The paper is targeting the important question of CO2-emissions from drained forested peatlands and their potentially higher emissions than undrained forested peatlands. This is important in the context of GHG reporting (UNFCCC). They establish extensive systems for measurements during a 2-year period in a total of 26 sites (n=19 are drained) distributed to all three Baltic countries. They underline the uncertainty of the current default IPCC emission factor for the region and that the transition between temperate and boreal zone (hemiboreal) may be poorly represented by the current IPCC default. Thus, I very much welcome and acknowledge the effort and recognize its importance. Particularly in light of future policy demands on the LULUCF sector emissions/sequestrations and the need to increase Tier levels and enhance documentation.

It is my opinion that the paper still needs some work to be ready for publication. This relates to the clarity in methods, thorough discussion of uncertainties including reference to magnitudes and drivers found in other studies (Rtot, Rhet), the use of extensive soil chemical data and on the text and priorities of both introduction and discussion. See specific comments below.

Thank you for the positive view, the very constructive comments, and the in-depth review of our article! Your comments are well-received and will greatly help us improve the article.

Specific comments:

The introduction largely refers to IPCC guidelines and national scale emission estimates based on UNFCCC (national submissions). I miss an effort to link the study to existing research literature on what the important drivers have been found to be, if emissions are mainly climate-driven or if they have also been found to be driven by vegetation, history, geology/landscape etc. It is stated that earlier results are "inconclusive" but you should at least mention what other studies have expected to find and what was concluded. A paragraph on how forested peatlands in the Baltic region may differ from those used in the IPCC default EF as well as the potential methodologies would be interesting in this context. The study focuses on nutrient rich sites. While the study has an apparent aim to contribute to higher certainty for the UNFCCC reporting for the Baltic countries, it is not shown to what extent the selected sites are area-representative for the drained peatlands for which these countries need to report.

We agree that more strongly addressing the scientific relevance of the study, including the current scientific understanding and knowledge gaps, can be beneficial. We will consider how to incorporate all the listed aspects in the Introduction.The availability of previous research results on the key drivers of organic soil CO_2 emissions allows for their association with climate, land use history, and soil nutrient status, provided that the amount of studies in the respective climate zone permits it (Hiraishi et al. 2013). While the IPCC (2014) evaluated soil nutrient in the boreal zone only, the ammount of studies conducted in the last decade allows for further EF stratification by categories of afforested sites and forestry-drained sites, as well as more specific site nutrient status and productivity categories based on timber production potential also in the temperate zone (Jauhiainen et al. 2023).The primitive approach clearly indicates necessity to seek for broader understandement. Boreal vegetation zone has the most extensive and consistent data; however, in the temperate climate zone (including hemiboreal vegetation zone), the amount of research results remains significantly lower, resulting in a greater uncertainty (Jauhiainen et al. 2019, 2023). The contradictions in research results are likely due to this uncertainty, which has prevented a consensus on whether organic soil is a source or sink. It has been observed that forest soils in moist and cool climates do not necessarily lose carbon stock. Studies show that drained organic forest soils C stock is stable or even increasing (Butlers et al. 2022; Minkkinen et al. 2018), particularly in nutrient-poor or moderate nutrient-rich soil (Lazdiņš et al. 2024). Nonetheless, there are observations that a higher risk of soil C loss can be observed in mostly nutrient-rich soils, often showing a C loss (Lazdiņš et al. 2024). However, observations indicate that C stock of such soils can also be preserved (Butlers et al. 2022). The available data is still insufficient to confidently identify the key influencing factors contributing to the inconclusiveness of these results. This is largely due to the specific local conditions of the study sites, such as meteorological conditions, water table level, vegetation, and soil properties. Quantifying the impact of the variability of these conditions on soil carbon balance remains complex.

An emission increase gradient from Nordic countries to Central Europe can be observed based on earlier studies. This climatic gradient also explains the differences in peatlands in the Baltic states. Vegetation, soil properties, temperature, and groundwater level dynamics are all influenced by climate, which is why the peatlands in the Baltic states differ from those characterized by the IPCC default emission factor. The IPCC factor represents geographic regions with warmer and drier climates.

We will also deal with the site representativeness issue in the Material and Methods section, and briefly return to it in Discussion for improved transparency. We focused on nutrient-rich soils because they are associated with higher emissions; it seems rational to begin improving the accuracy of such soil emissions assessments. By applying the obtained knowledge to all areas, we can avoid underestimating emissions while later working on understanding nutrient-poor soils.

Methods should include the history of the sites studied (I don't find this), particularly when they were drained (and perhaps drainage channels were maintained over time as a typical management activity through time), their LU/area characteristics before planting (if they were planted). What does one know in terms of expecting that these forested peatlands were similar to the undrained forested peatland that are included in the study? The undrained sites in general show a higher tree basal area than the drained sites – do the undrained sites represent sites that would have been selected to be drained historically? In the paper you mention several places the "effect" of drainage. I claim that you are not measuring an effect of drainage, but you are comparing (contrast) two types of forest with apparent different management over time – and most likely the drainage happened long ago.

Thank you for your detailed observations. We agree that information about historical land use would be beneficial for transparent scientific reporting and facilitating interpretation. Unfortunately, historical land use is often poorly documented or not documented at all in the Baltic states. While we intended to include such information, it is unfortunately unavailable. We can, however, provide a general description of typical land use history in this region, and include the best possible evaluation of what has been done previously at our study sites. Water table level can be elevated in drained areas and lowered in undrained areas. To characterise the functioning of drainage system and soil moisture conditions we aimed to describe our observations regarding water table levels in detail. This will help indicate whether existing drainage systems are functioning and what the water table characteristics are. We observed that in undrained areas, the average measured water table level was above -30 cm. However, the drained study sites showed a wide gradient of average water table levels, ranging from approximately -30 cm to more than -100 cm. Extensive drainage was conducted around the mid-20th century. The drainage was typically not performed for afforestation but to improve tree growth conditions in already forested areas. This means that forest development in all areas (both undrained and drained) occurred naturally as a natural succession initially. Due to the aforementioned reasons: these undrained sites included in the study can be expected to be similar to histroical forested peatlands; it is unlikely that the currently undrained areas were historically drained; and undrained sites represent sites that would have been selected to be drained as the water level was consistently elevated.

Indeed, the lower end of the basal area range is higher in undrained sites, which coincides with similar observations in forest stand ages. Such observation arises due to sites selected and does not imply that basal areas of undrained sites tend to be higher in undrained sites. We will consider how to concisely include this descriptive information in the article.

We agree that the term "effect" can be reconsidered, and comparisons between undrained and drained sites should be interpreted carefully here, given that the drainage occurred long ago. We can revise the phrasing to clearly indicate that we are comparing the two groups of sites, and that at best, we may be evaluating the long-term effect of drainage based on assumed similarity before drainage.

Also, sites likely do vary a little in ground vegetation composition – it would be timely to have a clear description of ground vegetation as you use only some of the vegetation components in the balance calculations.

That is correct, ground vegetation is typically different in drained and undrained areas, and there is, naturally, at least some variation also within the site groups. We have conducted vegetation surveys at sites in Latvia, so we can provide at least descriptive information to illustrate the vegetation present in these areas.

Method description of respiration (total, heterotrophic), litter input fluxes and C balances (forest floor, soil) need to be supplemented by a figure with the fluxes that are measured and estimated and how they are combined to calculate the soil and forest floor balances. I believe this will make the methods much more clear as well as shorten the text.

There is such a figure already elaborated, we will consider if it helps the readers, and check the potential to shorten the text while improving the clarity of method description using the figure.

You observe that measured Rhet results seem unreliable, very high compared to Rtot and with a poor correlation to temperature $r2 < 0.3$) relative to Rtot (r2 ca. 0.7-0.8). In some context (fx. line 462) you mention they are found to be in error. In other context you claim that they are unlikely to be subject to measurement errors but mention their likely influence from decomposing roots (line 499), or the lack of temperature and moisture measurements that reflect the actual measurement position (poor correlation to temperature, unknown potential effects of moisture fluctuations). As far as I can see you i) describe the field measurement methods for Rhet in detail, ii) discard the results (line 214) for use in the C balances, iii) do not present them in the Results but iv) refer to them with correlations with Corg%, C:N, porosity or BD (line 436 and onward…however, it is unclear to me if your reference to Rhet here is to the measured Rhet or the Rhet used in the C balance calculations (eq 2)). I realize you wish to present openly to the reader what you have done and which problems you encountered. I feel the balance in the paper of this challenge is wrong. I think what I would do would be: include the field method description of Rhet in the main text if your results from these measurements are still helping you in your research aims. If not – I would move most of it to the supplement.

I would use more effort when selecting the empirical relationship by selecting (reviewing) more than one. The chosen one is from boreal forest (you claim in the intro that hemiboreal EF is likely not represented by temperate EF, back up your choice of a Rhet estimation regression from a boreal study). If I understand correctly that you refer to measured Rhet in the discussion (correlations with Corg%, C:N, porosity or BD) I would like to know why you believe this is relevant given the likely effects of decomposing roots on measured Rhet. Given the clear effect that your decision has on not using the measured Rhet in C-balances I would like to see in the main text a figure with magnitudes and correlations for Rhet and Rtot (clarifying to the reader) and an opportunity to clearly state what you use the measured Rhet for and what not. What are the correlations to temperature and the magnitudes one can expect? And please use different abbreviations for the measured Rhet and the estimated Rhet from eq. 2. An uncertainty discussion should include the uncertainties inherent in the choice of using the eq2 (alternative Rhet).

Thank you for this observation, and the very helpful comments on how to improve the clarity and motivation of our choices. We now clearly see that there remained some confusion in the text relative to the challenges that we faced with measured Rhet, and the need to improve the clarity of presenting these findings. We will gladly follow the suggestions provided. You are correct in understanding that the measured Rhet results were not used in the C balance calculation, but they were used to describe specific relationships between emissions and soil quality indicators. We have outlined this explanation in L437-438, but it is evident that we should also include it in the sections where we describe Rhet (measurements rather than estimations) and their influencing factors.

We can improve the clarity in the article that these influencing factors were more closely correlated with the measured Rhet than with Rtot. Although the measured Rhet results were not otherwise utilized, we aimed to highlight potential uses of soil properties for characterizing soil emissions.

Since the direct Rhet measurements were not used in quantifying the C balance, we agree that some of this information can be moved to the supplementary material. We believe it is essential to inform the reader about this critical Rhet measurement aspect of our study by providing a brief outline in the main text to indicate that these observations were made, but the results were not used in the C balance calculation due to significant issues that justified this decision. A detailed explanation of how the measurements were taken and the associated problems can be moved to the supplementary material. This approach will help avoid overloading the main article while still informing the reader about these important observations.

Thank you for the suggestion to use a different term for the calculated Rhet. We will implement this change. We can indeed also add a figure with the magnitudes and correlations for Rhet and Rtot, and dedicate a subsection in the results for a concise Rhet description, including the current quantitative information from the discussion, while moving the rest to the supplementary materials. The Rhet overestimation compared to the Rtot data can be presented, e.g., in a boxplot format. We will also include a note about the quality issues of the Rhet measurements. We struggled with how to present the measured Rhet data in a meaningful way that would both add critical information and transparency in the paper and support future studies on these issues. We believe that the revision will lead to fulfilling both these aims.

Both results and discussion use considerable space on describing observed effects/relationships between fluxes and soil nutrient characteristics. I miss a much more clear direction on these tests and on the discussion of their results and this direction should be set in the introduction, preferably as specific research questions and/or hypotheses.

We will consider how to improve the presentation of these results. It is unlikely that we will formulate a specific hypothesis regarding soil quality, as we performed these analyses to characterize the study site conditions, while evaluating soil quality as such was not our aim. The opportunity to compare these results with emissions adds value as an attempt to find factors correlating with the fluxes and explaining variation among sites. We could describe this as our intention, to collect data to evaluate the potential use of such data as predictors for soil emissions in the future. Specifically, in this case, to assess the relationships between soil quality and emissions.

It is clear that studies of this scale will not yield specific results applicable in GHG reporting on their own. Even if practical applications were identified, they would not be feasible when detailed soil quality maps do not exist. However, they might exist in the future. Therefore, this contribution is an ongoing effort to gather data from study to study, describing what is observed in smaller studies, and later working towards more concrete solutions by combining the results of this and other studies. We could mention this in the introduction to explain our interest presenting these results, e.g., in the new section suggested for evaluating what is known about the drivers of the fluxes.

Application of statistical tests (methods) are not clearly described in terms of testing expected biological relationships; every test should be used for a clear purpose. Given the few sites some of the statistical methods rely on very few observations per strata. I think it would be better to limit analyses to the mixed model analyses. I find that the PCA analyses are not utilized to their potential – fx. one could use the PC vectors as explanatory variables and – if they express environmental variability that is possible to interpret in a meaningful way – they may help to find a pattern in how the many measured variables influence/drive emissions (example: Callesen et al. 2006. Growth of Beech, Oak, and Four Conifer Species Along a Soil Fertility Gradient. Baltic Forestry Vol. 12, No. 1 (22)).

We will revise the description of the statistical analyses, adding the specific purpose of each test. We agree that the observation count is indeed small in some cases; however, we used different approaches that provided confluences towards building observations and conclusions. Therefore, we considered it valuable to show that different methods provide the same indications of relationships. However, we will reevaluate this carefully, and may move some confirmative tests and their results to supplementary material.

We used PCA in primarily to visualize the covariation in the data, and to evaluate whether there were any clear patterns of country or dominant tree species affecting soil emissions or other quantitative characteristics of the study sites. One of our aims was to evaluate if dominant species or country should be used to stratify emission factors. Ultimately, the confluences of results from various statistical methods led to the conclusion that it would not be scientifically justified, and a single emission factor should be used for all drained or undrained soils. We agree that PCA could be used further for evaluating the relationships between the variables. We will carefully consider the suggestion in the context of this article. Thank you for providing the literature.

The discussion starts by targeting the errors observed in one of the flux-methods (unclear if Rhet is the measured one or the one from eq. 2). Rather, I believe the discussion should start by referring to the results actually obtained in the study on the outcomes you have targeted in your study aim (probably the balance rather than any specific flux?). And target uncertainties on specific fluxes (parts of the balance) in later sections. As an example, in the very last sentence of the section on soil heterotrophic respiration interpretation you state that roots cut in the process of installation was most likely the major reason for your errors…if this is the most important contribution then you should start the section on soil heterotrophic respiration error-discussions with this. Also, I would expect that you would find several studies who have found a similar challenge and it would be timely to refer to such.

We agree that the logical flow can be improved here. Indeed, there are previous studies discussing trenching issues, and we will seek to include some references to show that similar discussions have been raised before. As previously agreed, based on your valuable suggestions, discussions related to directly measured Rhet will likely be moved to the supplementary materials. We agree that the article should focus on data directly used in C balance estimation. However, since the Rhet-related findings are likely valuable for some of the readers, we will include in the article the reasoning for data exclusions, with a note indicating that a more in-depth discussion on this issue can be found in the supplementary materials.

Thank you for providing many suggestions for technical corrections. We genuinely appreciate the time you have dedicated to helping make our article as readable and accessible as possible for the reader. We will address all of these suggestions.