strengthen our choice.

The separation of NEE into daytime and nighttime subsets does not constitute a valid flux partitioning. Daytime NEE cannot be assumed to represent GPP, and nighttime NEE cannot be equated with ecosystem respiration without an explicit partitioning approach. Respiration occurs continuously throughout the day and night, and

eddy covariance does not directly measure GPP. Consequently, cumulative NEEday and NEEnight sums cannot be interpreted as physiological components of the carbon balance. The justification based on changing wind direction does not resolve this conceptual limitation, and the resulting interpretation of metabolic fluxes and annual component budgets therefore requires substantial revision.

We totally agree with Referee 2, that is why we finally carried out the partitioning of simulated NEE into GPP and Reco for each of the 8 sectors in Fig. 2a as well as for the entire wheat sector. All this work will be presented in the coming revised MS.

Given the uneven data coverage, the lack of flux partitioning, and the absence of uncertainty quantification, the study would benefit from a clearer discussion of limitations. Without uncertainty estimates, it is difficult to assess whether differences between crops or modelling strategies are meaningful, particularly during disturbance and post harvest periods when eddy covariance data are highly non stationary.

We thank the reviewer for this valuable comment. We agree that the limitations of the study should be more explicitly discussed. In the revised manuscript, we will expand the Discussion section to better acknowledge the uneven distribution of observations among vegetation types and the implications of using NEE without flux partitioning.

In addition, as detailed in our response concerning uncertainty quantification, we will include a bootstrap-based uncertainty analysis and report confidence intervals for the reconstructed fluxes. This additional analysis will provide a more robust basis for evaluating differences among vegetation types and modelling approaches, particularly during disturbance and post-harvest periods when fluxes are highly variable and non-stationary.

Finally, some of the literature cited in support of the interpretations is more than 15 years old. While not inherently problematic, more recent studies, particularly those addressing grazing systems, may provide more relevant context given the management practices present at the site.

We thank Referee 2 for this suggestion. We agree that incorporating more recent literature would strengthen the discussion and place our findings in a more up-to-date scientific context. In the revised manuscript, we will review and update the references used to support our interpretations, with particular attention to recent studies addressing grazing systems, grasslands, and managed agricultural ecosystems. We believe that these additions will improve the discussion and better contextualize our results within the current literature.

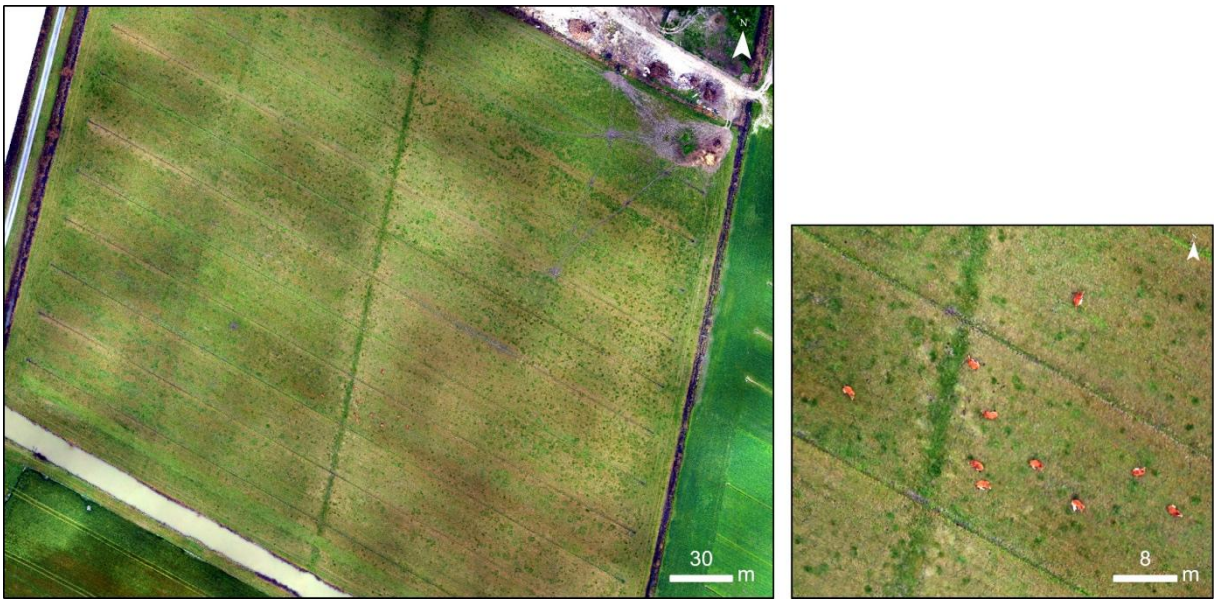
Specific notes

Line 43: Intensive agricultural is typically associated with heavy inputs, which generally don't correlate with increased carbon storages in soils. Perhaps a clarification of which systems are meant here would help

This clarification will be done in the revised version of the MS.

Line 135 please indicate how many cattle were present, as their density might influence the CO2 balance as well.

This information will be specified in the revised manuscript. 10 in February 2023, 41 in April 2023 and 43 in July 2023 cattles were present in the grassland (see the attached February 2023 file for example below).



L 161, measured fluxes were averaged every 10 minutes. This needs more expansion

Indeed, information given above will be added in the revised manuscript.

Section 2.3.2 The previous section describes in great detail how different fields were isolated based on turbulence wind sector analysis, which included some areas such as shading by the anemometer, etc, but in this section the first scenario RF is phrased as if all sectors were now included. But shouldn't at least the ditch and the tower shaded area be excluded from this model? Perhaps this is what was meant here. Further explanation is needed.

Further explanation will be given in the revised MS as rightly suggested by Referee 2 and as explained in our Referee comment responses.

Results

Section 3.1 The manuscript states that vegetation type is included in the global RF model via a qualitative variable (Scenario 1), but this variable is not clearly defined in the predictor list. Clarification is needed, which should be included in the methods section 2.3.3.

Indeed, clarification will be done in the revised manuscript.

Line 369: The term maximum measured productivity could be misleading, please use the standard NEE instead. Further, since this is mentioned for Scenario 1, it should also be mentioned for Scenario 2, as the goal is to compare their performance.

We thank Referee 2 for her/his suggestion that will be used in the revised MS.

Line 373: sentence starting with "However, " do you mean here actually observed in the field, or in a dataset. If dataset, which one is referenced here?

We mean in each corresponding hybrid and mixed grained dataset. This will be more clearly specified in the revised MS.

Line 377: less productive is again a bit misleading, perhaps better "reduces photosynthetic activity"

We thank Referee 2 for her/his more appropriated suggestion here that will be used in the revised MS.

Section 3.3 Given the noise and non-stationarity of EC data, particularly during disturbance and post-harvest periods, the absence of uncertainty quantification limits interpretation and makes it difficult to assess the significance of differences between vegetation types and modelling strategies.

We agree with Referee 2 and as explained above and in our Referee's responses (Referee 1), all these aspects will be added in the revised MS.

Line 422 The description of cumulative NEEdaytime is confusing and potentially misleading. The reported "peak" value refers to a cumulative minimum rather than a flux, and the shift from -844 to -764 gC m^{-2} indicates a reversal toward daytime carbon release after harvest, not merely a slowdown in accumulation. Moreover, interpreting daytime NEE as productivity or photosynthetic uptake is conceptually incorrect without flux partitioning.

Following our partitioning of simulated NEE into GPP and Reco done for these reviews, section 3.3 will be completely rewritten in the revised MS.

Section 3.4 Again for this section including uncertainties is very important in order to distinguish between sinks and sources.

Indeed, and accordingly all these aspects done will be added in the revised MS.

Discussion

Line 509 The interpretation that reduced model performance during storm periods is "not problematic" because the RFs do not reproduce "physically unrealistic fluxes" is not justified. Failure of a Random Forest model to predict certain flux values indicates that these values are not explainable by the chosen predictors, but does not demonstrate that they are artefacts rather than genuine storm-driven ecosystem responses or unmodelled processes. Using model behaviour to infer data validity is circular and should be avoided. This passage would benefit from substantially more cautious language and explicit acknowledgement that these periods represent increased uncertainty rather than invalid fluxes.

This passage will be re-written more cautiously in the revised MS as rightly suggested by Referee 2.

Line 516 This should be noted in the methods, that this period was excluded from cum sum evaluations in section 2.4. Further description is also needed regarding how the models were run with their exclusion.

It will be noted in the M&M section of the revised MS along with the further model descriptions as suggested by Referee 2.

Line 521 The discussion of static and automated chamber measurements is confusing, as no such data are presented or used in this study. While chamber studies from the literature are cited to illustrate potential agreement with EC fluxes, this does not constitute validation of the EC measurements or RF model applied here. This section should be removed or rephrased.

Referee 2 is right and we better specified this paragraph in the revised manuscript as our intention was to say deploying simultaneously static chambers at our study site at the different seasons according to associated crop management and phenology could have been useful to validate our modelling approach and this method should be done in the future.

"In the future, to validate the EC-based flux measurements and Random Forest (RF) modelling approaches at the studied site, static and automated chamber measurements could be deployed simultaneously over each crop at different locations inside the footprint and key time (seasonal, diurnal) periods linked to phenology and management. Indeed, some studies indicate that chamber measurements can reliably reflect EC fluxes at plot scale and can also capture temporal dynamics (e.g. drought responses or post-harvest changes). For instance, Wang et al. (2013) reported near-identical NEE values using EC and automated chambers in winter wheat, with seasonal EC estimates of -251 gC m^{-2} and chamber-derived values of -205 gC m^{-2}. Similarly, Drewer et al. (2017) found consistent patterns between chamber and EC fluxes in intensively managed grasslands. Although chamber data have limited spatial coverage, along EC measurements, both techniques run simultaneously can offer crucial

plot-scale insight and could help verify RF model accuracy, especially under flux footprint conditions that are highly localized."

Line 546 I'm not sure that Reichstein et al. 2005 is an appropriate citation here, since the focus is on riparian zones and higher soil moisture contents.

Referee 2 is right; we will remove the corresponding citation here in the revised MS.

Line 596 The term "behaved" is imprecise and anthropomorphic when referring to a winter wheat carbon balance.

Indeed, it will be modified in the revised manuscript as followed: "By contrast, the established winter wheat field was almost neutral in full carbon accounting."

Line 596 the harvest data should be introduced in the results section.

It will be introduced in the result section of the revised MS as proposed.

Line 604 this sentence is very hard to understand and should be broken up into different sections. Further, it is hard to see the connection to the present study, when the measurement timeframes differ by at least 5 years. Further, different flux pathways, spatial domains, and time scales are mixed without clear signposting, giving the impression of a quantitative comparison with cumulative NEE that is not actually performed. Clearer framing is needed to distinguish contextual background from results included in the present carbon balance.

All this section will be modified accordingly in the revised MS.

Conclusion

The concluding claim that the approach provides a "scalable template for integration into larger-scale carbon inventories" appears overstated, as no spatial or temporal scaling is demonstrated and key components of full carbon accounting are acknowledged to be incomplete. This statement would benefit from more cautious, prospective wording.

It will be rephrased accordingly in the revised MS.