Referee comments on "Water table driven greenhouse gas emission estimate guides peatland restoration at national scale" – bg-2023-23

Within the article the authors aim to calculate national peatland greenhouse gas (GHG) emissions for Denmark and potential rewetting emission reductions following various rewetting strategies. The approach consists of three steps: (1) Modelling average water table depth (WTD) statistics for organic soils, (2) fitting a relation between WTD and GHG observations and (3) estimating total GHG emissions and (4) exploring emissions reduction pathways for Denmark.

The authors bridge multiple research disciplines, outcomes and approaches and present a clear and valuable narrative. The results could directly support peatland management decision-making processes aiming to minimize the contribution of the Danish peatlands to climate change. As the article might impact peatland management on a national scale it is important to address shortcomings and discuss results with nuance, which is something that the authors clearly do. However, I have several additional concerns, which I listed below. I recommend that these concerns are addressed before publication of the manuscript.

Major concerns:

Long term average summer WTD

The authors used WTD data to train the machine learning model for predicting average summer WTD for the complete Danish lowland/peatland areas. More or less half of this dataset originates from median well measurements. Besides, almost half of this WTD dataset originates from soil auger campaigns that consisted of one or two auger measurements. It is not clear from the text what methodology was used for determining the auger-campaign-WTD; (1) were oxidation marks used that reflect historic WTD characteristics or (2) was the WTD measured directly? In the latter case, summer WTD may fluctuate substantially and an auger WTD measurement is highly dependent on the timing measurement, and therefore the question remains if it is really valid to include this data within the research. To demonstrate the accuracy or uncertainty that is introduced, I suggest that the authors compare summer WTD averages of wells positioned nearby auger measurements.

Furthermore, it is unclear how many wells were actually positioned within in peat soils (OC > 12%). Are the well measurements representative enough for peatland WTD's, or mainly located within the lowland valleys?

The model was assigned to predict a draining effect of ditches by setting draining dummy points, which were found to function successfully within the SHAP analysis (Line 210). However, the authors did not discuss the spatial occurrence of ditches and explain why a ditch WTD of 1.21 (Line 101) would be a valid assumption. Are ditches always dry and thus draining the soils? Or do they also supply water during dry conditions? Is this more or less the same throughout the country?

Model performance and ecosystem carbon balance

A mean error of the trained model of -0.20 m for WTD estimates between 0-0.5 m (Table 3) seems quite high, given the high sensitivity of CO_2 emissions predicted for this WTD range (Fig. 7a) and the high percentual area with WTD within this range (Line 257). This seems problematic within further analysis. The authors are open about these high uncertainties (i.e. Line 275) and proceed the analysis on a national scale. The mean error of -0.20 m means that the modelled WTD is more shallow as compared to the observed WTD. Would this mean that national estimates of total emissions are highly underestimated, and that emission reductions are underestimated as well? The MAE of 0.08 m (Line 324, Table 3) is based on validation including dummy points. As these points represent a large portion (1/3) of the dataset the model performance statistic seems biased. Therefore, I would suggest to exclude dummy points by default in Table 3 (and refer to these in text) and report performance statistics including dummy points between brackets instead.

It is exciting that the authors found very similar parameterization for the adopted Gompertz function used in Tiemeyer et al. (2020) (Fig. 7a) (the function to estimate NECB based on WTD). However, the amount of net ecosystem carbon balance (NECB) datapoints in the sensitive WTD range is quite low. As the authors already address, many linear relations have been fitted between WTD and NECB within literature. Except for one observation with a WTD > 1 m, a linear relation would not misfit the data in this research. Such a less steep linear relation would distribute the weight away from the currently very sensitive (and steep) 0- 0.5 m WTD-region and result in carbon balance estimations (and emission reduction estimations) that are less vulnerable to apparently quite substantial errors in WTD estimations (especially for 0-0.5 m WTD). I suggest that the authors conduct two sensitivity analyses: (1) the effect of water table bias on GHG emissions, and (2) the effect of Gompertz function on the GHG emissions by stretching, squeezing and using a linear relation.

Covariate analysis

In the results the authors discuss the impact of land surface temperature (LST) independently of ranked wetness. However, to what extent are these two covariates independent? A crop with deep WTD generally evaporates a higher amount of water, leading to more latent cooling. I would suggest to check the independency of the covariates. Furthermore, the small SHAP values for LST might not justify the high amount of attention in the results section (i.e. Fig. 3 and 5, from Line 228 onwards). In addition, the lower LST in areas of high saturation might not only be caused by more evaporative cooling, but maybe also due to induced heat transport towards deeper soil layers (wet peat soils conduct heat much better).

Total GHG emissions and reduction pathways

Given the urgency of climate action, I would suggest to include 25 year global warming potential to calculate total GHG budgets complementary to the 100 year timeframe that is currently applied in Fig. 8 and 9. In this case, the research would address both the short- and long-term effects of rewetting strategies.

Other remarks:

- I would suggest aligning the x-axis of Fig. 4 and the legend of Fig. 5 in order to facilitate a constant direction of thinking (i.e. negative SHAP values are related to shallow WTD).
- Figure 6 gives insight in the general performance of the GBDT models. I suggest to additionally show how well the models perform within the most important and sensitive WTD range of 0-1.0 m (maybe with grids of 0.1 m or 0.05 m).

- It is unfortunate that peat depth could not be taken into account within the analysis (Line 355, 376). Including such data would likely result in a lower prediction of emissions, which could be stated more directly in the text.