Discussion of "Seasonal variation in vegetation-climate interactions shape the CO_2 exchange in a degraded raised bog"

New title "The timing of warming matters as much as its intensity for the annual carbon balance of a degraded raised bog"

By Behrens et al.

Responses by the authors are marked in blue. Line numbers correspond to the revised manuscript with tracked changes.

Authors´ Response to Reviewer #1

This manuscript focuses on understanding the magnitude, seasonal dynamics, and main drivers of carbon fluxes from an abandoned drained peatland. Carbon fluxes were measured using an eddy covariance tower installed at the site, and net ecosystem exchange was further partitioned into gross primary production and ecosystem respiration. The authors then applied a range of statistical analyses to disentangle the effects of key climatic drivers, with particular attention to separating the influences of air temperature, vapor pressure deficit, and radiation. Overall, the manuscript is clear and well written, methods are rigorous, robust and described in a way that allows for replication. The figures are clear and show relevant information, and the analyses performed are adequate for the objectives presented in the paper. I think this manuscript would be a very nice contribution to Biogeosciences as it fits into its scope and would also be interesting to the readers of the journal as it looks at an underrepresented site in the literature. However, there are some points that need to be clarified before publication.

Response: We thank the reviewer for their positive feedback. We carefully considered the points raised and adjusted the manuscript in the relevant sections.

In the following is a point-by-point response in which we outline how we have revised the manuscript in accordance with the reviewer's comments.

While the authors note that data availability is insufficient to robustly estimate seasonal CH4 fluxes and the drivers, it is a bit unclear why there was not a lot of data available (was it precipitation? Bad quality? Below detection limits for the most part?), and which periods were considered for the CH4 budgets presented. Where the gaps evenly distributed across time-seasons? Or where there periods of bigger gaps like the one you mentioned due to rodent damage and lightning? A time series in the supplement would help to see this a bit better.

Response: Thank you for this suggestion. We have added the time series of the CH_4 fluxes in the revised version (Supplementary Fig S2).

The main issue preventing the calculation of a reliable annual CH_4 budget is that fluxes were very low (close to the detection limit) and random fluxes, not relatable to any measured climate variable. The measurements are collected with an open-path gas analyzer (as mentioned in line 187) and thus measurements are affected by bad weather conditions (e.g., during rain events). The fluxes missed due to bad weather conditions can be filled if they can be modelled based on their relationship with climate variables. In our case however, fluxes were too low and random to relate them to climatic conditions and thus the time series could not be gapfilled reliably. This justification is outlined in lines 420 to 424.

Figure 4 shows normalized daily mean GPP, but there is no mention in the methods section of how the normalization was done.

Response: We added the calculation of normalized daily mean GPP in the Methods for the derivation of seasonality (please see section 2.4).

Flux results are split into monthly aggregates, or seasonal values (non-growing season, growing season) but there are also mentions of early and late growing season stages. While each perspective is individually informative, the combined presentation leads to some redundancy and makes it harder to identify the core results. It would be good to merge or summarise the results a bit better to avoid redundant information both in the text and in the figures, as they all convey similar information (Table 1, Figure 5 and Figure 6). A more selective presentation that highlights only the most informative figures in the main text and relegating secondary results to the supplement would improve readability and focus.

Response: Thank you for this suggestion. We agree that there is duplicated information in Figure 5 and 6. We therefore removed Figure 5.

Table 1 is the main source of information on annual and seasonal budgets while Figure 6 visualizes the differences of monthly flux budgets across the three years.

The filtering and anomaly-based analysis substantially reduces confounding by seasonality, radiation, and the strong covariation between TA and VPD, and represents a thoughtful attempt to isolate short-term driver effects. However, despite these efforts, relationships involving Reco (and to a lesser extent GPP) are still partly conditioned by the partitioning approach, as the nighttime NEE partitioning explicitly assumes a temperature dependence of respiration. As a result, the derived sensitivities should be interpreted as method-conditional rather than as independent estimates of ecosystem process controls. I think this limitation should be more explicitly discussed, or at least acknowledged, in the Discussion, with clearer bounds on the mechanistic interpretation of these results.

Response: We thank the reviewer for pointing out this crucial point. In the context of the mechanistic interpretation of the results, we added text in the Discussion to acknowledge the artifacts introduced by night-time CO_2 flux partitioning on the temperature-response analysis and on the mechanistic interpretation of the results such as the rising Reco with rising TA (line 637-639) or the stomatal closure effect on reducing GPP (lines 670 to 671).

Minor comments:

- L 90 -95: I understand the meaning of this sentence: “To understand what drives carbon fluxes from abandoned drained bogs and to delineate the mitigation potential of restoration measures under a warming climate demands both precise flux measurements and a rigorous analysis of their drivers.” It is a bit confusing and perhaps missing a comma or two. I think it needs rewording. Here is my suggestion: "To understand what drives carbon fluxes from abandoned drained bogs and to delineate the mitigation potential of restoration measures under a warming climate, both precise flux measurements and a rigorous analysis of their drivers are required."

Response: We acknowledge that the original wording was complicated. The suggestion sounds more concise and we incorporated it in the final manuscript (lines 96-97).

- Please check the manuscript for ‘CO2’ use rather than CO_2

Response: We corrected the formatting throughout the paper.

- Could you provide coordinates for the location of the site?

Response: We added the coordinates of the site in the site description.

- Figure 2 looks great, a minor improvement would be to align the x-axis of figure 2d and 2b. I think that would allow the reader to see the variations on WTD due to precipitation a bit better.

Response: We thank the reviewer for pointing out the misalignment. We aligned the x-axes of Figure 2d and 2b.

Author Response to Review #2

This article presents an analysis of vegetation-climate interaction from measurements at an abandoned, degraded raised bog. I found the article interesting and the analyses were thorough, and I think the content would fit nicely within the scope of Biogeosciences. However, some revisions and clarifications are required.

Response: We thank the reviewer for their positive feedback. We carefully considered the points raised and adjusted the manuscript in the relevant sections.

The introduction is well written and provides the necessary background information on the topic. You could include the "low hanging fruit" of rewetting/mitigation potential message from your abstract in the introduction since that was missing. This message is important considering that across several countries there are abandoned drained fields that may be persistent sources of emissions.

Response: We thank the reviewer for this suggestion. The aspect of the low-hanging fruit is incorporated in the Introduction in lines 60-62.

In the first paragraph you can consider introducing natural peatlands as sources of methane, since in the second paragraph you talk generally about carbon emissions and not just CO_2 emissions. Carbon balances can/should also encompass lateral losses. While in temperate peatlands this loss may be small, it can decide whether a site is a net C sink. I recommend to either adjust this, or be more specific that you are writing about CO_2 fluxes only in this paragraph.

Response: We added the mention of CH_4 emissions from natural peatlands in the Introduction (lines 38). Additionally, we expanded text on the lateral carbon loss as an additional carbon flux (lines 629-631). This is added in the Discussion in Chapter 4.1 where the annual budgets are discussed to provide a comprehensive picture of the carbon balance. We further mention that the quantification of lateral carbon loss was beyond the scope of this study in the same lines.

The methods were generally clear, but some more detail is required: see my line comments below.

The seasonality detection was nice and avoids subjectivity to define season stages.

For the driver analysis in section 2.6, is there a reason why SWC wasn't included? It could have been an interesting addition for the driver/anomaly analysis and perhaps provided some insight into flux regulation that WTD doesn't provide. But I could see that it may have had high collinearity with WTD.

Response: SWC was not included because the measurements had gaps due to sensor failure. The WTD timeseries, which correspond closely to SWC, were more reliable. Thus we used WTD as a proxy of moisture availability in the peat. We clarified this point in the Methods section 2.6 (lines 301-303).

Assuming you used the REddyProc implementation of the nighttime partitioning method and didn't implement your own, Reco is fit on night NEE and TA using sliding 7 day windows in steps of 4 days to define the parameters. It is modelled for day and night, i.e. it doesn't keep the observed night NEE in the timeseries. For the driver analysis, since Reco is fit using TA, it will of course be a dominant driver (as you show later) and therefore those results are unsurprising. Because of the fitting windows/steps, your daily flux anomalies are likely smoothed.

Response: We thank the reviewer for pointing out this crucial point. As mentioned in our response to Reviewer #1 we added text in the Discussion to acknowledge the artifacts induced by the usage of partitioned data. We make sure to make this point early on in the discussion of GPP and Reco responses to TA (lines 637-639) and later when discussing the response of GPP to high VPD as a limitation (lines 670-671).

A complication that isn't considered/discussed is effect of gapfilling and partitioning on the driver/anomaly analysis. Only the night-time conditions are considering during the fitting of Reco, but then you are examining daily anomalies of the fluxes and drivers. I think the effect of e.g. daily VPD may not truly be represented for Reco considering the way it is fit and the typical diurnal cycles of VPD, though there are likely memory effects. This has some carry over effects to GPP. There is also the effect of the gapfilling of NEE on the daily driver analysis, since some of the drivers were also used as predictors for the gapfilling. These issues are hard to avoid unless you stick to only analysis of the measured data – and generally for understanding drivers using the direct flux observations would be the best. Unfortunately then you would only get the effect of drivers on daytime and nighttime NEE and not the partitioned fluxes. The analysis is still interesting and you can derive insights from it, but you should mention these circular/confounding issues in the Discussion.

Response: We thank the reviewer for pointing out this important aspect. We see that the true effect of VPD on Reco, e.g., in form of light-induced suppression of respiration during daytime, may not be truly represented in partitioned data. We incorporated the limitation arising from analyzing partitioned data in the Discussion in lines 637-639 and 670-671 as well as in lines 684 and 700.

For the driver analysis we only included days with minimum 50% measured data. We provided proof of the stability of our results by testing how the threshold for maximum gapfilled data per day affects the driver analysis results. To do this, we repeated the analysis for thresholds between 10 to 80 percent and provide a comparison of the resulting driver strengths (Supplement Fig. S1). Further, we implemented a bootstrapping approach in the regression analysis, repeatedly resampling the data and running the multiple linear regression for each threshold, so we can provide confidence intervals for each driver strength for each threshold of allowed gapfilled data.

In the revised manuscript, we explained this procedure in the Methods (Section 2.6, lines 294-299). We mention the stability of the coefficients with respect to the amount of included gapfilled data when presenting the results of the MLR analysis in Section 3.5 (lines 515-517).

It was unclear sometimes in the results if you were using the nighttime partitioned data introduced earlier, or your own partitioning you introduced in section 2.7, e.g. the results in Section 3.7 and Figure 9 (clearly your own method was used for the monthly responses in Figure 8). Or even if the gapfilled NEE/NEP data was used in e.g. Figure 9 for the analyses.

Response: This is an important point and we have made it clear in the revised manuscript. Our models introduced in Section 2.7 were only used to derive the parameters Q_{10}, A_{max} and α, not to partition CO_2 fluxes. We clarified this point in Section 2.7 lines 318-319 to avoid confusion.

Further, for the analysis regarding the response of fluxes to high TA and VPD conditions (Sections 2.8, 3.7 and Figure 9), we only used observed CO_2 fluxes, and the partitioned GPP and Reco related to those observed CO_2 fluxes. We clarified this point in the Methods Section 2.8 in line 341.

CH_4 budgets were included in a table but were not discussed or referred to in the text, though you noted the issues with gapfilling earlier. You could refer to them in Section 3.2 when you write that gapfilling performance was poor / uncertainties were too high (or exclude them entirely). See my line comments about the low number of data points.

Response: In Section 3.2 we mentioned that models for gap filling CH_4 fluxes failed (lines 420-422). We removed the unreliable CH_4 budgets from the table and added the complete time series of measured CH_4 fluxes in the Supplementary (to show the gaps and randomness of the fluxes). We clarified the justification for not including CH_4 in the carbon budget estimation in Section 3.2 lines 420-422.

The article was generally written well with good figures but there were some grammar issues, incomplete sentences, and incorrect units applied which need to be corrected. I pointed out some but probably didn't catch all of them – please check carefully.

Response: We thank the reviewer for pointing out these oversights. We have read the text carefully and made sure all mistakes were corrected in the revised version.

Units: See line items. Spaces between the units (e.g. g C m^{-2} a^{-1}, not gC). Annual flux units should have a time unit a^{-1} or yr^{-1}. CH4 flux you had units of mmol m^{-2} s^{-1}, which I assume should be nmol m^{-2} s^{-1} since otherwise that is a huge flux and would exceed your detection limit. A couple of issues with radiation and VPD units.

Response: We carefully checked all units and corrected any issues.

Inconsistent uses of subscripts at times, e.g. Q_{10} and Q10, CO2 and CO_2.

Response: We thank the reviewer for pointing out these errors. We thoroughly checked the text and corrected any formatting mistakes.

Check language around when discussing WTD, it can be confusing to know if you are talking about shallower or deeper water tables at times when it is "less" or "more" etc. Sticking to e.g. deep and shallow, raised or lower etc. is clearer.

Response: To avoid confusion, we changed from water table depth (WTD) to water table (WT) in the revised version. Using WT allows us to use "low" for deep and "high" for shallow water tables, as this terminology is less confusing for the readers. Also, we made sure the terminology is used consistently throughout the text.

The tense in the MS was not always consistent, check carefully.

Response: We made sure the tenses are used correctly and consistently throughout the revised version.

Line comments

Abstract

L9: The first sentence does not flow that well, I would rephrase it. Carbon dioxide is also not defined yet if it needs to be.

Response: We rephrased this sentence as: "Pristine peatlands function as natural carbon dioxide (CO2) sinks, but anthropogenic drainage turns them into sources of CO2, responsible for 2–5% of global annual greenhouse gas (GHG) emissions." (lines 11-12).

L14: methane (CH4) if it needs to be defined

Response: We added the definitions of CH_4 at the first mentions in the Abstract and Introduction (lines 18 and 38).

L15: northwest Germany

Response: We corrected this (line 19).

L16: Perhaps clarify the units that 15 cm is below ground and that WTD is in the notation of positive units -> deeper. Or just write water tables consistently deeper than 15 cm if you don't have space here.

Response: We used WT (water table) instead of WTD (water table depth) to avoid confusion (please see our response to a previous comment).

L17: gC m-2 a-1 -> g C m-2 a-1

Response: We corrected all units in the revised manuscript.

L18: greening or growth?

Response: Given that we estimated the start of the growing season from GPP, we have changed this statement into "[...] triggered an earlier onset of CO_2 uptake and substantially increased maximum CO_2 uptake capacity from April to June. [...]" (lines 22-23).

Introduction

L31: carbon dioxide should be defined again if it needs to be

Response: We added the definitions of CO_2 at the first mentions in the Abstract and the Introduction (lines 11 and 36).

L34: "leading to carbon losses in form of CO2": be more specific here, the aeration enables microbial peat oxidation and this is the largest C loss pathway. Carbon losses could also entail DOC/POC losses, but microbial peat oxidation enabled by drainage is normally the greatest.

Response: We highlighted the high proportion of carbon lost through drainage-induced aeration more by rephrasing to:
"However, drainage of peatlands for peat extraction, forestry, or agriculture (Leifeld &

Menichetti, 2018) results in the aeration of the peat column. This eases the constraints on microbial decomposition, leading to peat oxidation and carbon emissions in form of CO_2." (lines 38-41).

L68: "when WTD is consistently high (> 20 cm)" do you mean consistently deep? Intuitively when one talks about high water tables you think of it being close to the surface.

Response: We agree that the usage of "WTD" is more confusing than "water table" (WT). In the case of WTD, if a "depth" has "high" values, it means the water is furthest from the surface, thus a high water table depth translates to a low water table. To avoid future confusion, we switched to the use of "WT" in the whole manuscript. The sentence now reads "However, the relationship between WT fluctuations and Reco can break down when WT is consistently low (deeper than -20 cm),[…]" (lines 72-73).

Methods

L110: in this first paragraph I would write the former land use of the site, based on L115 it was for peat extraction so clarify this.

Response: We added the former land use of the site (peat extraction) in the Methods (line 117).

L120-125: consider adding the common names of the plants

Response: We added the common names for all plants mentioned in the site description (lines 133-137).

Figure 1: include the footprint model that you used to calculate the climatology in the caption.

Response: We added the mention of the Kljun flux footprint model as well as the reference for it in the caption of Fig. 1.

L135: in the meteo measurements, you have included the company names and locations for some of the instruments but not all. E.g. for Onset Hobo, you have included Bourne, MA, USA but for Hukseflux, Stevens Hydraprobe and others they are not included.

Response: We thank the reviewer for catching this inconsistency. We added the missing information for all instruments.

L146: german -> German

Response: This is corrected.

L154: "The first": the MX2001?

Response: We clarified this by mentioning the MX2001 logger directly:
"The MX2001 WT data logger had several power outages" (lines 174-175).

L155: I assume what happened was the Thünen instate logger was corrected for linear offsets relative to the MX2001?

Response: The time series of the complementary WTD data spanned from 19.09.2023 to 05.03.2025. Therefore data from the MX2001 from 2023 and from 2025 are needed. Because the MX2001 logger was moved in January 2025 the data from 2023 may differ from that in 2025. The complementary data was used as reference and data from the MX2001 before 19.09.2023 and after 05.03.2025 were individually debiased with OLS. We made sure this point is clarified in the text by adding in lines 173-179:

"[...] The MX2001 WT logger had several power outages and was moved in the beginning of 2025, resulting in data gaps and a non-continuous time series. The complementary WTD logger covered the period from 19.09.2023 until 05.03.2025. To create a full time series spanning from the beginning of 2023 until the end of 2025, data from the complementary data was used as a reference, and the time spans before 19.09.2023 and after 05.03.2025 were inserted from the MX2001 data. To remove offsets due to the sensor positions, data from 2023 and 2025 from the MX2001 were individually fitted to the complementary WT data, with an ordinary least squares (OLS) regression (R^2 = 0.91 for 2023 and R^2 = 0.88 for 2025)."

L158: R in R^2 should be in italics

Response: This is corrected everywhere in the manuscript.

L160: LI7200RS -> LI-7200RS

Response: This is corrected.

L162: LI7700 -> LI-7700

Response: This is corrected.

L163: strictly the fluxes were not measured at 10 Hz, rather the 3D wind speeds and gas concentrations were. You calculated half-hourly fluxes.

Response: We changed the wording to: "Gas concentrations and wind speeds were measured on a tripod at 2.77 m height and with a frequency of 10 Hz."

L164: include the full EddyPro version number, most probably v7.0.9

Response: That is correct, we inserted the full version number (line 190).

L165: raw 10 Hz data was->were filtered

Response: This is corrected.

L166: the simple covariance maximisation technique is generally inappropriate for H2O with the LI-7200RS due to the dependence of lag time on relative humidity. Did you check if the lag windows were appropriate? The automatic timelag optimisation is a better approach for non-passive gases with closed path instruments in EddyPro. The impact on your results is probably not large since you didn't make extensive use of LE outside of the EBC, but it is used in other corrections such as the WPL. See Sabbatini et al. DOI: 10.1515/intag-2017-0043.

Response: We thank the reviewer for highlighting this point. We agree that the effect on CO_2 fluxes is likely small. In our case, water vapour fluxes which would most be affected by the maximisation technique, were only used in the WPL correction of the CH_4 fluxes, since CO_2 fluxes were measured with a closed-path analyzer. However we do not anticipate that the lag determination would affect the CH_4 fluxes substantially given that measured fluxes were already small and close to the measurement uncertainty. The WPL correction would improve any overestimation rather than increasing the uncorrected fluxes. In the revised version we applied the automatic lag detection instead of covariance maximization and mentioned this in line (195).

L169-170: the correct reference for the high-pass filtering effects is Moncrieff et al. (2004), the year is wrong, see DOI: 10.1007/1-4020-2265-4_2

Response: We corrected the reference (line 197).

L170: I assume the spectroscopic correction was also applied for CH4 with the LI-7700, see McDermitt et al. 2011 DOI:10.1007/s00340-010-4307-0. Again, the LI-7700 measures gas concentrations (well molar densities actually) of CH4 and not fluxes.

Response: Spectroscopic corrections were indeed applied during flux processing in EddyPro. We added this in lines 198-199:
"CH_4 concentrations measured with the open-path analyser were corrected for density fluctuations using the WPL correction (Webb et al., 1980) and spectroscopic effects (McDermitt et al., 2011)."

L175: did you filter the LI-7700 using signal strength? If so what threshold?

Response: We filtered the CO_2 flux data based on a signal strength of 80% and the LI-7700 based on a signal strength of 10% (McDermitt et al., 2011). We added this information in the Methods (Section 2.3, lines 191-193).

L179: After the absolute limits removal, the remaining spikes…

Response: This is corrected.

L182: The RStudio version is not important, rather the version of R is if you want to include something.

Response: We added the correct R version instead of the RStudio version (line 213).

L184: What was the determined threshold? How many quantiles were extracted in addition to the determined threshold, just 0.05 and 0.95? Or e.g. 40 quantiles spaced evenly as in OneFlux?

Response: We added the annual u*-threshold values in line 217-218. We also clarified that we used 37 evenly spaced quantiles between 0.05 and 0.95, in spaces of 0.025 (line 215).

L203: I assume these other uncertainties were also added in quadrature? Was the uncertainty due to u* threshold computed as the variance of the different gapfilled totals for each u* quantile threshold?

Response: We did add the other uncertainties in quadrature as well. To make this more clear we changed the wording in lines 242-245 to:

"Uncertainty of the annual CO_2 balance was computed by adding in quadrature the aggregated uncertainties arising from random measurement errors (Finkelstein & Sims, 2001), u*- threshold selection, quantified as the standard deviation of annual balances derived from gap-filled flux time series generated using alternative u* thresholds (Pastorello et al., 2020), and aleatoric and epistemic gapfilling model uncertainty (Vekuri et al., 2025)."

L208: NEE introduced but not defined. Did you add the single point storage estimate to get NEE or is just the CO2 flux? I assume there was no profile since the tower was low at 2.77 m.

Response: We added the definition of NEE here as the net ecosystem CO_2 exchange.

There was no profile measurements as the EC instruments were installed at 2.77 m. Storage fluxes are low (abs. median of 0.06 µmol m^{-2} s^{-1} compared to an abs. median CO_2 flux of 1.23 µmol m^{-2} s^{-1}), and were not included in the NEE calculation. In the revised version we added the storage flux to the CO_2 flux to calculate NEE. This is now clarified in lines 202-203.

L211: convention is to use G as the soil/ground heat flux, as you did in Figure S3.

Response: We changed to "G" for soil heatflux throughout the manuscript.

L231: all available observations – only observed and not gapfilled?

Response: The anomalies are calculated on the daily scale from all data, measured and gapfilled. In the subsequent driver analysis (Section 2.6) we imposed a threshold which filters out any days that have more than 50% of gapfilled data. As we are using a daily timescale only a small fraction of days have no gapfilled data. Removing all days with any gapfilled data would result in large data loss as well as a skewed filtering. Gapfilled data that is created close to measured data is generally more robust than gapfilled data in long gaps of several days or weeks (Vekuri et al., 2025). As mentioned in response to a previous comment, we examined the effect of the daily threshold of gapfilled data in the supplementary (Fig. S1) and referred to it in Section 3.5 (lines 515-517).

Additionally, we made the derivation of anomalies on the daily scale using all data clear by rephrasing in lines 287-288 to:
"Daily anomalies were calculated as the deviation of each daily data point from the mean of the daily data within a ±3 day-of-year window across all three years. "

Further, we added in lines 296-298:
"To minimize the effect of gapfilling on the driver analysis we used a threshold of maximum 50% gapfilled data per day. We tested the influence of the percentage by repeating the analysis for thresholds from 10 to 80%. The results are presented in the supplementary material."

L250: Q10 was earlier with subscript 10 in L105

Response: Throughout the revised version, we now consistently use Q_{10}.

L253: All Reco is modelled using the temperature relationship determined from NEE night, not just Reco day if you used the REddyProc implementation. You can see that in your response functions in Figure 9

Response: We rephrased this line as:
"NEE_{night} is used instead of the partitioned Reco because Reco is already modelled." (line 321-322).

L261: Sentence starts lower case and missing 'to', should it be "To determine the light response of the ecosystem fluxes , …"?

Response: We thank the reviewer for catching this. The missing "to" was added.

L264: why is night time defined as < 10 W m-2 but daytime is > 50 W m-2?

Response: This is indeed a mismatch between the post-processing pipeline and the data analysis pipeline which is now corrected. We consistently used a 10 W/m² threshold to determine day/nighttime conditions (lines 209 and 333).

L266: µMol -> µmol, fix here and other usage

Response: We fixed this typo throughout the manuscript.

Results

L290: be consistent with the number of digits for reporting R^2. Instead of writing "it was" be more specific, e.g. the R^2 was

Response: We reported with a consistent number of digits and corrected the wording as suggested.

L295: between 11.0 and 11.6ºC

Response: This is corrected.

L296: W m-2 instead of kW m-2

Response: This unit error is corrected.

L297: total rainfall (Fig, 2b), respectively,

Response: We adjusted the wording as suggested.

Figure 2: figure resolution needs to be increased

Response: We increased the resolution of the final figures to at least 300 DPI.

Figure 3: nice Figure. Daily P might look better as a barplot rather than line

Response: We changed daily P to a barplot.

L324: The year 2025 was by far the driest…

Response: We agree that this wording is more fluent and incorporated it.

L330: What is the average closure of FLUXNET2015?

Response: The average closure is provided in the Results (line 403).

L336: "The ensemble neural networks filled the gaps with an average R2 across all u*-scenarios of 0.85±0.03, uncertainty in the R2 expressed as the 95% confidence interval across all models."

This should be rephrased. You cannot know the R2 of the gaps since you don't have the data (unless you are talking about artificial gaps, but I don't recall you introducing them). Rather, is this the average score of the validation sets? Reporting the RMSE and bias would also be useful.

Response: We agree with the reviewer. To provide a robust estimate of the performance of the gapfilling model we added an iterated temporal-split validation with temporal blocks of different sizes being held out during training to give reliable validation metrics for different gap lengths. Additionally we added the RMSE and the bias in Table 1.

L339: this is a high proportion of data points filtered out for CH4. Was it because the data quality/signal strength was too low often? Note that if the fluxes are regularly below the detection limit you should not use the stationarity flag of CH4 (included in the Mauder and Foken test), rather you should apply the CO2 stationarity flag. See Nemitz et al. 2018: Filtering and gap-filling section on page 536. DOI: 10.1515/intag-2017-0042.

In any case, given the low fluxes the overall impact on your study will be low.

Interesting that you didn't see regular CH4 fluxes given the water levels could be elevated at times with warmer temperatures, but also considering the presence of drainage ditches. Particularly the summer of 2023 there was a wetter period, but you did not have measurements then.

Response: We thank the reviewer for pointing this out. In fact there was an error in the calculation. The correct percentage of filtered data is 58% which is closer to typically expected filtered out amounts of data. This is now corrected in line 416.

L340-341: mmol m-2 s-1 -> nmol m-2 s-1, otherwise those are large fluxes of CH4!

Response: We checked any mention of CH_4 fluxes throughout the manuscript and corrected the units to nmol m^{-2} s^{-1}.

L349: "reducing all correlations controlled for to maximum remaining Pearson correlation coefficient of 0.13" please rephrase

Response: We rephrased this statement in lines 566-568 to:
"The filtering methods successfully isolated the effects of TA and VPD. The remaining correlations of fluxes with the controlled variables did not exceed a Pearson correlation coefficient of 0.13 (Fig. S6)."

L368-369: CO2, 2 should be in subscript

Response: We fixed subscript errors throughout the manuscript.

Table 1: how did you put CH4 into units of g CO2-C eq. m^{-2} yr^{-1}? If a GWP value was used, state which and what timeframe in the figure caption. Be consistent with the number of decimal places for each value in the table. Your Reco and GPP totals add up to NEE for 2024 and 2025, but there is small difference for 2023.

Response: In the revised version we do not provide annual budgets of CH_4, because the fluxes could not reliably be gapfilled. We only mention CH_4 in flux units of nmol m^{-2} s^{-1}. This is explained in lines 420-422.

We checked the exact flux sums for NEE, GPP and Reco for each year (and season) and provided the exact matching sums in the revised version (Table 2).

Section 3.5: It's interesting that the climate anomalies didn't explain much of the flux in EGS, however there was clear diverging behaviour/conditions as shown by parameters in Figure 8 – particularly in 2024.

Response: We added this observation in the Discussion in section 4.3, lines 720-721 where the variations in P_{max}, α and Q_{10} in the early growing season are discussed.

Fig 7: Z_WTD had positive effects on Reco and GPP but a mixed response on NEE, why? (Okay, I see a reason in discussion later).

L411: Interesting that lower WTs enhanced Reco but VPD decreased it. Which growing season for the VPD impact, both EGS and LGS (seems that only EGS from the figure)?

Response: Indeed Reco is only suppressed during the EGS and not in the LGS, which is also visible in Fig. 9 e. We changed the wording in line 530 to:
"Conversely, positive VPD anomalies suppressed Reco during the EGS."

L419: missing full stop at end of sentence.

Response: We added the missing full stop (line 540).

Fig 7: black orders -> borders.

Response: This is corrected in the caption of Fig. 7.

Figure 9: VPD units should be in hPa? Add regression goodness of fit the plots as well? Would be good to use consistent y-ranges for the plots and extend Reco and GPP to 0. Ranges may be consistent but the Reco right side plot is missing the 5 tick.

Response: The unit for VPD was changed to hPa in Fig. 9. We added the regression goodness of fit (R^2) to the plot and made the y-ranges consistent and correct.

Discussion

L461: include the uncertainty interval of these totals, units should be g C m-2 a-1

Response: We added the uncertainty ranges and corrected the unit (lines 591-593).

L474:476: be more specific about the budgets of the other studies, do you mean GPP and Reco are both similarly within that range?

Response: The sums encompass both GPP and Reco, which we are now explicit about. We rephrased this in lines 606-607 as: "where annual sums of GPP and Reco typically range between 600 and 800 g C m^{-2} (Humphreys et al., 2014; Hurkuck et al., 2016)."

L490: Neither of those studies included Molina caerulea, though they did write about the effect of pressurised flow systems of vascular plants and CH4 emissions. You could also add a citation that showed the effect of M. caerulea, e.g. Leroy et al. 2017 DOI:10.1016/j.soilbio.2017.01.005

Response: We thank the author for pointing us towards this specific reference and added it to support our text (line 622).

L475: Is the higher nutrient availability, as indicted by your elevated N and lower CN, also a reason for the higher productivity compared to the semi-natural sites and enabled the dense shrub cover?

Response: That is possible, however a direct attribution of nutrient composition and vegetation productivity is not possible based on our data. We included this point in the Discussion in lines 615.

L490: do you know anything about the water / peat chemistry? Could there be other terminal electron acceptors that suppress CH4 emissions here?

Response: In Amtsvenn, the peat below 30 cm depth contains low concentrations of inorganic alternative terminal electron acceptors (Lemmens et al., 2026), yet ombrotrophic peat typically contains redox active organic moieties, contributing considerable electron accepting capacities (Guth et al., 2023). Thus a suppression of methane formation by alternative electron acceptors during the warm summer period may occur. Although nitrogen and phosphorus concentrations are elevated in the upper 10–15 cm of the peat, peak water levels only rarely reach these layers. Moreover, when they do, low temperatures are likely to co-limit methane production. We have added this discussion in line 623-628.

L496: this should be self-evident since Reco is defined by the relationship with TA

Response: We acknowledge this in the context of Reco partitioning in lines 683-684.

L501: I'm not sure I agree with your phrasing about the contrasting role. WTD does or should control emissions across sites since it limits the oxygen intrusion into peat and as a result limits decomposition, and it forms a nice relationship at the seasonal/annual scale with net emissions typically. I think to find a short-term effect it will be tricky with NEE if you don't examine day/night separately – as you note with the effect of Reco/GPP and WTD cancelling out. And hence why you find a better relationship in the NGS.

Response: We acknowledge the differences in spatial and temporal scale of the inferences made by the referenced scales and rephrased this as follows in lines 647-651:
"While WT is widely accepted as the primary driver of CO_2 emissions on the annual timescale and across sites (Evans et al., 2021; Ma et al., 2022; Tiemeyer et al., 2020), short-term anomalies in WTD had little influence on daily NEE (Fig. 7)."

L573: To the best of our knowledge, there…

Response: The missing comma was added (line 736).

Supplement

Figure S2: ET isn't in the figure?

Response: ET was removed from the caption in Fig. S2.

Figure S3: fix the date range(2023–202)

Response: This was changed to (2023-2025).

Figure S4: no (a) and (b) labels in figure

Response: The (a) and (b) labels are added to the figure.

Author Response to Review #3

This paper is well-written, methodologically strong, and engaging. Although it presents only three years of eddy covariance data from a drained, shrub-dominated raised bog, the authors maximize the value of this dataset. This type of abandoned, unmanaged peatland is both climatically relevant and significantly underrepresented in the literature. The analytical framework is thorough for a single-site study: the combination of anomaly regression, ecophysiological response curves, and the sequential filtering approach to separate TA and

VPD effects demonstrates strong methodological ambition and careful attention to confounding factors. The central message, that the timing of a warm spring matters as much as its intensity for the annual carbon balance, is clearly conveyed and well supported by the data. The figures are effective, the writing is precise, and the authors are transparent about the limitations of a three-year dataset. The comments below primarily clarify and suggest ways to tighten the framing, none of which undermine the core findings.

Response: We thank the reviewer for their encouraging and positive feedback. We would like to take the opportunity to streamline the title to better reflect the main message by changing it to:
“The timing of warming matters as much as its intensity for the annual carbon balance of a degraded raised bog.”

General

G1. Length of the measurement record

The eddy covariance record spans only three years, which limits the ability to draw robust conclusions about long-term drivers or to place the observed interannual variability in a broader climatic context. The authors are aware of this and acknowledge it in the conclusions, which is appreciated; this is a classic, project-related constraint. That said, the three years captured do include substantial climatic contrasts (wet/warm 2024 vs. dry 2025/2023), which partially offset one another. No additional action required, but explicitly framing this as a limitation when discussing generalisability, particularly for the spring warming and VPD findings, would strengthen the paper.

Response: We agree that the three year span is a limitation of the paper. We explicitly mention this limitation in the conclusion in lines 759-763:
“While this study spans three years, the period captured substantial inter-annual variability, providing valuable insight into peatland responses to the climate. Continued observations over longer timescales would further strengthen the assessment of whether the observed patterns—particularly the spring warming response and the seasonal divergence in VPD effects—represent persistent characteristics of this ecosystem.”

G2. Absence of remote sensing or greenness indicators for phenology

Given that vegetation phenology is a central element of the study, it is somewhat surprising that no remote sensing datasets or optical greenness indicators (e.g., NDVI, PhenoCam, or near-surface RGB camera indices) were used to independently corroborate the GPP-derived phenological transition dates. Such datasets are now widely available at high temporal resolution and would provide an empirical cross-check on the SOS/POS/EOS dates derived from the double-sigmoid fit. The authors should, at a minimum, discuss whether such data exist for the site and, if not, why they were not used, or acknowledge this as a gap.

Response: We agree that the inclusion of other flux-independent sources of phenology would strengthen the analysis. There is in fact a phenocam installed at the site. However, the data has issues in 2023 and a data gap near the expected transition point in February/March 2024. Further, we tested remote sensing indices like daily MODIS NDVI to additionally derive the phenological transition dates but the lower temporal resolution and missing data due to noisy pixels in the satellite data product provided a worse resolution than deriving phenology from our GPP measurements. We added these points into the Methods section 2.4 (lines 260-264) where we justify our method for deriving phenology

G3. Scale mismatch in the WTD discussion

The manuscript argues (Section 4.2.1) that the minor role of WTD anomalies on daily NEE contrasts with “its widely accepted role as the primary driver of CO_2 emissions across sites.” However, the cited cross-site studies (Evans et al., 2021; Ma et al., 2022; Tiemeyer et al., 2020) establish WTD–CO_2 relationships at the interannual or cross-site scale, not at the ±3-day anomaly scale used here. These are fundamentally different analytical contexts. A lack of daily covariation in anomalies does not contradict WTD as the dominant structural control on absolute emission levels. The framing should be revised accordingly.

Response: As stated above we rephrased this text in lines 647-651 as follows: “While WT is widely accepted as the primary driver of CO_2 emissions on the annual timescale and across sites (Evans et al., 2021; Ma et al., 2022; Tiemeyer et al., 2020), short-term anomalies in WTD had little influence on daily NEE (Fig. 7).”

Additionally following the suggestion of Reviewer #1 and #2 we changed the notation of WTD to WT throughout the revised manuscript.

Line Comments

l. 130 - Site heterogeneity

The site is described as moderately heterogeneous, with varying cover of Calluna, Molinia, Erica, graminoids, and individual Betula trees. Nothing wrong with that for an EC study, but it is somewhat surprising that spatial heterogeneity does not appear to be addressed further, for instance, through footprint-weighted vegetation composition or a comparison of flux behavior across dominant footprint sectors. This may be out of scope for this paper, but a brief sentence acknowledging it as a caveat would be appropriate.

Response: While there is some moderate heterogeneity, especially through some patches of dwarf birches, the vegetation is not well separated in sectors. This is visible to some extent in Figure 1. The canopy consists of ericaceous shrubs with interspersed graminoids such as *Molinia caerulea* and birch seedlings throughout the footprint. We therefore do not expect different flux behavior from different sectors of the footprint. We included this point in the Site Description (lines 131-132 and 138-139).

l. 140 Disappearance of SWC and TS as covariates

Soil water content (SWC) and soil temperature (TS) are measured at the site (two profiles at 5 cm depth) and reported here, but they do not appear as predictor variables in the gap-filling or driver analysis frameworks. Given that SWC and TS are direct controls on microbial decomposition and that Reco is included in the analysis, their absence warrants explanation. Was spatial heterogeneity too large for these point measurements to be representative of the footprint? Or were they excluded because of data gaps? Please clarify.

Response: This is a point that warrants clarification. During inspection of the data we found that both TS and SWC sensors had more data gaps in them than our final product of TA and WTD data. Due to the collinearity of TA and TS (R= 0.83) and SWC and WTD (R= 0.56) we decided to only incorporate one of each into the driver analysis. We clarified this point in the Methods section (lines 301-303) to avoid any confusion.

l. 191 - No hydrological predictor in gap-filling

Water table depth is not included among the gap-filling predictors (TA, VPD, SWIN, hour, month). For a deeply drained site where WTD fluctuates substantially (Fig. 3e), this is surprising, even if WTD is ultimately found not to be a dominant daily flux driver (Section 3.5), excluding it from the gap-filling model means the model cannot learn any WTD signal that may exist, which could become circular when the driver analysis subsequently finds no WTD effect. At a minimum, the rationale for excluding WTD from the gap-filling predictors should be explicitly stated, even if it is not standard for non-peat-related sites.

Response: In the revised version we included WTD as a predictor and re-ran the gapfilling model.

l. 192 - Test data split for gap-filling evaluation
"20% of the data was set aside as test data to evaluate the model's performance."

Please clarify whether this 20% was drawn by random sampling or a temporal block holdout. Random sampling of half-hourly data yields optimistically biased R^2 due to autocorrelation between adjacent time steps. A block holdout would better reflect real gap-filling skill, especially for the long equipment-failure gaps (19 and 29 days). Again, all depends on the gapfilling objective; it is very legit for small gaps.

Response: We acknowledge this difference in testing scopes for ML inference models. In the current version the test samples were randomly drawn. We added an iterated temporal-split validation with temporal blocks of different sizes being held out during training to give reliable validation metrics for the model for different gap lengths. This is now explained in lines 236-241. Additionally reported the RMSE and the bias in Table 1.

l. 195 - Neural network architecture

Two hidden layers with 50 nodes each represent considerable model capacity for a single-site problem with 5–6 input predictors. The ensemble approach and Laplace uncertainty quantification partially address the risk of overfitting, but it would help to be more explicit about what this model is being asked to do in practice. Outside the two major equipment failures, gaps are presumably short (hours to a few days), for which a simpler model would likely perform equally well. Does the reported $R^2 = 0.85 \pm 0.03$ reflect performance on short gaps, long gaps, or both combined? Clarifying the distribution of gap lengths and model performance by gap-length category would help readers assess whether the architecture is justified.

Response: As mentioned in response to a previous comment, we incorporated a more robust testing procedure, providing information on how well the model filled gaps of different lengths in the flux data, and added these results in Table 1.

l. 222–224 - Description of L1 and L2 in Eq. (2)

"L1 and L2 are the times of the early summer and later summer plateauing of GPP."

This is incorrect. In Eq. (2), L1 and L2 are the asymptotic amplitude parameters (same units as GPP, not time). The timing parameters are t0 and t1. Please correct.

Response: Thanks for detecting this error. The description of the function parameters was corrected in the revised version (lines 268-271).

l. 305 - Figure 2: no long-term SWIN reference

Long-term climatological context is provided for TA and P from the nearby DWD station, but not for SWIN (Fig. 2c). Is this because SWIN is not available from the meteo station over a comparable period, or is it because the data are not comparable? If so, this should be stated - without a long-term reference, it is harder to assess whether some of the radiation conditions in 2023–2025 were anomalous.

Response: Indeed, there is no radiation data available from a nearby DWD station. We clarified this in the caption of Figure 2. Additionally, we added the long-term mean temperature and precipitation in line 123.

l. 315 - Figure 3: 2024 WTD

The 2024 WTD time series (Fig. 3e) looks markedly different from 2023 and 2025, with a not-so-deep summer drop. A good reason to investigate this effect for the 2024 carbon balance, which looks very different from the others.

Response: Indeed the WTD in 2024 was markedly higher in July and August compared to the other years due to some high precipitation events. Looking at Figure 6 we found that this is not the period where NEE differs from the other years, therefore it seems that WTD did not rise enough to effectively suppress soil respiration. In contrast, Reco was lowest in 2025 (Fig. 6), when WTD dropped to its lowest levels (Fig. 3). We included this point in the Discussion in lines 663-666.

Fig. 4 (l. 353–360) - Double-sigmoid fit: envelope, bias, and phenological metric

The double-sigmoid fits capture the overall seasonal envelope of GPP well across all three years. However, the fitted curves tend to run slightly above the bulk of the observed daily values (blue dots), suggesting a modest but systematic overestimation of GPP around peak season, visible in all three panels. Since SOS, POS, and EOS are derived from the third derivative of this fit, a consistent upward bias in the envelope need not affect transition timing per se, but it may influence the partitioning of flux budgets between EGS and LGS if the peak date shifts accordingly. The authors should comment on whether this affects the seasonal budgets in Table 1, and ideally report confidence intervals for SOS, POS, and EOS.

Response: In the revised manuscript we provided confidence intervals for the seasonal transitions derived by bootstrapping the double sigmoid function and the respective derived third derivatives, explained in the Methods section (lines 280-283). We presented the confidence intervals for the transition dates by adding error bands in Figure 4 and reporting the values in the text. We tested how much the flux budgets varied when using the maximum versus minimum lengths of the derived seasons, and reported this in the Results section 3.4 (line 484), where the seasonal budgets are presented.

Additionally, the phenological transitions are expressed in day-of-year. Given that the spring onset in this study is strongly linked to accumulated warmth (the warm spring of 2024 being the key example), it would be worth discussing whether expressing the phenological axis in accumulated degree days - standard in crop phenology - would provide a more mechanistically interpretable picture of the growing season dynamics.

Response: We agree that accumulated degree days can provide a more mechanistic representation of phenological development, particularly in systems where temperature accumulation strongly controls seasonal transitions. However, in this study we chose to express phenological transitions as day-of-year to maintain consistency with the broader peatland and ecosystem flux literature, where DOY-based metrics are more commonly reported, to enable comparability across years and sites.

In addition, our objective was to characterize the temporal dynamics of ecosystem fluxes and seasonal boundaries rather than to develop a temperature-driven phenological model. While the unusually warm spring of 2024 likely accelerated the growing season onset through enhanced heat accumulation, translating the seasonal metrics into degree days would introduce an additional modeling layer and assumptions regarding temperature thresholds and accumulation periods that are beyond the scope of the present study.

l. 375–378 - Table 1: season date ranges

Given that SOS/EOS dates differ substantially across years (SOS ranges from DOY 80 to 103), it would help readers to include the actual DOY ranges for NGS, EGS, and LGS for each year directly in the table as additional rows, rather than requiring a cross-reference to Sect. 2.4.

Response: Thank you for this suggestion. We added the information of the DOY ranges to Table 2.

l. 420 - Reco–VPD relationship: confounding with SWC/WTD

Caution is warranted when interpreting the relationship between Reco anomalies and VPD, as VPD and SWC/WTD are typically correlated during dry periods, making it difficult to isolate a direct VPD effect on Reco from an indirect moisture-limitation effect. The authors acknowledge the negative SWC–flux correlation in the filtered data (Fig. S4, l. 544–545), but this concern applies equally to the Reco–VPD relationship highlighted here. The reader would benefit from a clearer statement that the VPD effect on Reco cannot be fully decoupled from concurrent soil moisture conditions, which is also the context eventually provided at l. 540.

Response: We employed a predictor selection process to limit the inclusion of confounding variables. This is outlined in the Methods (Section 2.6 Driver Analysis). Working on the daily scale and with anomalies, further seems to reduce the coupling of VPD and SWC. We checked for the correlation of SWC and VPD and the correlation of their anomalies in all seasons. The correlations between the actual measurements of SWC and VPD are 0.12, -0.4 and 0.06 while for the anomalies they are -0.08, -0.21 and -0.03 in the NGS, EGS and LGS respectively.

We have provided this information in the Supplementary (Potential correlation between SWC and VPD during periods of high VPD).

l. 475 - High GPP and footprint heterogeneity

The attribution of high GPP to the dense shrub canopy is plausible, but a footprint-based analysis - comparing flux signatures under wind directions dominated by the denser shrub patches versus the Betula patches or Molinia areas - could provide more direct empirical support. The authors may well be working on this as a separate study, and if so, a brief note to that effect would be appropriate. If not, the claim should be framed as a suggestion rather than a conclusion.

Response: Perhaps our description of the site was misleading. As mentioned in our previous response, the canopy within the footprint is not distinctly divided into separate sections according to vegetation type. It is rather a patchwork structure largely dominated by shrubs with intergrowing grasses and individuals of trees. The dense shrub canopy is therefore not a local aspect of the ecosystem but the dominant characteristic of it overall. One stand of birches in the north-west of the tower is the only region that differs from the rest of the footprint, however the main wind directions are SW and NE, thus the impact of it on measured fluxes is marginal.

We have rephrased our site description in lines 131-132 and 138-139 to make sure this point is clear.

References

Guth, P., Gao, C., and Knorr, K.-H.: Electron Accepting Capacities of a Wide Variety of Peat Materials From Around the Globe Similarly Explain CO2 and CH4 Formation, Glob. Biogeochem. Cycles, 37, e2022GB007459, https://doi.org/10.1029/2022GB007459, 2023.

Lemmens, M., Teickner, H., Lamentowicz, M., Thomas, C. L., Glatzel, S., Gałka, M., Draga, M., and Knorr, K.-H.: Multi-proxy high-resolution geochemical analysis reveals ecological baselines and evaluates potential restoration trajectories in European ombrotrophic peatlands, Ecol. Indic., 183, 114648, https://doi.org/10.1016/j.ecolind.2026.114648, 2026.

McDermitt, D., Burba, G., Xu, L., Anderson, T., Komissarov, A., Riensche, B., Schedlbauer, J., Starr, G., Zona, D., Oechel, W., Oberbauer, S., and Hastings, S.: A new low-power, open-path instrument for measuring methane flux by eddy covariance, Appl. Phys. B, 102, 391–405, https://doi.org/10.1007/s00340-010-4307-0, 2011.

Smith, C., Nicholls, Z. R. J., Armour, K., Collins, W., Forster, P., Meinshausen, M., Palmer, M. D., and Watanabe, M.: The Earth’s Energy Budget, Climate Feedbacks, and Climate Sensitivity Supplementary Material, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, Y., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom, 2021.

Vekuri, H., Tuovinen, J.-P., Kulmala, L., Aurela, M., Thum, T., Liski, J., and Lohila, A.: Improved uncertainty estimates for eddy covariance-based carbon dioxide balances using deep ensembles for gap-filling, Agric. For. Meteorol., 371, 110558, https://doi.org/10.1016/j.agrformet.2025.110558, 2025.