

Discussion of “What matters when? Temporal development of drivers and sources of nitrous oxide emissions in winter wheat”

Author (Turco et al.) response to Referee 1 and Referee 2 comments

Dear Editor and Reviewers,

We thank the referees for their constructive and detailed comments. We have revised the manuscript accordingly and provide below a combined point-by-point response. Reviewer comments are shown in blue italics; author responses are shown in normal font. Section, line numbers, table, and figure numbers refer to the revised manuscript unless stated otherwise.

Overview of main revisions

- Results and Discussion were shortened, and the narrative was sharpened to emphasize the temporal interplay among N₂O drivers rather than the general identification of already known controls.
- The site- and year-specific nature of the study is now stated more clearly, and emission factors and GHG budgets are presented as site-specific estimates.
- The former background-flux calculation based on a 30-day post-fertilization exclusion window was removed. The manuscript now reports an apparent emission factor and explains why this is not an IPCC Tier 1 EF.
- RF gap-filling performance metrics are now reported for N₂O and CH₄ fluxes; information on gap-length is summarized in Sect. 2.2.3 and Table A2; u* filtering sensitivity results were added in Table A3.
- The XGBoost workflow was revised to use a chronological hold-out evaluation, with feature selection and hyperparameter tuning restricted to the training period.
- The interpretation of GPP was revised: GPP is now transparently described as a proxy for crop activity and N demand, not as an independent causal driver separated from soil mineral N availability.
- The previous binned WFPS-GPP analysis was replaced by a SHAP dependence plot for GPP (Fig. 5), with additional SHAP dependence plots in Fig. A4.
- The isotope methods and interpretation were clarified by specifying the March 2023 sampling window, reporting the reduced fraction as (1-r), adding the alternative FRAME scenario in Fig. A6, and discussing uncertainty at low fluxes.
- The former dedicated mitigation subsection was removed; relevant contextual statements were integrated into the other Discussion subsections with more cautious wording.
- **Additional changes:** We also revised the calculation of winter wheat N export from harvest. In the original manuscript, N export was derived from aboveground biomass sampling, whereas in the revised manuscript we calculated N export from the actual grain and straw harvest data. We consider this approach more appropriate because it directly reflects the harvested N removal from the field. This correction changed the calculated wheat NUE.

Response to Referee 1

General Comments:

This manuscript is a thorough and insightful exploration into the drivers of soil N₂O emissions. Building on a well-researched field, the paper employs strong and novel methods to demonstrate this idea through use of spatio-temporally integrated flux measurements and frequent measurement of crop growth dynamics to estimate soil nitrogen dynamics, plant nitrogen uptake, and N₂O emissions over time. The experiment appears well conducted and to have yielded a wealth of data, to which the authors add investigative statistical methods which contribute a strong

picture of the drivers of N₂O and how these drivers shift over time, which I found particularly insightful. The manuscript is also clearly written, if overly descriptive and verbose at times.

We thank the referee for the positive assessment and constructive suggestions. We shortened and reorganized several sections of the manuscript, especially the Results and Discussion, to improve focus and readability.

However, I have key concerns regarding the methodological rigor and the resulting interpretations in three key areas. First, there is significant opacity regarding the data handling process during machine learning, including a high risk of temporal data leakage during model validation. Second, the methods used to classify 'background' vs. 'hot-moment' emissions are tenuous and lack a mechanistic basis. Third, while the authors identify Gross Primary Productivity (GPP) as a preeminent negative driver of emissions in their discussion of results, the exclusion of soil mineral N concentrations from driver analysis creates a risk of omitted variable bias, in which GPP may act as a proxy for N exhaustion. To justify their major claim that GPP is a suppressor of N₂O emissions and that N synchronicity is a path toward N₂O mitigation, the authors should make efforts to decouple GPP from simple substrate limitation in their statistical investigation of drivers.

We have addressed these points in detail below.

With a tighter text and these improvements to statistical methods, I am confident this manuscript will represent a strong and novel contribution to understanding pathways to agricultural GHG mitigation.

Thanks. We have shortened Results and Discussion and sharpened the narrative throughout.

Major Comments:

L55-56: Is this sentence referring to nitrogen cycling or crop growth?

Thanks for pointing out this ambiguity. It did refer to N₂O emissions. This has been clarified in the revised manuscript (lines 58–59).

L80-84: This is true, strictly speaking. However, it is also possible to infer the dominant N-transforming process based on the rate of N₂O production, as hot moments have been shown to be overwhelmingly the result of denitrification. Relatedly, a recent work used ML to differentially model nitrification-dominant and denitrification-dominant emissions (Lussich et al., 2026) <https://doi.org/10.1002/jeq2.70126>

We thank the reviewer for this suggestion. While it is generally accepted that high N₂O fluxes often reflect denitrification-dominated conditions, this inference does not necessarily hold across all sites or years. For this reason, we consider stable isotope approaches essential for robust source attribution and have therefore retained this methodological distinction. We consider machine learning approaches complemented by stable isotope approaches very powerful, also for future studies. In this regard, we recognize the value of innovative ML approaches, so we have added a sentence and the reference to the revised manuscript (lines 84–86).

L84-90: I'll use this passage to illustrate a broader observation that characterized much of this manuscript. The authors frequently provide detailed explanations, such as the principles and mechanics behind stable isotope analysis. While thorough, these explanations are often characterized by a poor economy of words, include details which are not directly pertinent to the narrative at hand, and add up to a manuscript of considerable length (I count nearly 13,000 words). This work would significantly benefit from greater concision, illustrating only the most pertinent details and taking advantage of the expected level of familiarity of a biogeoscientist audience.

We have shortened the text throughout the manuscript, improving logical flow while providing enough details so that biogeochemists not familiar with highly specialized topics, like SP, still understand our approach and results.

Table 1: slurry was applied the day after mineral N fertilizer was applied. Is this typical practice in the region? This creates perfect C and N chemistry to produce N₂O. A study focused on understanding N₂O mitigation opportunities, this is not an efficient fertility management decision. Can slurry be applied in fall, before wheat planting?

We clarified that this practice can indeed occur in integrated Swiss crop–livestock farms and is compliant with Swiss agricultural regulations, while also acknowledging that the near-simultaneous application of mineral N and slurry likely created favorable conditions for denitrification and N₂O production (Sect. 4.3, lines 689–691; Sect. 4.4, lines 741–744).

L162-163: Why were two separate outlier detection methods used to filter erroneous fluxes?

We clarified that different filters were applied sequentially because they capture different types of anomalies, and not all filters were required for all fluxes (lines 171–173).

L173-184: Regarding the RF gap-filling: was there any analysis of error propagation? Or an assessment or justification of the accuracy of RF as a gap-filling method? Accurate gap filling N₂O emissions is still a topic of much research, and has resulted in mixed success. One of the papers cited here reports R² of gap-filling for N₂O between 0.6 and 0.76, using only 15% of ‘missing values’ as the test set, in comparison to ~47% of the time period missing in this study. Other work has reported lower R² values from 0.42 (Taki et al., 2019, 10.1139/cjss-2018-0041) to 0.66 (Goodrich et al., 2021, 10.1016/j.agrformet.2020.108280), also using just 15% of the data as ‘missing values’ to be filled. The other paper the authors cited here performed no analysis on the accuracy of the RF gap-filling method used. The success of ML gap-filling has also been shown to be related to the length of gaps (Taki et al., 2019), yet there is no discussion of typical gap length or the impact of lengths beyond reporting the aggregate percentage of missing values. While I acknowledge that flux gap-filling is a major challenge which has yet to be solved, and that work must go on in the meantime, nevertheless I feel it is important to acknowledge the limitations of gap-filling high resolution N₂O flux data and the effect that these limitations might have on this study.

We fully agree that the flux community has not yet settled on one method to gap-fill non-CO₂ greenhouse gas fluxes. We have now added quantitative RF gap-filling performance metrics in Sect. 2.2.4 (R², MAE, and RMSE for N₂O and CH₄ RF gap filling; lines 194–195) and statistics on data-gap structure are given in Appendix Table A2 and referred to in Sect. 2.2.3 (lines 177–180). We acknowledge that uncertainty as a function of gap length was not assessed but note that most gaps were short. While we did not conduct a full analysis of error propagation, we now added a sensitivity analysis of cumulative fluxes to the u* filtering threshold, reported in Appendix Table A3 (the approach is introduced in Sect. 2.3 and the results are given in the text of Sect. 3.3).

L183-185: It seems here like the authors estimated the background flux level by just excluding 30-day post-fertilization flux data with assumption that fertilization effect lasts only for 30 days. First of all, I don't think this assumption is correct and your own data did not support the assumption. Fertilization can have long-lived impacts on N₂O emissions, particularly during dry periods, offseason, etc. As well, not all hot moments are fertilizer-driven, as those like Claudia Wagner-Riddle's group have shown. Moreover, not all fertilizer-driven N₂O emissions are caused directly by fertilizer at all: the excess N added to the soil system by fertilization may be temporarily captured by plant or microbial biomass and mineralized months or even years later, thus driving emissions that would not occur in a natural, unfertilized system but which nevertheless occur distantly from any fertilization event. The authors' data shows this: the large peaks, as big as post-fertilization, from mid-June to August is well beyond 30-day post-fertilization period and unlikely to happen in a truly unfertilized control treatment. Moreover, the SHAP analysis (Fig 4) also shows that fertilization effect was a dominant factor for almost two-months after fertilization. I have serious reservation about this method of distinguishing between HM and BG emissions by simply excluding fluxes within 30 days of fertilization. This potentially inflated the background emissions and perhaps underestimated emission factors. Authors might consider alternative methods of contextually distinguishing background emissions from hot moment emissions based on outlier detection. Please see Ackett et al., 2025; doi.org/10.1029/2025JG008953.

We agree that the previous use of “background emissions” was confusing. We also agree that we cannot reliably separate fertilization-induced from non-fertilization-induced emissions in the absence of an unfertilized control treatment. We therefore removed the background-flux estimation and no longer calculate a background-corrected emission factor. Sect. 2.3 now defines an apparent emission factor as cumulative N₂O-N emissions divided by fertilizer N input (lines 222–229). Sect. 3.3 reports a winter wheat apparent EF of 4.0% and an annual apparent EF of 5.1%, explicitly noting that non-fertilization-induced emissions could not be subtracted (lines 488–489).

L199-201: Spatially aggregated flux measurements are captured across a field using EC, yet a single point measurement is used for soil moisture content and temperature? Soil moisture is also highly spatially heterogeneous. This seems a potentially noteworthy limitation.

We agree and now explicitly frame soil moisture measurements as a temporal driver rather than footprint-mean spatial estimates. Sect. 2.4 now acknowledges the limitation of a single soil profile and discusses spatial representativeness (lines 242–243). It also notes that an independent nearby sensor showed similar temporal SWC dynamics (line 244). We also state that WFPS conversion used bulk density values from earlier measurements (Emmel et al., 2018), leading mainly to uncertainty in absolute WFPS values (lines 245–247).

L303-305: Again, the persistence of elevated N availability is highly variable and can be much longer than 30 days.

We agree. We clarified this in Sect. 2.7.1 and now the sentence reads:

“To represent recent fertilization inputs, we used the cumulative fertilizer N applied over the preceding 30 days (total N; kg N ha⁻¹). This window was chosen to capture the short-term post-application period commonly considered in event-based reporting, not accounting for longer-lived or legacy fertilization effects (Bouwman, 1996; Cowan et al., 2020; Qian et al., 2025).”

L318-325: A more complete description is needed about the data splitting process is needed in order to verify its validity. In this custom time-block method, how large were the time chunks? Was the model trained on data further in the future than the test data? When working with a single time series, the most correct method of cross validation is to use a method similar to that employed by the TimeSeriesSplit function in scikit-learn. By this method, the timeseries is split into n+1 chunks, and the model is trained on all prior data within the timeseries, including an initial runup chunk (i.e. an expanding window). This ensures that the model is never trained on future data, which would constitute a form of data leakage.

AND

L328-331: I have searched through the document and could not find a description of what the authors here call the “final model,” nor what the “test set” was. There seem to be a lot of mixing of terminologies regarding data handling and model evaluation, leading to a very opaque picture of the actual methods used. Starting at the beginning of this section:

“Following variable selection, model hyperparameters were optimized using 10-fold cross-validation. To account for temporal autocorrelation and avoid overfitting, while also providing representative coverage of the measurement period, we employed a custom time-block strategy. This approach involved an 80/20 split between training and validation...”

This seems plainly contradictory. If a 10-fold cross validation were used, then across each fold 90% of the data would be used for training and 10% for validation, yet the authors claim an 80/20 split. My best guess might be that the 10-fold cross validation might be a separate process exclusively for hyperparameter tuning, for which no details were provided, and then an abrupt shift to a new data handling process involving an 80/20 split of some kind takes place? Yet descriptions of each process are incomplete and poorly differentiated in their purposes, leading to a jumbled passage.

“with the validation set comprising randomly selected, non-overlapping time blocks that together represented 20% of the available data.”

The authors here claim that time blocks were randomly selected, but a timeseries should not be randomly split. The authors seem to acknowledge this idea with their reference to temporal autocorrelation, but their description of their methods here does not give confidence that they have properly dealt with this challenge.

“This splitting strategy was consistently applied throughout the modeling workflow, including during cross-validation and final model evaluation.”

This does not make sense to me. There have been no clear definitions of what constitutes the “cross-validation” and the “final model evaluation.”

“Cross-validation results showed a R² of 0.60 and a RMSE of 1.1 nmol N₂O m⁻² s⁻¹ on the validation set, while the

training set showed a R^2 of 0.98 and a RMSE of $0.29 \text{ nmol N}_2\text{O m}^{-2} \text{ s}^{-1}$. The final model, trained with early stopping (10 rounds) to prevent overfitting, achieved a R^2 of 0.70 and a RMSE of $1.14 \text{ nmol N}_2\text{O m}^{-2} \text{ s}^{-1}$ on the test set (Fig. A2)."

This passage invites more confusion. Beyond the unclarity regarding CV vs final evaluation, the authors describe model performance on "the test set," "the validation set," etc. Was there a singular test set, as the language here suggests, or are the authors instead referring to the average model performance on holdout data across the 5 or 10 folds?

To avoid temporal data leakage, we revised the XGBoost modelling workflow. The final 20% of the time series is now reserved as an independent chronological hold-out set, while feature selection and hyperparameter optimization are conducted only on the first 80% of the data using 5-fold cross-validation, described in Sect. 2.7.1 (lines 375–382), and model performance is shown in Figure A2. We now report the optimized hyperparameters in Table A5. For the SHAP analysis, the model is trained on the full dataset because the aim is interpretation of the observed temporal patterns rather than prediction of unseen data, as now described in Sect. 2.7.2 (lines 389–390).

L333-349: In contrast to the preceding passage, this description of SHAP values is exhaustive, and may be excessive.

We shortened the SHAP description while retaining the essential information needed to interpret SHAP values and normalized relative driver contributions (lines 384–395).

L360-384: This passage is a lengthy textual description of figure 2, which I'm not sure is necessary given the length of the manuscript.

We shortened the Results description of environmental and vegetation dynamics (lines 407–418).

L452-453: The average background flux estimate of $0.44 \text{ nmol m}^{-2} \text{ s}^{-1}$ is equivalent to $\sim 11 \text{ g N}_2\text{O-N/ha/d}$, which is substantially (3-4 times) larger than globally estimated background emissions of ~ 2.5 to $3 \text{ g N}_2\text{O-N/ha/d}$. This could be due to the reasons I explained above.

As explained above, we have removed the background-flux calculation from the revised manuscript.

L462-466: "In contrast, the cover crop acted as a net CO_2 source, emitting $108 \text{ g CO}_2\text{-C m}^{-2}$ "

The cover crop phase was only measured for 2 months, which was influenced by cultivation after wheat harvest. Is this the typical cover crop growth duration? If not, then this is misleading information that provides a snapshot of the rotation. Cover crop active growth phase is often known to sequester C (NEE sink) and negligible N_2O emissions. As the authors correctly point out in their discussion, it is not the cover crop per se that is acting as an emitter of CO_2 , but rather this is the response to harvest of the winter wheat and the bare soil. It is tricky, but I wonder if there is a better way to convey that the cover crop is actually reducing emissions, not magnifying them, even though these emissions are taking place during the cover crop's establishment.

We agree that our text was not clear. We revised the text and clarified now in the Discussion (Sect. 4.3) that the observed net CO_2 release reflects NEE during a short, disturbed post-harvest establishment phase, not an intrinsic property of cover crops (lines 721–722). We also added context indicating that cover cropping likely reduced post-harvest emissions compared with a bare-soil situation at the same site (Sect. 4.3, lines 723–725).

L480-483: in L452, the authors mentioned background flux of $0.44 \text{ nmol m}^{-2} \text{ s}^{-1}$. This is confusing as it seems different subset of data are being used in places to estimate background flux.

We revised Sect. 2.7.2 (lines 391–393) and Sect. 3.4.1 (lines 520–521) to better clarify what the SHAP baseline represents.

L486-509: SHAP values are best interpreted by relating the direction/magnitude of impact alongside the magnitude of the factor value. For example, the authors write: " N_2O emissions were mainly suppressed by WFPS." The fact that soil moisture had a large negative impact on flux predictions, as described here, leaves a lot of room for interpretation. Were these negative influences related to high or low soil moisture?

We added SHAP dependence plots: Fig. A4 shows dependence plots for drivers other than GPP, while Fig. 5 shows the GPP dependence plot with a LOWESS smoother. We also revised the text to link the sign of SHAP values to the corresponding driver ranges (Sect. 3.4.1, lines 531–533).

L510-515, Figure 5: The binning strategy used here to quantify the relationships between WFPS, GPP, and flux, is statistically fragile. The bins create artificial step functions in what is likely a linear relationship, and also reduces the number of data points for statistical analysis. It also is able to demonstrate an interaction exists, but does not quantify that relationship in terms of strength or behavior. A better approach might be to model a multiple linear regression with an interaction term. Partial r^2 could also be used to more precisely determine the relative explanatory power of each factor and the interaction.

We agree that the binning strategy was statistically fragile. We considered the suggested multiple linear regression with an interaction term, but the data did not meet the assumptions required for linear regression, so we instead replaced the binned analysis (former Figure 5) with a continuous SHAP dependence plot for GPP including a LOWESS smoother (new Fig. 5). This avoids artificial binning and retains all observations. We have also removed all statements about the interaction effect between GPP and WFPS, as this interaction was based on the binned analysis and is no longer supported by our revised approach.

L559-560: True, but could this also be due to increasing contribution of nitrification to the net N_2O emissions? How could you be so sure about Fig 7B secondary vertical axis?

The secondary axis in Fig. 7B represents the reduced fraction ($1 - r$) derived from the FRAME model based on a closed Rayleigh formulation ($M = M_0 + E \cdot \log(r)$), where r is the residual N_2O fraction remaining after reduction. Thus, $(1 - r)$ quantifies the fraction of N_2O that has been reduced to N_2 . We have clarified this explicitly in both the main text (lines 297–299) and the figure caption to avoid ambiguity.

Regarding the isotopic pattern: while an increased contribution of nitrification could indeed explain higher site preference (SP) values, it would not account for the pronounced enrichment in $\delta^{18}O$ observed. Therefore, the simultaneous rise in both $\delta^{18}O$ and SP, particularly given the magnitude of the $\delta^{18}O$ shift, is more consistent with enhanced N_2O reduction rather than a shift toward greater nitrification alone. This interpretation is further supported by the FRAME model results, which indicate that the relative contributions of the different production pathways remained stable during the last three sampling dates (Fig. A6).

L565-566: Relatedly, could not the increased role of nitrifier denitrification also indicate a decreased rate of denitrifier denitrification overall?

As clarified above, $(1 - r)$ represents the fraction of produced N_2O reduced to N_2 before emission. We revised the wording to avoid ambiguity, replacing “more complete denitrification” with “a larger fraction of N_2O reduced to N_2 ” (line 606). In general, nitrifier and bacterial denitrification are very difficult to distinguish, as their source signatures overlap (Figure 7a in the revised manuscript). We now clarified this in lines 598–600, reading: “However, following Buchen et al. (2017), we pooled heterotrophic bacterial denitrification with nitrifier denitrification, and nitrification with fungal denitrification, due to the large uncertainty in separating these processes given their similar endmembers. All four processes are nonetheless shown separately in Fig. A6.”

L646-650: The N_2O offset effect on net CO_2 is based on NEE and not actually on long-term C sink, which is soil C stabilization. Not all of the NEE sink from a growing season is translated into soil C gain that is stabilized for longer period beyond the growing season. The studies using long-term experiments showed much higher N_2O offset effects in fertilized systems in other regions (see Dhaliwal et al., 2025; doi.org/10.1002/jeq2.70046). Therefore, the short-term nature of the study may have underestimated N_2O effects.

We agree and now distinguish EC-based net GHG exchange from the net GHG budget including lateral C fluxes. Sect. 2.3 defines net GHG exchange and net GHG budget separately (lines 209–211). In Sect. 3.3 and Sect. 4.3, now both quantities are discussed, and we included lateral C fluxes as well as the net GHG budget in Table 2.

L656-661: Mineral N sampling also showed modestly high N availability during this period, likely related to low uptake and mineralization from residue and microbes

We incorporated this point into the interpretation of post-harvest N₂O emissions (line 715).

Section 4.4, L705-708: I will take this line as an opportunity for a discussion on the idea of synchronicity between N supply and demand as a key feature in determining N₂O emissions, and as a key idea in this manuscript. The authors place much significance on this idea throughout their introduction, use of methods, and the discussion here. In my view, the evidence for nitrogen synchronicity as the preeminent determinant of N₂O emissions, rather than simply N availability, is a bit scattered. For example, Figure 4 excellently illustrates flux and GPP, yet mineral N availability is missing as the third key component needed to make this case. From Fig 4 there certainly does seem to be an inverse relationship between flux and GPP, but I wonder the extent to which GPP is also negatively correlated with N availability. Given the inverse relationship between GPP and mineral N (the nitrogen is moving directly from the soil and into the plant biomass), more evidence is needed to prove that GPP is independently negatively affecting N₂O emissions, beyond soil N.

This leads me to my concern about the authors' interpretation of GPP's role in the SHAP analysis. Because soil mineral N concentrations were not included as input features in the Random Forest model, the model likely suffers from omitted variable bias. While 'days since fertilization' is included as a predictor, it is a very rough and linear proxy that fails to capture the dynamic, non-linear depletion of the soil N pool. In the absence of a direct N-substrate variable, the machine learning algorithm quite possibly utilizes GPP as a better mathematical proxy for the diminishing N pool. Consequently, the high importance attributed to GPP in the SHAP analysis may simply reflect N exhaustion rather than a mechanistic suppression of N₂O production by plant uptake (synchronicity). To truly substantiate the claim that synchronicity is key to emissions, the authors would ideally show that for a given level of soil mineral N, higher GPP results in lower emissions. Without including measured soil N data in the modeling workflow, the current results show correlation with the crop's growth phase, but do not sufficiently decouple plant demand from simple substrate limitation to make strong claims such as this one.

As noted above, we revised all relevant interpretations to state that GPP is a proxy for crop activity and plant N demand, and that the observed relationship likely reflects substrate limitation rather than direct causality. The role of GPP is now extensively discussed in Sect. 4.2 (lines 670–680).

Response to Referee 2

General and specific comments

The study reports GHG fluxes by ECT in winter wheat following 2-3 year grass clover cropping and measurement of vegetation dynamics, soil moisture, and mineral N to (i) quantify GHG fluxes, (ii) identify drivers of N₂O fluxes (iii) and determine N₂O pathways.

A novel machine learning (ML)-based modeling approach was used to identify drivers for N₂O peak events and to model N₂O fluxes.

The ECT approach was state of the art and together with the soil and vegetation parameters determined the data-set is probably unique and well suited for model testing. Isotopic fluxes were determined during 3 weeks after early spring fertilization with mineral and organic fertilizer to determine N₂O pathways and reduction to N₂.

While the evaluation of ECT data and the measured control factors using the KI model is certainly innovative and helpful to identify drivers of N₂O in that study.

Thanks for this positive feedback.

But the study presents only data one site and one under soil conditions and fertilization regimes which are quite special (one-year, one soil, without replicates, special case of grass-clover as preceding crop with very high expected N-mineralisation and mixed organic/mineral fertilization, in a clay loam soil which is quite prone to intense denitrification), the study thus does not expand our knowledge on the processes and their control. It only confirms

with several observations the well-known regulation of N₂O fluxes by the interplay of fertilization, soil moisture, mineral N and crop growth. Similarly, the determination of emission factors and of the GHG budget must be seen as individual observation of that site and year that add little to regional or global estimates of these values.

We agree that our study represents only one site and one winter wheat season under specific soil and management conditions, and we have made this limitation clearer throughout the revised manuscript (e.g., lines 749–752). However, the combination of methods and thus data streams used, i.e., integrating fluxes, stable isotopes, soil, plant and management information, is still rare. Moreover, we also address the question of how the relative importance of different N₂O flux drivers changes over the course of a cropping season, an aspect often not considered in other studies. Nevertheless, we have now removed all claims that our emission factors and GHG budgets are directly transferable to regional or global inventories. At the same time, we have sharpened the manuscript's main contribution: it does **not** aim to identify well-known N₂O controls in isolation, but to resolve how their relative contributions change over time by combining high-resolution EC fluxes and explainable ML (e.g., lines 622–630). We revised the Abstract, Discussion, and Conclusions accordingly.

The new ML-model was used to identify drivers of N₂O fluxes. Traditionally, this is done by conventional statistical models or using process-based biogeochemical model like, e.g., DNDC. There were no attempts to compare the new model to traditional approaches. Moreover, the model is not addressed in the discussion. From my view the data-set and the evaluation reported in the paper are suitable for a methodical paper that focusses on the new model and how it relates to previous models. But the discussion largely neglects the new model and focusses on the question (1.) how the fluxes were regulated, (2) how the magnitude of N₂O fluxes relates to previous studies and (3.) what we can learn from the data on N₂O mitigation strategies. While the first aspect, as said before, does not add any new insights, the study does not add any basis for the third aspect.

The paper is thus not suitable for publication in the present form. I suggest major review which should include model comparison as explained above and removal of the current discussion on emission factors, flux regulation and N₂O mitigation.

We appreciate this suggestion, but such a comparison of ML with other statistical or process-based models would require data from multiple sites/years and was therefore beyond the scope of this study. As pointed out above, our dataset is very comprehensive and targeted towards process interactions, not a methodological comparison. In a new Discussion subsection dedicated to the role, strengths, and limitations of the implemented explainable ML framework (Sect. 4.1, lines 622–639), we clarified that XGBoost–SHAP is used as an interpretative, explanatory tool rather than as a new model to be benchmarked. We also removed the stand-alone mitigation subsection and integrated the relevant mitigation-related statements into Sects. 4.2 and 4.4.

There are some further shortcomings in the paper regarding the GHG budget calculation, the determination of emission factors, the calculation and interpretation of the isotopic data, the presentation and discussion of the N budget and of NUE. These points are addressed in the detailed comments below.

Details:

Line 45f pH as control of denitrification product ratio missing

Role of N₂ related to total N loss, yield and GHG cost for replacing loss of fertilizer N and role of N₂O reduction to N₂ to understand N₂O regulation is missing in introduction and also in the evaluation of the N₂O drivers

We have revised the Introduction to clarify that environmental controls of nitrification and denitrification also affect the reduction of N₂O to N₂, the final product of denitrification (lines 47–49). While pH is an important control of N₂O reduction, we avoid presenting it as the sole driver because other factors are relevant for this step as well. We also agree that N₂ losses are relevant for total N loss, yield, and the GHG costs of replacing fertilizer N. However, our study focuses on N₂O emissions and their drivers, and N₂ losses were not directly measured.

L90 relevance for isotopocules to determine rN₂O missing

We have revised the sentence (lines 94–95), which now reads:

“Combined with nitrogen and oxygen stable isotope ratios ($\delta^{15}\text{N}$ and $\delta^{18}\text{O}$, respectively), SP can be used to trace

the temporal development of N₂O producing processes and constrain the extent of N₂O reduction to N₂ (Verhoeven et al., 2019; Yu et al., 2020).”

L91-96 need to stress that process identification with stable isotope was only for a few weeks.

We have revised objective (iii) to clarify that isotope-based process identification was conducted during the weeks following a fertilization event, not throughout the full study period (lines 101–103).

L103 supply full soil info: pH, bulk density, total C, total N, soil classification.

We have complemented the site description with the requested soil information, including pH, bulk density, and C and N stocks, while retaining the existing soil classification and soil texture information (lines 110–113).

L110-118 give more details on tillage technique and straw management after wheat as this is a key control of N₂O

We have added further details on the tillage technique after winter wheat harvest in Table 1. Straw management was already specified (line 124).

Table 1: unit of Nin/out missing

We have revised Table 1 by moving the unit from the caption to the corresponding table heading, so that the units for Nin/out are now explicitly stated.

Interesting that NH₄NO₃ and slurry were applied simultaneously. Consider that labile C of slurry might create hot spots of denitrification from NO₃ of min. fertilizer. And that O₂ consumption during nitrification of NH₄ might also cause short-term anaerobic hot spots.

We added context that this practice can occur on integrated Swiss crop–livestock farms and noted that the combination of mineral N and labile C likely promoted denitrification (lines 740–744).

More details on slurry needed (volume Norg and NH₄ content, and “+18” is not clear to me

We added the slurry application volume to Table 1 and, to avoid confusion, give the estimated NH₄⁺ content of the applied slurry in the caption.

2.2 skipped, not my expertise

L185-190 background flux calculation not clear. I think it is not possible to get a robust estimate of N₂O fluxes with zero fertilization from the EC data of one field. Please discuss uncertainty arising from this simplification and clarify in all parts of the paper that the reported EF is an approximation.

We removed the background-flux correction entirely. Sect. 2.3 (lines 222–229) and Sect. 3.3 (lines 487–490) define and report apparent EFs and state that non-fertilization-induced emissions could not be subtracted.

L199-207 one moisture profile is not enough for one field due to variability. Not adequate to use old data for BD in top soil. Please discuss these limitations

We agree and now explicitly frame soil moisture measurements as a temporal driver rather than footprint-mean spatial estimates. Sect. 2.4 now acknowledges the limitation of a single soil profile and discusses spatial representativeness (lines 242–243). It also notes that an independent nearby sensor showed similar temporal SWC dynamics (line 244). We also state that WFPS conversion used bulk density values from earlier measurements, leading mainly to uncertainty in absolute WFPS values (lines 245–247).

L 243 need to calculate uncertainty of soil emitted isotopes ((Wu et al., 2019). SD of isotope analysis should be reported

Technical replicates were not available for each sample due to the method used, so sample-specific SDs could not be reported. We clarified the quality-control procedures used to assess instrument performance (lines 272–275).

L254 need to report results of all scenarios and uncertainty given by Frame

Sect. 3.5 now reports mean $\pm$ SD (lines 602–604) and notes that the alternative scenario did not change the interpretation (lines 604–606). Fig. A6 shows results from both scenarios.

L423 better keep consistent wording (N2O fluxes) to avoid misunderstanding

AND

L436 keep consistent wording. This sentence suggests that net ecosystem flux is different from emission, but you use both expression for the same flux. So what do you mean by “indicating”

We standardized the terminology throughout the manuscript. “Fluxes” is used for both positive and negative net exchange, while the term “emissions” is used when referring specifically to losses to the atmosphere.

Section 4.3.4: why was there continuous positive CH₄ flux? The soil was mostly rel. low in WFPS, eg clearly above 0.8 and N₂O flux was low. Thus it would be expected that the soil acts mostly as CH₄ sink. This questions validity of EC results. Or was there evidence for positive CH₄ fluxes from the chambers? An explanation might be the decay of the grass-clover sward in the clay loam leading to CH₄ production. But this would indicate strong reducing conditions which would be highly relevant for the N₂O production and reduction. So this should be discussed.

We clarified the low mean measured CH₄ flux, median signal-to-noise ratio and high uncertainty of CH₄ fluxes in Sect. 3.2.3 (lines 471–473). We discussed likely site-related explanations in Sect. 4.3 (lines 708–711), reading: “The field acted as a small net CH₄ source, consistent with results from other agricultural soils (Cowan et al., 2021). CH₄ emissions occurred mainly during wet periods, while dry periods showed low uptake, indicative of aerobic oxidation (Le Mer and Roger, 2001). The net CH₄ source likely reflected the poorly drained silty clay soil at our site, which likely promoted anaerobic conditions favoring methanogenesis.”

L450 please check units: 0.8 nM/m²/sec gives 5.6 g N/ha/282 days so your numbers seem to be mMol/m²/sec

We have rechecked the unit conversion and confirm that the reported fluxes are/were correct and are expressed in nmol N₂O m⁻² s⁻¹. The cumulative emission over the 282-day period is calculated as:

0.81 [nmol N₂O m⁻² s⁻¹] * 282 [days] * (60*60*24 [seconds per day]) * 10⁻⁹ [nmol to mol] * 28.014 [molar weight N₂] * 104 [m² in 1 ha] * 10⁻³ [g to kg]

Resulting in values expressed in kg N₂O-N ha⁻¹ in 282 days.

L278 need to explain how the root simulators are different in estimating plant available N, since N_{min} is plant available. Also explain why 0-0.1 m was chosen as rooting depth is much deeper

We clarified that PRS probes provide a time-integrated estimate of ion supply, whereas N_{min} is an instantaneous extractable inorganic N pool (lines 327–329). The 0–0.1 m depth was chosen because it is strongly affected by management and contains high root density and thus nutrient uptake activity (lines 331–332).

L 303 not clear why N fertilization was used as a proxy for N availability instead of measured N_{min} in cores or accumulators. To train models of N₂O, typically N_{min} is measured in top soil at each gas sampling event.

We clarified that N_{min} was measured only on discrete sampling dates and could not be meaningfully interpolated to the subdaily EC time scale (lines 360–362). Fertilization variables and GPP were therefore used as proxies for recent N input and crop activity/plant demand (lines 362–364).

Unit nM/s²/sec: please refer to identical units: fertilized area is kg /ha. N₂O should be reported accordingly.

We are not sure what this comment refers to in the manuscript. We use $\text{nmol m}^{-2} \text{s}^{-1}$ when referring to instantaneous or average N_2O fluxes and $\text{kg N}_2\text{O-N ha}^{-1}$ for cumulative fluxes, e.g., for comparisons with N inputs and outputs.

L326,327 not clear what these variables mean

We moved model hyperparameters to the Appendix (Table A5).

L453 – 450 Calculation of EF based on estimation of unfertilized N_2O fluxes from phases of background flux is not adequate as this is affected by seasonal effects (temp, moisture, Nmin of unfertilized from mineralization-nitrification-immobilisation. Please delete estimation or explain that a very rough approximation was used to get an EF estimate.

We removed this part, as described above.

L 466 CO_2 uptake during cropping is not a climate benefit as the fixed C is not stable unless grain and straw were long-term deposited which is quite improbable. Thus it does not make sense to be compared with GHG effect of N_2O flux. Clearly, not NEE but NEE minus net C export must be used in GHG budgets.

Please correct or delete paragraph and delete last line of Table with the GHG budget and report only the fluxes. Please double check the CH_4 fluxes. Positive flux in wheat is not plausible. Please explain reasons for net source (shallow groundwater, stagnic soil properties?)

We agree and now distinguish EC-based net GHG exchange from the net GHG budget including lateral C fluxes. Sect. 2.3 defines net GHG exchange and net GHG budget separately (lines 209–211). In Sect. 3.3 and Sect. 4.3 now both quantities are discussed, and we included lateral C fluxes as well as the net GHG budget in Table 2.

On the magnitude of CH_4 fluxes, see our response on the previous page.

L512 avoid unclear interpretations in results. PS is not a direct regulator of N_2O . It is not PS itself but the multitude of plant growth effects (exudation, rhizosphere respiration, water uptake, Nmin uptake, etc.), please address in discussion accordingly

We have revised the wording and expanded the Discussion in Sect. 4.2 (lines 670–675) to better contextualize the observed association between GPP and N_2O fluxes. Specifically, we have emphasized that GPP is used as a proxy for plant activity and N demand, and that the relationship likely reflects indirect plant-mediated effects (notably N uptake and thus change in soil N availability), rather than a direct control of N_2O emissions.

Section 3.5: Stable isotopes also yield info on N_2O reduction, thus this aspect should also be covered by the heading.

We retained a process-focused heading but explicitly describe and discuss the reduced fraction ($1 - r$), i.e., the fraction of N_2O reduced to N_2 (lines 606–607).

Please state and discuss whether and why you use the mixing-reduction or reduction-mixing scenario in FRAME and how results would change if you assumed the other scenario. Please also discuss the rising uncertainty in isotopes of soil emitted with decreasing flux. As far as I know, the FRAME model does not yet add this effect to the reported uncertainty. But the effect is addressed in detail in Wu et al 2019. $r\text{N}_2\text{O}$ of march 17 might be highly uncertain due to the low flux.

We used the mixing–reduction scenario as the main framework and tested the alternative reduction–mixing scenario. The alternative scenario produced strongly overlapping source-fraction distributions and did not alter the interpretation. We also added a discussion of increasing isotope uncertainty at low fluxes. Sect. 2.5 (lines 299–301), Sect. 3.5 (lines 604–605), and Sect. 4.4 (lines 748–749) were revised accordingly. Figure A6 now shows results from both scenarios.

Fig. 7 Not clear what is shown here residual N₂O fraction or unreduced N₂O fraction? r explained in the methods is: “The contributions of different N₂O production pathways and the residual fraction of unreduced N₂O (r) were estimated with the FRAME software”

The residual N₂O fraction is equivalent to the N₂O/(N₂+N₂O) ratio, ie lowering values show increasing N₂O reduction to N₂.

But in Fig 7ba it can be seen from the smaller circles along the reduction line that rN₂O should be lowering over tie. Hence it seems you show 1-r which corresponds to the expression in the caption “reduced fraction”.

Thanks for pointing out that this was not clear. We have now clarified that the plotted quantity is (1 – r), the reduced fraction, while r is the residual unreduced N₂O fraction (lines 298–299; 606; caption Fig. 7).

L556: to be clear better state: d180 and SP of emitted N₂O

We revised the wording accordingly (lines 591–592).

L 557 better name it “heterotrophic bacterial denitrification” since AOB are also bacteria

We adopted this terminology across the revised manuscript.

L 583 not plausible that GPP is a direct driver of N₂O but rather its relation to N uptake, exudation and water uptake

As mentioned above, we have revised the text to clarify this point. We expanded the Discussion in Sect. 4.2 (lines 670–675) to address this ambiguity.

4.1 This descriptive discussion does not yield any new insights but just summarizes that N₂O fluxes were controlled as known since long.

We shortened and refocused this section on the temporal interplay of drivers, which is the central contribution of the study.

L 640-644 not clear which numbers are compared current study to Maier 2022? High NUE in this study is probably not related to wheat cropping but to previous crop (grassland conversion) delivering high amounts of N_{min} from sward mineralization. Please note that 3 year grass-clover is often used in organic agriculture to feed 2 follow-up harvest of winter wheat without adding mineral fertilizer, which illustrate the high delivery of plant available N from this previous cropping season.

We clarified that the maize NUE comparison was derived from values reported in Maier et al. (2022) (lines 696–697), and we now caution that the apparent wheat NUE may be overestimated because N mineralized from the preceding grass-clover ley was not included in fertilizer input (lines 699–700).

L646 revise CO₂ in GHG budget (see above)

We revised this aspect by adding the distinction between net GHG exchange and GHG budget (lines 209–217), as described above.

L 654 role of vegetation not new at all

We assume this refers to the statement around line 694 (not line 654), as vegetation is not discussed at line 654. This subsection was replaced with a discussion of the explainable machine learning framework (new Sect. 4.1).

L 700-703 only one peak event investigated thus not representative for the cropping period

We agree and now explicitly restrict isotope-based interpretation to the March post-fertilization period (lines 731–733). We also quantify that March accounted for 45% of total winter wheat N₂O losses to clarify the importance of this time window.

L703f not adequate to derive mitigation strategies based on a one year study with only treatment. A multitude of field plots studies has investigated this before. The current study may be in agreement with the previous ones, but this single observation can't be used to confirm the previous results, and certainly not to propose N₂O mitigation strategies.

The former “Implications for N₂O mitigation” subsection was removed. Remaining management implications are integrated into Sect. 4.2, 4.4 and framed more cautiously.