

RC1: 'Comment on egusphere-2026-524, Anonymous Referee #1, 17 Mar 2026

This manuscript presents a novel gap-filling method for CO₂ fluxes measured by a single eddy covariance tower located in the middle of four fields with different vegetation types. The proposed approach relies on machine learning techniques, specifically a random forest algorithm, to reconstruct fluxes when wind direction does not originate from the target area. The random forest model is trained using a set of environmental variables, including temperature, radiation, soil moisture, as well as information related to crop management practices and the day of year. This approach shows strong potential for reconstructing CO₂ fluxes, as well as energy fluxes, in heterogeneous landscapes using measurements from a single flux tower, and is therefore of great interest for extending the spatial representativeness of Eddy Covariance.

We are grateful to Reviewer 1 for her/his positive and encouraging feedback on our work and for all her/his constructive following comments we addressed below.

However, although the method is promising and deserves particular attention, the manuscript would benefit from more details regarding the technicality of the method. It would in particular benefit from the computation of the uncertainties in reconstructed/gap-filled CO₂ flux in each wind sector. This would allow to assess the statistical significance of the differences in CO₂ fluxes observed in each wind sector (specific to a vegetation/soil type).

We thank the reviewer for this constructive suggestion. We agree that the manuscript would benefit from a clearer description of the reconstruction procedure and from an explicit quantification of uncertainty associated with the gap-filled CO₂ fluxes. In the revised version, we will add a detailed description of the sector-specific modelling framework and we will estimate uncertainty for each reconstructed wind sector using a bootstrap approach. Specifically, we will refit the random forest models on multiple bootstrap resamples within each sector, generate a distribution of reconstructed fluxes, and derive 95% prediction intervals from the empirical distribution. This will also allow us to test whether the differences in reconstructed CO₂ fluxes between wind sectors are statistically significant. We will report these uncertainties and clarify their interpretation in the Methods and Results sections of the revised MS.

The authors could also try to use spatially explicit data to evaluate if the observed CO₂ flux dynamics is sound. The use of NDVI or EVI from Sentinel data may be of interest. Due to the lack in uncertainty and significance analysis, in its current form, the discussion remains somewhat speculative regarding the reconstructed CO₂ fluxes for the plots surrounding the tower.

We agree that additional independent information on vegetation dynamics would help support the interpretation of the reconstructed CO₂ fluxes. To address this point, the revised manuscript will include a new analysis based on Sentinel-2 NDVI time series for the three study plots over the entire study period as shown in Figure 1. below:

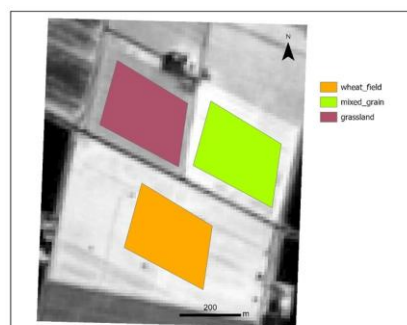


Figure 1. Polygon delineation uses for NDVI estimation

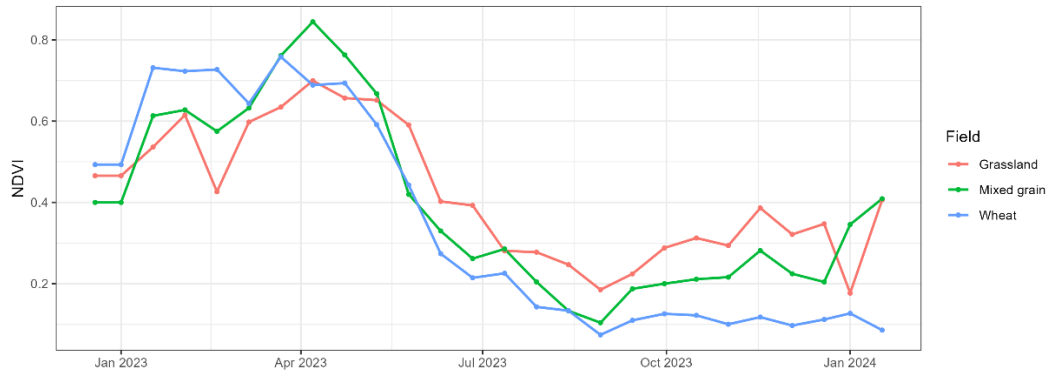


Figure 2. NDVI values of the tree different fields

Also, as shown in Figure 2, we will provide a detailed description of the phenological trajectories associated with the different land uses and management practices. In particular, it clearly highlights the contrasting dynamics of the wheat and mixed grain fields. The mixed grain field exhibits a sharp decrease in NDVI immediately after harvest on May 24th, while the vegetation is still green and physiologically active. Following the autumn sowing, NDVI increases again, reflecting the rapid re-establishment of vegetation cover. In contrast, the wheat field maintains high NDVI values until late June before gradually senescing. Harvest occurred on July 7th when the crop was already dry, resulting in a much smaller NDVI discontinuity. Furthermore, unlike the mixed grain field, no autumn sowing was performed, leaving the soil largely bare until the following spring, as reflected by persistently low NDVI values.

Importantly, these NDVI time series are computed on well-defined “pure” agricultural parcels (see Figure 1, showing the three polygon delineations), and therefore do not include the influence of the riparian vegetation located along the two canals. This distinction is essential, as this riparian vegetation is part of the EC footprint in certain sectors (as discussed in Figure 2a in the submitted MS), but is explicitly excluded from the NDVI-based analysis. For this reason, these NDVI observations cannot be directly used as predictors in the RF model, which is trained on footprint-integrated flux measurements rather than pure-field vegetation indices.

These contrasting vegetation trajectories provide an independent framework for interpreting the reconstructed CO₂ fluxes. For example, the abrupt reduction in vegetation cover caused by the mowing of the mixed grain field should coincide with a marked decrease in CO₂ uptake, whereas the delayed harvest of the wheat crop is expected to have a more limited impact because photosynthetic activity had already largely ceased before harvest. Similarly, the autumn regrowth observed in the mixed grain field and the prolonged bare-soil period observed in the wheat field should be reflected in the reconstructed carbon exchange dynamics.

I would therefore recommend major revision by integrating uncertainties and significance analysis, before interpreting and discussing potential differences between crop types.

Again, we thank Reviewer 1 for all her/his constructive following comments we addressed below to be able to submit after the revised manuscript according to the Co-Editor decision.

Specific comments:

L155-164: in equation ρ should be "dry" air density (ρ_d). and there is a conversion from g to mol missing. Also give units of w and s.

Referee 1 is right and ρ given in the equation (l.155, p.6 in the submitted manuscript) was actually the dry air density air (ρ_d). We modified accordingly the equation and gave w and s units as requested:

“The method relies on measuring the covariance between fluctuations in vertical wind velocity and scalar concentrations within these eddies. The CO₂ flux can be expressed as:

$$F = \overline{\rho_d w s} \approx \overline{\rho_d w' s'}$$

where the overbar indicates the time averaging and the prime represents fluctuations around this average (Burba, 2013; Reynolds, 1886). Here, (ρ_d) denotes the dry air density (kg m^{-3}), (w) the vertical wind velocity (m s^{-1}), and (s) the dry mole fraction of CO_2 (mol mol^{-1} , commonly expressed in $\mu\text{mol mol}^{-1}$ or ppm). The covariance between these variables yields a mass flux with units of $\text{kg m}^{-2} \text{s}^{-1}$. For CO_2 exchange studies, this mass flux is converted into a molar flux by dividing by the molar mass of CO_2 ($M_{\text{CO}_2} = 44.01 \text{ g mol}^{-1}$) and is therefore commonly reported in $\mu\text{mol m}^{-2} \text{s}^{-1}$. In this study, measured fluxes were averaged every 10 minutes to maximize the number of observations. Negative CO_2 fluxes indicate the ecosystem acts as a CO_2 sink, while positive fluxes indicate ecosystem source behaviour.”

L190: why using a time delay with colocated $\text{CO}_2/\text{H}_2\text{O}$ and sonic instruments?

Referee 1 comment is fair especially in the case of the Campbell Irgason sensor (spatially close CSAT3A and EC150 sensors) as used in the present study. Actually, in EddyPro, we used the covariance maximization option for the time lag detection method that calculates the most likely time lag based on the circular correlation procedure that is recommended in most situations. Thus, using this option instead of no time lags compensation method at all, we avoided possible flux underestimations in the case of existing potential time lags.

L200: the u^* threshold is usually adapted to local conditions. Justify the use of a single value.

As responded to Referee 3 too, in the revised manuscript, we applied additional quality-control procedures to the CO_2 flux time series to remove physically unrealistic values that were not adequately detected by the standard filtering approach. While the conventional filters based on the median absolute deviation (MAD) and the commonly used threshold of $u^* < 0.1 \text{ m s}^{-1}$ successfully removed a large fraction of low-quality measurements, several extreme flux values remained in the dataset, particularly during periods of strong meteorological disturbances.

To address this issue and as it will be presented in the revised MS, we now introduced hard limits based on the observed range of realistic ecosystem fluxes. Before harvest, NEE values were constrained between $+20$ and $-50 \mu\text{mol m}^{-2} \text{s}^{-1}$, whereas after harvest, when ecosystem activity was substantially reduced, a range of $+10$ to $-10 \mu\text{mol m}^{-2} \text{s}^{-1}$ was applied. These limits removed a small number of spurious observations, including the most extreme values associated with the November storm period (see Fig. 3).

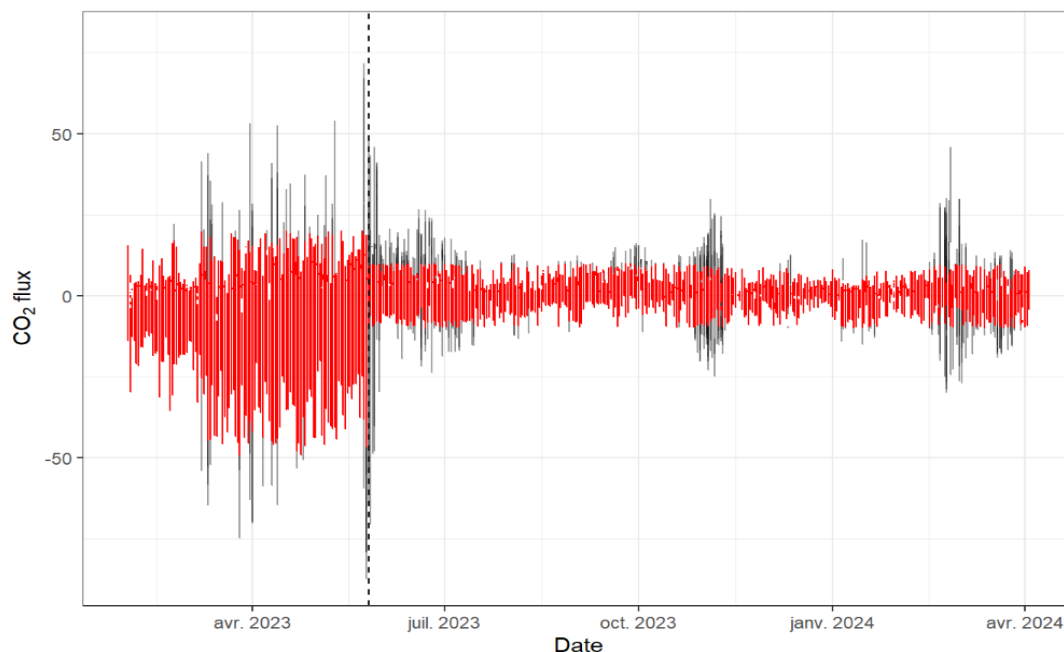


Figure 3. CO_2 fluxes presented in the submitted manuscript (black) and CO_2 with hard limits filtration (red) introduced in the revised manuscript.

The effectiveness of this additional filtering is particularly evident for the wheat sector after harvest, which was the most strongly affected by these extreme values. In the original analysis, the RF model achieved an R^2 of 0.44

for this period. After applying the hard limits, the model performance increased to an R^2 of 0.97, indicating that a limited number of unrealistic observations were disproportionately degrading the model calibration.

Finally, to better understand the RF model's performances on wheat, we tested it separately on the 3 subsectors we had defined in Fig. 2a (Sector 4: 133°-173°; sector 5: 173°-213°; sector 6: 213°-283°), including sector 5, that we previously excluded due to u^* contamination from the instrumentation wake; this inclusion required recalculating the u^* values, needed to filter the NEE data. At the end, the comparison of these 3 subsectors with those of the entire wheat sector (133°-283°) is now possible. All these changes will be specified in the revised MS.

Figure 3 (L219): you could select only neutral conditions here to have a closer relationship with the roughness z_0 .

Estimating the roughness length z_0 from the roughness coefficient u^*/u would indeed require using the following equation under neutral conditions:

$$u(z) = (u^* / k) \ln((z - d)/z_0)$$

However, that's not the focus of this study, where we need to determine the spatial origin of the measured CO_2 fluxes (NEE) under all atmospheric stability conditions, with Reco being linked to stable night-time conditions and GPP being the main component of daytime NEE under unstable conditions. Using only neutral conditions to determine u^*/u could therefore give a biased view of the sectors where CO_2 fluxes come from.

L244-250: this is somewhat redundant with previous paragraph.

We will modify it in the revised MS to avoid possible redundancy with regards to this information.

L265-270: the temporal subdivision, and all parameters in the RF approach, should be explicated in a table (in appendix eventually).

We will explicit it adding a new table in the revised MS as suggested.

L280: could you explicit what is a qualitative day-night indicator?

The qualitative day-night indicator distinguishes periods of daytime and nighttime activity by assigning "day" to time steps with $PAR > 10 \mu\text{mol m}^{-2} \text{s}^{-1}$ and "night" to those with $PAR \leq 10 \mu\text{mol m}^{-2} \text{s}^{-1}$.

L282: PAR also captures seasonal changes in radiation regimes? Is this not redundant?

The choice and the importance of all these variables used in each model was checked as specified in the submitted MS and the associated Appendix A.

Section 2.3.3 : in the predictive model, a measure of the leaf area index would be very valuable and would integrate spatial variations. I would recommend trying using an NDVI or EVI high resolution satellite product (from Sentinel data).

Please see our response and information given above according to this comment.

Section 2.3.4: this is the critical point in this manuscript. I would suggest the authors to compute an uncertainty for the RF reconstructed CO_2 fluxes. This would be necessary to interpret the results and determine if significant differences can be retrieved from each wind sector. I am not an expert in RF modelling but I guess some uncertainty could be computed. At least from the comparison between model and observed data uncertainties could be deduced.

Please see our response and information given above according to this comment too.

L314-318: I would recommend avoiding any use of GPP here as NEE_{day} is not GPP. This paragraph could be rephrased to clarify that NEE was simply split between day and night

Since we finally carried out the partitioning of simulated NEE into GPP and Reco, following Referee recommendations, this paragraph will be completely rewritten in the revised MS.

3.1 section: The Figures 7 and 8 would benefit from showing the difference between the modelled and observed CO₂ flux rather than each flux on top of each other. Actually the two may be shown : the modelled fluxes and the difference with observations. This graph should include uncertainties in the modelled fluxes.

We thank Referee 1 for her/his comment and modified Figures 7 and 8 accordingly adding modelled-measured flux differences.

The rest of the results and the discussion provides valuable arguments but would really need to have uncertainties to evaluate if the differences discussed are significant or not, especially when comparing annual budgets.

We agree with Referee 1 with regards to the need to add uncertainties for annual budget comparisons; accordingly, we have done further computations in this way to address this comment and be quantitatively more accurate showing atmospheric CO₂ flux dynamics over each parcel and corresponding annual budgets (see information within our response given above).