

## **“Water table driven greenhouse gas emission estimate guides peatland restoration at national scale” – bg-2023-23**

[Reviewer comments in *italic blue font*; author replies normal font]

*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.*

**Reply:** We would like to thank the reviewer for their very fair, thoughtful, and constructive review of our submission. We are glad that the novelty and relevance has been generally acknowledged by the reviewer. The reviewer states several very relevant points. Therefore, we will gladly pick them up during the revision to enrich the discussion of the manuscript. We totally agree, this article might impact the peatland management at national scale and uncertainties and shortcomings need to be addressed and discussed. Please find our point by point replies and concrete plans for the revision below.

*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.*

**Reply:** The soil auger campaigns were conducted with the purpose to both, collecting soil samples for subsequent laboratory analysis as well as allowing for a WTD measurement. The WTD was measured directly in a 120 cm deep borehole left by the auger, the borehole was open approximately 30 minutes from augering to measurement. We agree that a single direct WTD measurement is associated with uncertainties due to groundwater fluctuations. However, the soil

auger based WTD data is the more or less the prime datasource we have for organic rich peatlands. Wells are often placed in the fringe areas or at deeper depths. We agree, a comparison of soil auger based WTD data and well based WTD data is currently missing.

**Plan for revision:** We will compare auger based WTD data with well based WTD data that are in close vicinity of each other. Furthermore, we will highlight the uncertainties of using temporal snapshots of WTD for building a steady-state WTD model. The analysis will feed into an updated data description as well as discussion.

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

**Reply:** We agree that information is very relevant and currently not provided in the manuscript. However, we are not very concerned that wells from the fringe areas in the lowlands will affect the WTD in peat soils (OC > 12%). First of all, wells were selected based on a 200m buffer analysis as well as a well filter depth constraint of 5 m was applied. Second, the covariates will provide the ML model with information that will distinguish the WTD training data with respect to how representative they are for OC > 12%.

**Plan for revision:** In the data description we will provide an overview of the spatial distribution of wells with respect to OC > 12 %, lowland & river valley and 200 m buffer zone.

*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?*

**Reply:** First of all, a WTD of 1.21 m was used for the dummy points in ditches as well as for drained forests to provide the model with information on deep WTD under those settings. Given the applied WTD transformation, WTD variability below 1 m depth is very insignificant. A value of 1.21 m corresponds to the dry borehole WTD measurements from the soil auger campaigns. Second, the spatial occurrence was derived from a very detailed national surface water dataset. From that dataset, all small to large streams were removed. For this task we used all streams that are currently represented in the national water resources model (DK-Model). Lastly, yes, it is our assumption that a WTD of 1 m or more right next to ditches is realistic for Danish conditions. In most cases the ditches drain the surrounding land, also in summertime, and ditches in Denmark are generally not used to supply water to fields during dry periods, but entirely to drain water away from fields. We also want to highlight that the applied ML model does not directly simulate whether the ditches drain or supply water from the surrounding fields. In fact, both situations are possible, depending on the training data provided. In case training data indicates a WTD > 1 m in vicinity of ditches, the ditches may be supplying water. Vice versa, shallow WTD training data in vicinity of ditches will likely result in a situation where the ditches drain the surrounding fields.

**Plan for revision:** Since it is not explicitly modelled whether ditches drain or supply water, we cannot conduct a quantitative analysis. However, we will revisit the presentation of ditches in the data section as well as add some details to the discussion.

### *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<sub>2</sub> 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.*

**Reply:** We totally agree that the bias is substantial given the implications of small uncertainties when applying the Gompertz function for CO<sub>2</sub> upscaling. We relate the bias to the fact that the training data are associated with a considerable uncertainty. This uncertainty mainly originates from the fact that we use observations from temporal snapshots to model a steady state mean summer WTD. The bias was calculated as obs minus sim and therefore the bias of -0.2 m in the uppermost depth interval means that the model predicts a WTD that is generally deeper than the observations. This has implications for the GHG emissions. For CO<sub>2</sub>, the baseline emissions can potentially be overestimated which in turn means the reduction potential may have a tendency to be overestimated.

We have investigated to build a bias correction model on top of the ML model. However, this proved to be unfeasible since there is not correlation between simulated WTD and bias, which is not surprisingly, because if there was a correlation, the ML model would have picked that up.

Table 3 reports the performance for both all training data and well&auger data only (in brackets). We believe it is fair to report the performance including dummy points, because this is what the ML model trained/optimized against. However, we agree that the performance of well&auger data is the more relevant performance quantification for the given application.

**Plan for revision:** In table 3, the performance based on all data will be stated in brackets instead of the performance based on well&auger auger data. In the text, we will highlight that the latter is the more relevant performance quantification for the given application. Uncertainties of the training data that cause the bias will be further emphasized in the discussion.

*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.*

**Reply:** We do not have much to add to this well formulated and spot-on comment. Reviewer #2 also pointed out the need to consider a linear model. We agree that the two suggestions for sensitivity runs will be insightful. Due to the WTD transformation, the model's reliability for WTD greater than 1 m is somewhat limited. This limitation works very fine with the Gompertz model that flattens around 0.4 m, but it may have some implications to the GHG emissions when applying a linear model.

**Plan for revision:** In the revised manuscript we plan to add a new discussion section where we will explore uncertainties and their effect on the overall GHG estimation as well as for the rewetting potential. We will stick to the Gompertz model for presenting the key results. We will 1) explore the possibilities to fit a linear model to the Danish GHG - WTD data and if it is suitable for GHG upscaling and 2) investigate ways to bias correct the GHG sensitive WTD interval.

#### *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).*

**Reply:** We agree, LST and ranked wetness are to some degree related. However, land cover heterogeneity will overrule their relationship. For example, forests have a generally lower LST due to their higher transpiration, but in most cases a low wetness rank. In figure 3, there exists a large, connected patch of low LST in the northern part of the zoom, which shows a low wetness rank, i.e. dry conditions. This patch is forested and here the linkages between LST and ranked wetness are different as over grass or cropland, where we generally see low LST collocated with high wetness ranks, i.e. wet conditions. The ranked wetness is based on the registered crop type by the farmers, which has been analyzed over a period of 5 years to aggregate the data to a single wetness rank. Uncertainties lie in the crop registration as well as in the simplification to derive a ranked wetness. The wetness rank is constant for an entire field whereas the LST shows variability from cell to cell, which can inform the model of subfield WTD variability, which becomes relevant for larger fields. Based on these arguments we believe that both, LST and wetness rank are of high relevance in the applied ML model.

Yes, the SHAP values for LST may not be the most dominant ones, but based on the top six covariates presented in figure 4, it is among the ones that has a high spatial variability, potentially revealing sub field variability, which should not be forgotten when analyzing the results from the SHAP analysis.

With respect to the evaporative cooling, we interpret the data in the way that a shallow WTD sustains a high water supply for plants for transpiration. We have not considered variations in heat conductance of dry and wet soils.

**Plan for revision:** We will emphasize that evaporative cooling is not the only process that can cause low LST. Moreover, the reasoning of using both, LST and ranked wetness in the model building will be discussed. The importance of both, SHAP values and spatial variability will be addressed in the revised manuscript.

#### *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.*

**Reply:** We agree.

**Plan for revision:** The implications of 100- and 25-year global warming potentials will be presented in the results and in the discussion section.

#### *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).*

**Reply:** We agree, the legends can be aligned better.

**Plan for revision:** Legends in figure 5 and figure 6 will be adjusted so that negative SHAP values are on the left and positive SHAP values in the right.

*- 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).*

**Reply:** We agree.

**Plan for revision:** The figure will be adjusted in a way that the top panel shows all WTD data and the bottom panel will focus on well&auger observations between 0 – 1.0 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.*

**Reply:** We agree, such a dataset is crucial since peat depth, in combination with the WTD, marks the lower boundary of the unsaturated peat. In Denmark, mapping of peat depth is currently undertaken, and the potential of using the WTD map proposed herein in combination with the peat depth map for an improved national GHG inventory from 2024 and onwards is currently investigated.

**Plan for revision:** The importance om peat depth will be further underlined in the revised discussion.