

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 – Second revision

Dear Editor and Reviewers,

We thank both referees for their continued careful reading of the manuscript and for the additional comments on the revised version. We have revised the manuscript accordingly and provide below a combined point-by-point response to the second-round comments. Reviewer comments are shown in blue italics; author responses are shown in normal font. Section, table, figure, and line numbers refer to the revised manuscript, unless specified.

Overview of main revisions

- The uncertainty of the isotopic signature of emitted N₂O is now propagated through Eq. 3 by Gaussian error propagation, using conservative precisions of 1‰ for the isotope measurements and 2% (relative) for the N₂O concentrations (GC), based on literature. The resulting sample-specific uncertainties are now used as measurement uncertainties in the FRAME model and are reflected in the reported posterior distributions (Sect. 2.5, 3.5; Fig. 7b; Fig. A6).
- SHAP dependence plots for drivers other than GPP (Fig. A4) are now colored by another driver, making the relationships between drivers (e.g., WFPS and recent N input) visible. The text describing Fig. 4 has been revised to clarify that positive contributions of GPP to N₂O fluxes (i.e., positive SHAP values) correspond to lower GPP values (i.e., limited crop activity and thus plant N uptake), not to high crop activity, addressing the apparent implausibility of GPP dominance as a driver in winter.
- Discussion Sect. 4.1 was substantially expanded into a dedicated assessment of the explainable-ML (XGBoost-SHAP) framework: its intrinsic capture of predictor interactions, the limits imposed by the 30-day windows used for the fertilization and disturbance predictors, the redistribution of unrepresented drivers (N mineralization, labile C) onto other predictors, and a comparison of the framework’s role, strengths, and limitations relative to process-based biogeochemical models, including a concrete proposal for future integration.
- The statement attributing the high N₂O emissions mainly to the use of EC instead of chambers was reframed in Sect. 4.3: methodological differences are now presented as one of several possible contributing factors, alongside site- and management-specific conditions (N legacy of the preceding grass-clover ley, near-simultaneous mineral and slurry fertilization, and the poorly drained soil at the site).
- FRAME results for the reduced N₂O fraction (1 – r) are now reported for both the mixing-reduction and reduction-mixing scenarios (Fig. A6).
- Added an explanation of why XGBoost was chosen over recurrent architectures such as LSTM, given the limited training data available from a single site-year.

Response to Referee 1

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. [...] Authors revised the manuscript and solved many of the issues raised in my first review, but some still need the authors attention before the paper is suitable for publication. As mentioned in my first review, 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, the study presents only data from one site and one season under soil conditions and fertilization regimes which are quite special [...], the study thus does not expand our knowledge on the processes and their control. While the revised version downscaled some of the previous highlights of the flux dynamics, some aspects still overemphasize the relevance of the findings with respect to

adapted soil management (see details). The previous section on emission factors has been adequately adapted, but the discussion of the data is still directed towards general conclusion on mitigation, i.e. on the relevance of N synchronicity. Again, the one year and one treatment approach is not suitable to verify that effect. Moreover, previous studies and reviews have repeatedly demonstrated this. Thus it is not adequate to present this as a finding of this paper. The discussion should be limited to the fact that known regulation of N₂O fluxes was also observed in the data-set.

We thank the referee for this assessment and agree that the one-site, one-season, unreplicated nature of the study limits the generality of our findings, despite presenting a highly comprehensive dataset. We have further tightened the wording around N synchrony in Sect. 4.2 (L714-716): the relevant passage now states that our results are consistent with the established “N synchrony framework” rather than presenting this as a novel finding or as evidence that the synchronization theory was validated in our study.

Moreover, in the conclusion it was stated that in view of the higher fluxes in this one year wheat study compared to average fluxes of previous reports, authors claim that ETC is superior to chamber based approaches and the lower fluxes of the latter could be due to missing peaks between chamber measurements. Please remove this new statement (which was not claimed in the conclusions of the first submission). In addition to the insufficient relevance of a single observation, it is not adequate in view of the extreme site, extreme previous crop and unusual mineral-organic fertilization, all factors that lead to higher than usual fluxes.

We revised this passage in Sect. 4.3 (L723-731). Methodological differences are now presented as only one of several possible contributing factors, alongside the site- and management-specific conditions the referee lists: N mineralization from the preceding grass-clover ley, the near-simultaneous application of mineral N and slurry, and the poorly drained silty clay soil at the site. The residual methodological point is now supported by a citation to a recent review on soil N₂O research (Butterbach-Bahl et al., 2026).

Moreover, I am not aware that ETC has been independently evaluated for N₂O quantification but this would be another precondition for such a comparison. As far as I know, the ECT-CO₂ community still struggles to validate their high determined losses of soil-C, since long-term C-stocks show a different story. Thus, without validation, I would not exclude that ECT-N₂O might over or underestimate N₂O fluxes.

We appreciate the referee’s caution regarding independent validation of EC-based N₂O fluxes. EC and static-chamber N₂O measurements have in fact been directly compared in several independent studies, generally with reasonable to good agreement (e.g., Jones et al., 2011; Merbold et al., 2021; Pan et al., 2022; Pihlatie et al., 2005; Tallec et al., 2019).

Authors also did not follow my suggestion to compare the new ML-model to established statistical models or to process-based biogeochemical model like, e.g., DNDC. The future integration with process-based models is only mentioned in one sentence. But to better highlight the ML-part of the paper it would be good to get more in detail on this. There is need to show the potentials and limitations of the new model. This is not possible without discussing what its advantages could be compared to traditional approaches. Authors responded that use of process based models was beyond the scope of the paper as it needs multiple site and year data. Here I don’t fully agree. Using the data, it would be possible to run eg DNDC in default mode and check how variation of drivers identified in the ML model would change N₂O fluxes from DNDC. This could be used as a validation of the ML model. I would accept if authors still see this beyond the scope of the current paper. But to show how the ML model can be used in future it should be discussed more in depth, how it is different from other models, how further refinements (eg including mixed effects) could be useful, or if combination of the ML model with other models would lead forward, eg by modeling mineral N using process based models and use that data as additional driver in the ML model. Maybe authors have further ideas how to go forward which they could share in the paper.

Comparing the sensitivity of DNDC to the drivers identified by our ML model, as suggested by the referee, is not straightforward, as SHAP values and process-based model outputs are not directly comparable quantities. Process-based models such as DNDC include many more drivers than our ML model and resolve soil processes that we

could only represent by proxies: DNDC simulates mineralization rates, whereas we use days since soil disturbance, and it simulates soil mineral N, whereas we use engineered fertilization variables. Therefore, we consider this comparison beyond the scope of the present manuscript. Nevertheless, we have substantially expanded Sect. 4.1 to discuss the potential and limitations of the explainable-ML framework relative to process-based models in more depth.

I asked authors to show uncertainty of the input data of the FRAME evaluation. This has not yet been done and must be addressed in the revision (see details). The paper is thus not yet suitable for publication in the present form. I suggest major review taking into account the points above and below.

This is now addressed. Please see our response to the corresponding item below.

Details

L11 correct statement to clarify that isotopes were only measured in March

We revised the Abstract to state explicitly that the stable isotope analyses were conducted in March 2023 following a fertilization event, rather than throughout the winter wheat season (L12-13).

L12 SHAP needs to be explained in abstract

The Abstract now spells out and introduces SHAP (Shapley Additive exPlanations) as the method used to interpret the XGBoost model and obtain a time-resolved assessment of driver contributions (L13-14).

L30 $N_2 + N_2O$ loss should also be mentioned, eg. <https://doi.org/10.1016/j.cosust.2020.07.005>

We revised the Introduction to mention $N_2O + N_2$ losses and added the suggested reference (L32-33).

L60 I disagree with this statement in view of the multitude of modeling studies looking at the control of N_2O fluxes and how it can be explained, see eg Conen et al. 2001, <https://doi.org/10.1046/j.1365-2486.2000.00319.x>. Please revise.

Our statement refers specifically to the relative importance of individual drivers and how their contributions shift across a cropping season. The suggested study describes the general relationship between drivers and N_2O fluxes and predicts fluxes over a season but does not quantify how the relative importance of individual drivers changed over time – the very aspect our statement addresses. We therefore retained the original statement.

L91: Toyoda and Yoshida is not a suitable reference here as it addressed the N_2O isotopomer analysis only but not the source partitioning.

We replaced this citation with Pérez et al. (2001), which specifically addresses N_2O source partitioning using isotopic approaches (L93).

L93 in view of the uncertainties of using SP only “robust” is misleading. Suggest to name it: “can contribute to source partitioning..”

We adopted the suggested wording: the text now states that SP “contributes to the partitioning of N_2O sources” rather than describing the approach as “robust” (L95).

L97 again, clarify that isotopes were limited to March

The study objectives now state explicitly that isotope-based process identification was carried out during a sampling campaign in spring, rather than over the full study period (L100-101).

L223f There is need to address additional N from grass-clover decomposition, here or in the discussion.

We addressed this in the Results: the apparent emission factor is now explicitly described as also reflecting N_2O emissions from sources other than fertilizer application, such as the mineralization of residues from the preceding grass-clover ley (L495-499).

L275f Please report precision of isotopomer analysis for all values and also for GC analysis

As noted in our response to the first review, technical replicates were not available for each sample given the analytical method used, so sample-specific precision could not be determined empirically. We have, however, now used conservative precision estimates based on earlier studies to propagate uncertainty through the calculations (see next point).

L280 please state if the uncertainty of the isotopic values of emitted N₂O was calculated based on the uncertainties of the quantities in Eq. 3 and if it was implemented in the FRAME modeling to obtain correct uncertainty of the output (see Lewicki et al., 2022), eg the violin plots. If this was not done, please state in that Figure that it is excluded and the total uncertainty is larger than shown by the violin plots.

We have now propagated uncertainty through Eq. 3 by Gaussian error propagation, using the four measured input quantities and their assumed precisions. Because no replicate samples were available to determine analytical precision empirically, we adopted the conservative assumptions of 1‰ (1 SD) for the isotopic measurements and 2% (1 SD, relative) for the GC-based N₂O concentrations. The resulting sample-specific uncertainties are used as measurement uncertainties in FRAME and are therefore reflected in the reported posterior distributions, including the violin plots. Methods (L299-305), Results, and Discussion were updated accordingly.

L298 please explain why you report 1-r instead of r

We believe it is more intuitive, but we now mention the reason for this choice in L306-307.

L 519-551 As far as I can see, one major limitation of the current data evaluation is that interaction of drivers are not taken into account. At least, only the individual impact of drivers is shown in Figs 4 and 5, but not interaction of drivers, as it is done eg in statistical models. But it is clear that only the interaction of drivers leads to emission (eg high N and high wfps). Moreover, I found some unplausible observations in Fig. 4. E.g. during winter, GPP seems the dominant driver, but it is clear that due to the slow growth during winter this is not plausible. After fertilization there is a brief phase with negative WFPS effect. But it is clear that high fluxes following fertilization result from high Nmin and labile Corg in conjunction with high WFPS during that phase. It also appears that the definition of post disturbance and post fertilizer phases leads to artifacts, eg the immediate shift from fertilizer driver to WFPS in April and also the raise of GPP in Oct/Nov although it is clear that the effect can be only minor. I guess this results from definition of the diminishing disturbance, which neglects the huge impact from ploughing up grass-clover which may last during the entire growing season. These unplausible observations and limitations should be better addressed in the very brief discussion of the ML model performance.

We thank the referee for this detailed comment, which led us to revise Fig. A4 and to expand the discussion of the model's limitations.

On driver interactions: Gradient-boosted tree ensembles such as XGBoost capture interactions intrinsically, i.e., the effect of one predictor can depend on the values of others through the hierarchical structure of the tree splits. Interactions are thus already contained in the attributions shown in Figs. 4 and 5. SHAP attributions are additive by construction, however, so the contribution of such an interaction is distributed across the participating predictors rather than appearing as a separate term. The referee is right that our figures did not make this visible to the reader. We have therefore revised Fig. A4 so that each dependence plot is colored by a second driver, showing the conditions under which the contributions of a given driver occurred.

On the apparent dominance of GPP in winter and in October/November: We believe this comment reflects a misunderstanding of the sign convention of the SHAP values, which our previous wording did not make sufficiently clear. A positive SHAP value indicates that a driver increased N₂O fluxes relative to the baseline flux, and a negative value that it decreased them. In the specific case of GPP, positive SHAP values corresponded to low GPP (Fig. 5), i.e., to limited crop activity and thus limited plant N uptake, not to high crop activity. Furthermore, in Fig. 4 we show normalized SHAP values to analyze the temporal development of relative driver contributions. During winter and the autumn establishment phase, crop activity was indeed minimal, and the model attributed a positive

relative contribution to the absence of a competing N sink, independently of the generally low N₂O fluxes during that period. The attribution is therefore consistent with the mechanism described in Sect. 4.2. To avoid this misunderstanding, we have rephrased the relevant passages in Sect. 3.4.1 so that positive SHAP values of GPP are explicitly identified with low GPP (L546–548).

On the soil disturbance time window and the grass-clover ley: We agree that the 30-day windows shape the temporal pattern of some driver contributions, and that the influence of ley termination most likely extends well beyond 30 days, as does the input of labile organic C from the slurry application. Neither is represented in the model. Because SHAP attributes variance only to included predictors, contributions from unrepresented drivers are redistributed onto the other predictors. We now state this explicitly in Sect. 4.1 (L650–654)

.L606 but need to mention that the mixing reduction scenario yields lower values for r

Thanks for the suggestion. We now mention this (L614) and have added the reduced N₂O fraction to Fig. A6, so that both the mixing-reduction and reduction-mixing scenarios can be compared directly. The reduction-mixing scenario yields slightly higher reduced N₂O fractions than the mixing-reduction scenario used as our main framework.

L 639 a more in depth discussion on differences between the current ML-model and previous process-based models (mentioned above) would fit here.

This is addressed by the expanded Sect. 4.1 discussed above, which now explicitly compares our explainable-ML framework with process-based models such as DNDC, including their complementary strengths and limitations and a concrete proposal for combining the two approaches.

Section 4.3 still misses to state that this was untypical for winter wheat due to N mineralization of grass clover (e.g. doi.org/10.1080/03650340.2025.2545191), so not adequate to explain the high flux with continuous EC.

Thanks for the suggestion. We revised Sect. 4.3 (L725–731) to discuss the role of the preceding grass-clover ley, the near-simultaneous mineral-slurry fertilization, and soil type in shaping the observed high emissions. As one contributing factor among these, we retain the potential role of methodological differences between sparse chamber sampling and high-frequency EC, now supported by the review of Butterbach-Bahl et al. (2026).

L 749 This comes without context and should not be stated without discussing what these methods can or cannot reveal. Suggest to delete.

We agree and have deleted this sentence.

L 762 see comment on 4.3

We now mention the role of the N legacy from the preceding ley in the Conclusions as well (L796–798), consistent with the discussion in Sect. 4.3.

Response to Referee 2

I found the revisions to be comprehensive and findings robust; the measurements are extremely unique. I questioned the use of time series analyses from the SHAP results but feel that the partial dependence plot of NEE vs. SHAP N summarizes the major finding nicely. A question that I have is what approach was used to determine that XGBoost was an ideal ML method in this case? It tends to rise to the top of flux-related analyses, but with notable lags between driver and response (including in-soil transport) one might want to explore a LSTM. Is this something that was considered?

We thank the referee for the positive assessment. XGBoost was selected because tree ensembles perform well on tabular datasets of this size. Temporal dependencies between drivers and fluxes, including lags associated with in-soil transport, were therefore represented through explicit feature engineering rather than learned by the model architecture. Recurrent architectures such as long short-term memory (LSTM) networks could capture such lags directly, but a single site-year provides limited data for training sequence models. We now state this reasoning in Sect. 4.1 (L637-L639).

References

- Butterbach-Bahl, K., Philippot, L., Serra, J., Silver, W. L., Ogle, S., and Abalos, D.: Three decades of soil N₂O research—insights and gaps, *Glob. Change Biol.*, 32, e71038, <https://doi.org/10.1111/gcb.71038>, 2026
- Jones, S. K., Famulari, D., Di Marco, C. F., Nemitz, E., Skiba, U. M., Rees, R. M., and Sutton, M. A.: Nitrous oxide emissions from managed grassland: a comparison of eddy covariance and static chamber measurements, *Atmos. Meas. Tech.*, 4, 2179–2194, <https://doi.org/10.5194/amt-4-2179-2011>, 2011.
- Merbold, L., Decock, C., Eugster, W., Fuchs, K., Wolf, B., Buchmann, N., and Hörtnagl, L.: Are there memory effects on greenhouse gas emissions (CO₂, N₂O and CH₄) following grassland restoration?, *Biogeosciences*, 18, 1481–1498, <https://doi.org/10.5194/bg-18-1481-2021>, 2021.
- Pan, D., Gelfand, I., Tao, L., Abraha, M., Sun, K., Guo, X., Chen, J., Robertson, G. P., and Zondlo, M. A.: A new open-path eddy covariance method for nitrous oxide and other trace gases that minimizes temperature corrections, *Glob. Change Biol.*, 28, 1446–1457, <https://doi.org/10.1111/gcb.15986>, 2022.
- Pérez, T., Trumbore, S. E., Tyler, S. C., Matson, P. A., Ortiz-Monasterio, I., Rahn, T., and Griffith, D. W. T.: Identifying the agricultural imprint on the global N₂O budget using stable isotopes, *J. Geophys. Res. Atmospheres*, 106, 9869–9878, <https://doi.org/10.1029/2000JD900809>, 2001.
- Pihlatie, M., Rinne, J., Ambus, P., Pilegaard, K., Dorsey, J. R., Rannik, Ü., Markkanen, T., Launiainen, S., and Vesala, T.: Nitrous oxide emissions from a beech forest floor measured by eddy covariance and soil enclosure techniques, *Biogeosciences*, 2, 377–387, <https://doi.org/10.5194/bg-2-377-2005>, 2005.
- Talleg, T., Brut, A., Joly, L., Dumelié, N., Serça, D., Mordelet, P., Claverie, N., Legain, D., Barrié, J., Decarpenterie, T., Cousin, J., Zawilski, B., Ceschia, E., Guérin, F., and Le Dantec, V.: N₂O flux measurements over an irrigated maize crop: A comparison of three methods, *Agr. Forest Meteorol.*, 264, 56–72, <https://doi.org/10.1016/j.agrformet.2018.09.017>, 2019.