Author Comments – Response to Referee 3

Referee comments are marked in black and author responses are marked in blue.

Grasslands constitute a significant portion of agricultural landscapes in Europe and store substantial amounts of carbon. Effective management of these ecosystems serves as a nature-based solution for enhancing carbon sequestration. This study presents a unique long-term dataset spanning 20 years, encompassing flux, biometeorological, and detailed management data from an eddy covariance site in a Swiss grassland. It addresses the gap in understanding the temporal dynamics and development of CO_2 drivers by employing machine learning models, as opposed to the predominantly linear models used previously.

This research contributes to a detailed understanding of the impact of various management practices on the carbon budgets of mid-latitude grasslands. The use of state-of-the-art machine learning models provides an extended understanding of non-linear and dynamic relationships between CO_2 flux and its drivers. The study is methodologically sound, with comprehensive descriptions of the eddy covariance setup, the data processing, and statistical analyses. However, some concepts require further clarification. The results are well discussed, figures are well-described and visually comprehensible.

Dear Referee,
Thank you for your positive feedback! We appreciate the opportunity to improve the manuscript based on your thorough review and constructive comments. We have addressed your questions and suggestions in the responses below.

Specific comments:

Methods & Results:
1. Line 147: the soil variables are averaged across depths - what is the reasoning behind that? Especially for the soil water content measurements were taken in very different depths, and soil water fulfill different functions in the top soil layer than in the deep soil layer, depending on how available it is to microbes or plant roots-> why not use separately and see which layer is important in which times to gain more process understanding?

Thank you for the comment. The exact rooting depths of this grassland are unknown, as well as the exact depths of plant water uptake. Rooting and water uptake depths could also differ due to heterogeneity within the field. Furthermore, it is uncertain which soil depths are responsible for ecosystem respiration. With the EC measurements, flux data represent the integrated condition over the entire field. Therefore, we averaged the soil variables across depths to represent the integrated soil conditions over the entire profile, instead of focusing on specific depths. Nevertheless, we will adjust this sentence as "For soil temperature… averages across all depths were calculate and used in the final analysis to represent the overall soil conditions over the entire profile."

2. Line 114-115: DailyN is calculated as a daily average across the regrowth period. However, not the same amount of N would be available to the plants each day, maybe in the beginning more and later less, depending also on environmental factors such as strong rain events and leaching – please elaborate in how far this has implications on your results.

Thank you for the comment. As we responded to Referee 2, how to best represent N fertilization and "available N" in any model is indeed an ongoing discussion, also in our group. We agree

that the current value can be discussed, but currently we consider this as our best approach. During the majority of regrowth periods, there was only one fertilization event, and normally it happened relatively soon after the mowing. If there was a second fertilization event, we summed up the amount of fertilized nitrogen for that period and calculated DailyN based on the total amount. This way, we consider the supplied N as fully available for growth. To be extremely accurate, one would need to model the available N as suggested, e.g., with an exponential decline, considering many factors like soil properties, precipitation events and soil moisture dynamics, but also microbial activity, potentially delayed responses by soil moisture and soil temperature. Achieving an accurate estimate of this on a daily scale would either need another exhaustive study (with unknown methodology if data were to be collected daily in a real-world setting, not in a lysimeter) or would be based on many untested assumptions. Thus, we cannot solve this issue by changing the regrowth period average to an exponential decay or other dynamic functions without introducing large untested (maybe even untestable) bias. Furthermore, we did not find convincing alternative solutions based on the existing studies in the literature, which are currently still limited. Therefore, we consider our current way of representing real management as best as we can do, but would of course welcome more systematic approaches in future work.

3. Line 140: The gaps in flux data were filled with a random forest model. Might these confound the driver analysis that is conducted later? E.g. if gap-filling is based on other environmental variables such as Tair or VPD, of course the driver analysis shows high dependence of ecosystem fluxes on these variables. Please clarify, and if there might be confounding effects elaborate on that in the discussion.

Indeed, the half-hourly NEE data were gap-filled using a random forest model that takes into account a large number of factors, including timestamp (e.g., year, month, date, hour, DOY), meteorological variables that are used in traditional gap-filling marginal distribution sampling (MDS) methods (i.e., SW_IN, TA, VPD), and management events within the footprint (Hörtnagl, 2025), thus many more than were relevant in the driver analyses. We then partitioned the gap-filled NEE dataset using the nighttime partitioning method into GPP and Reco at half-hourly resolution. For the subsequent XGBoost models and SHAP analysis, we used daily aggregated values (mean or sum, depending on the variable). While we acknowledge the fact that certain variables are used in both processes (i.e., gap-filling and XGBoost models), the confounding effect is expected to be very small due to the following reasons:

(1) Beyond the meteorological variables, the random forest gap-filling also relies on temporal and management information. The gap-filling can capture short-term variability of the NEE, while the driver analysis focuses on overall driver importance based on broader temporal scales (20 years) for GPP and Reco.
(2) The driver analysis (XGBoost + SHAP) conducted based on daily aggregated data, therefore reducing potential correlation and confounding effects that might occur at finer time scales.
(3) In the driver analysis, PPFD_IN were used instead of SW_IN, and additional meteorological variables were used (e.g., TS, SWC). The variables used in the gap-filling process were found to have relatively low importance (current Figure A3b-d).

In addition, we cannot avoid gap-filling NEE: If we would use only measured, good quality data for the daily averages, we would overestimate the net CO_2 uptake as there are less data during the night when respiration is the dominant signal. We will add this information in Section 4.2 in the revised version of the manuscript.

4. Figure 4:
For the partitioning of GPP and Reco the nighttime partitioning method by Reichstein et al (2005) was used. In this method, this means, Reco is modelled depending on temperature. GPP is calculated as the difference between measured NEE and modelled Reco. Of course there will be dependencies then of the component fluxes to temperature and GPP, and their contribution in explaining the fluxes might be partly conceptually introduced. Does it "even out" since you integrate over regrowth periods?

I would suggest to quickly test another partitioning method (e.g. hybrid partitioning method introduced by Nguyen et al (2025), code available on github) to confirm if the main drivers are the same or if the results are confounded the partitioning method.
Nguyen, N. B., Migliavacca, M., Bassiouni, M., Baldocchi, D. D., Gherardi, L. A., Green, J. K., Papale, D., Reichstein, M., Cohrs, K.-H., Cescatti, A., Nguyen, T. D., Nguyen, H. H., Nguyen, Q. M., and Keenan, T. F.: Widespread underestimation of rain-induced soil carbon emissions from global drylands, Nat. Geosci., https://doi.org/10.1038/s41561-025-01754-9, 2025.

Thank you for the comment. During the preliminary exploratory data analysis, we compared different gap-filling methods, including nighttime (Reichstein et al., 2005), daytime (Lasslop et al., 2010), and the recent improved-daytime (Keenan et al., 2019) methods. In our PI dataset used in this study, we also included daytime partitioned fluxes as part of the flux product (Hörtnagl et al., 2025). On daily scale, the partitioned GPP and Reco had similar variation and dynamic/pattern, so we assume the result of our subsequent driver analysis to be similar as well.

In nighttime partitioning methods, GPP is calculated based on the difference between NEE and Reco, therefore it is less dependent on the variables used in modelling Reco. Thus, we chose to use this partitioning method for our driver analyses, also commonly used in other studies. Moreover, air temperature (TA) was used to model Reco in the nighttime partitioning methods. However, the XGBoost model for Reco used soil temperature (TS) as temperature-related driver, not TA. In the XGBoost model for GPP, TS turned out to be a better driver than TA in terms of importance (current Figure A3b), which also showed that the dependency of this component flux is stronger on TS than on TA.

The recent study by Nguyen et al. (2025) provides new insights, especially on machine-learning based partitioning methods. Our manuscript was submitted before the publication of this new method, therefore we did not include this method in the exploratory data analysis. We understand that the proposed partitioning method has a focus on global dryland, which is not the same ecozone/climate zone as this temperate grassland, and thus different processes are relevant in such semi-arid systems.

Technical corrections:

Line 95: how many times per year/ in which times fertilizer is applied? Please specify. One fertilizer application per regrowth period?

Yes, there was usually one fertilizer application per regrowth period (Fig. A2). The number of fertilizer applications differed among years, normally depending on the number of mowing events. We will update the sentence as "… (mainly as slurry, normally applied once after each mowing event…".

Line 128: "For more detailed information on instrumentation, see …"

Thank you for the suggestion. We will modify the sentence accordingly.

Line 205: mean annual temperature and precipitation is a repetition from lines 91-92, can be removed here or just put in brackets, e.g. "mild temperate climate (9.9 °C, 1147 mm)"

Thank you for the suggestion. We will shorten the sentence into "… mild temperate climate (MAT 9.9 °C, MAP 1147 mm)."

Line 354-356: the two parts of the sentences are not necessarily connected, I would suggest to split it up into two sentences. First part ("Often, studies focused on natural ecosystems or 355 forests at regional and global scales (Anav et al., 2015; Cai and Prentice, 2020; Davi et al., 2006; Norby et al., 2010)") fits better in the Introduction.

Thanks for the suggestion, but the info of the first sentence is indeed relevant for the second. To improve understanding, we will change the first sentence as "Often, studies focused on natural unmanaged ecosystems or forests…" to show that we want to emphasize the different ecosystems that have been studied.

Line 380-383: please improve for more clear language: "Such observations, albeit rare, are impactful in terms of C dynamics, and underscore the necessity to include such these infrequent destructive management practices in long-term flux studies and as well as modelling frameworks."

Thank you for the feedback. We will incorporate this change as suggested.

Line 392-393: sentence is unnecessarily long and not so well understandable, please split it up into two shorter sentences

We will change this sentence into two shorter sentences.

Line 428: what do you mean with "explanatory outputs which enhanced interpretability"? It sounds a bit generic, please specify

Here we meant the explainable machine learning tool like SHAP to improve the interpretability of the tree-based models. We will modify the sentence as "… but also provided explanatory outputs such as driver importance on a daily basis, which enhanced interpretability of the machine learning models and further improved…"

Line 436: in line 431 you state that during extreme events "SWC strongly reduced GPP", but later you state that "more significant negative effects of droughts on grassland CO_2 fluxes compared to those of heatwaves". It seems a bit contradictory, maybe you can clarify this section.

We meant to compare the effect of SWC and VPD. We will clarify this sentence as "… more significant negative effects of soil droughts (i.e., low SWC) on grassland CO_2 fluxes compared to those of heatwaves (i.e., high VPD) …"

Line 457-460: complicated sentence structure, please simplify or split into two sentences

We will split it into two sentences.

Line 462: please mention briefly what precision-farming is (and please mind consistent terms – in L.477 precision farming is written without "-")

Thank you for the reminder. We will add "… and spatially adjusted management practices such as precision-farming …".

References

Hörtnagl, L.: diive v0.87.1, https://doi.org/10.5281/zenodo.15648669, 2025.

Hörtnagl, L., Feigenwinter, I., Wang, Y., Buchmann, N., Merbold, L., Zeeman, M., Fuchs, K., and Eugster, W.: Eddy covariance ecosystem fluxes, meteorological data and detailed management information for the intensively managed grassland site Chamau in Switzerland, collected between 2005 and 2024: CH-CHA FP2025.3 (2005-2024), https://doi.org/10.3929/ethz-b-000747025, 2025.

Keenan, T. F., Migliavacca, M., Papale, D., Baldocchi, D., Reichstein, M., Torn, M., and Wutzler, T.: Widespread inhibition of daytime ecosystem respiration, Nat. Ecol. Evol., 3, 407–415, https://doi.org/10.1038/s41559-019-0809-2, 2019.

Lasslop, G., Reichstein, M., Papale, D., Richardson, A. D., Arneth, A., Barr, A., Stoy, P., and Wohlfahrt, G.: Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation, Global Change Biol., 16, 187–208, https://doi.org/10.1111/j.1365-2486.2009.02041.x, 2010.

Nguyen, N. B., Migliavacca, M., Bassiouni, M., Baldocchi, D. D., Gherardi, L. A., Green, J. K., Papale, D., Reichstein, M., Cohrs, K.-H., Cescatti, A., Nguyen, T. D., Nguyen, H. H., Nguyen, Q. M., and Keenan, T. F.: Widespread underestimation of rain-induced soil carbon emissions from global drylands, Nat. Geosci., 18, 869–876, https://doi.org/10.1038/s41561-025-01754-9, 2025.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grunwald, T., Havrankova, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.-M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., and Valentini, R.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, Global Change Biol., 11, 1424–1439, https://doi.org/10.1111/j.1365-2486.2005.001002.x, 2005.