The main contribution of this work is that the authors showed that geomorphological changes drive changes in salinization and reaction potential in beach aquifers using general 2-D numerical models. They showed that there is greater mixing of fresh and saltwater due to geomorphic change and this leads to greater reaction potential. They also demonstrated that hydraulic conductivity, dispersivity, and storm floods control the extent of the mixing zone and salinization under changing geomorphic conditions.

The authors provide an important contribution by incorporating geomorphic change into coastal groundwater models. The paper is technically sound and well written, but it could benefit from a few clarifications in the methods section. Therefore, I recommend minor revisions in accordance with the comments below.

Dear Reviewer #2 we thank you for reviewing our manuscript and appreciate the overall positive assessment. We believe that your comments and suggestions help to significantly improve our paper. We address your comments point by point as indicated by author comments (AC) in blue, planned changes to the manuscript are indicated by “speech mark”.

Kind regards, Rena Meyer

Minor Comments

Line 96 What is the datum that MHWL and MLWL are referenced to?

AC:. Mean high water line (MHWL) and mean low water line (MLWL) relate to the 10-years mean from 2010 to 2020, measured at the tidal station Wangerooge North (https://www.pegelonline.wsv.de/gast/stammdaten? pegelnr=9420030) and referenced to the normal sea level (Normalhöhennull, NHN) reference system. We will add this information to the methodology section.

“The mean high water line (MHWL), referring to the ten-year mean from 2010 to 2020 (Pegelonline, 2022), is located at 1.35masl and the mean low water line (MLWL) is at -1.35masl, referenced to normal sea level (NHN).”

Line 122 What is the specified flux that was used?

AC: We will add the specific flux of 0.5 m3/d/m to the respective sentence.

“Freshwater (salt concentration = 0 g/l) entering from the islands’ inland (freshwater lens) along the vertical Southern boundary was prescribed using a specified flux of 0.5 m3/per day per meter shoreline as estimated by Beck et al., (2017) and was uniformly distributed across the cells of the first column (Fig. 1).”

Line 123 What is the flux for the meteoric recharge boundary?

AC: The meteoric groundwater recharge at the upper beach is 400 mm/a. We will add the value to the methodology section.

"Meteoric groundwater recharge of 400 mm/a was applied at the upper beach above the MHWL."

Lines 124 Why was a general head boundary used?

AC: The general head boundary was used to avoid numerical problems that may occur when a constant head boundary is used in a highly transient system where the assignment of boundary condition would change with each time step (due to the temporal changes in beach morphology). To mimic a constant head boundary we, however, chose a very high conductance of 1000m^2/day.

We will add the reasoning to the methods section.

"A general head boundary (GHB) with a high conductance of 1000 m^2/d was specified along the seaside and intertidal zone (Fig. 1) to ensure good aquifer connection. The hydraulic head of the GHB boundary was set to 0 below the MLWL. This approach helped avoid numerical issues that could arise from using a constant head boundary in a highly transient system, where boundary conditions might change due to shifts in beach morphology."

Lines 128-130 How was the linear interpolation performed from the lidar scans (i.e. was each point in the domain sampled from the lidar scan or was it sampled at a different resolution), and how was the morphological change represented in the model (i.e. how was the interpolation done between each morphological realization, if erosion occurred was salt mass removed from the model and if deposition occurred was it assumed to be saline or fresh)?

AC: We understand the need to expand our explanation of how the variable beach morphology was derived from the LIDAR profiles. Each meter was sampled from the LIDAR scan. Interpolation between the morphological realizations was linear in daily steps, hence sudden erosion/accretion events are not displayed (therefore no additional salt mass was added or removed), as we outline in the limitations of our study.

We will extend the description of the interpolation of the morphology in the methodology section:

"In the intertidal zone (between the MLWL and the upper beach affected by storm floods) the beach surface was interpolated according to the methodology described by Greskowiak and Massmann (2021). Five cross-shore LIDAR scan profiles (as shown in Fig. 1b, obtained for Feb-July), were sampled at 1m resolution and interpolated to daily increments over a six-month period. The topography was then varied in daily increments for six month, and reversed for the second half of the year. The resulting annual topography was applied recursively over the 20a simulation period. Daily topography time series were then used to calculate the hydraulic heads using the tide-average head approach, which were assigned to the sea boundary in the intertidal zone (Fig. 1a, grey box)."

Line 133 Why was the simulated salinity assigned to the water discharging across the ocean boundary?

AC: With this we consider a so-called non-dispersive flux boundary condition. It's a 3rd-type transport boundary condition, where solute mass is transported out of the model domain via the product of discharging water and its respective solute concentration.

“In the intertidal zone at the GHB, saltwater inflow and outflow were modelled using non-dispersive flux boundaries. A solute concentration of 35 g/L was assigned to the inflowing saline water and the simulated concentration to the outflowing water. This third type of transport boundary condition transports solute mass out of the model domain based on the product of the discharging water and its respective solute concentration.”

Line 134 Was the simulation run to a steady-state salinity distribution before running the transient model?

AC: We did not run a steady-state salinity in our transient models. However, after 10 years of simulation the initial distribution of salinities (as TDS) and reactant were insignificant. Therefore, we chose to only use the last ten years of simulation period for our analysis to make sure that the initial conditions do not influence on our results and conclusion.

We will extent the explanation in the methodology section accordingly:

“The decision to assess SD TDS and RP_c based on the final 10 years of the simulation period was taken in order to circumvent the potential influence of the initial distribution of TDS, R_s and R_f. Hence, the first 10a of the simulation period serve as model spin-up.”

153-154 It seems like the initial distribution of the reactants is unrealistic given that there was no spin up to a steady-state salinity distribution. This should be mentioned as a limitation.

AC: We understand the necessity for clarification of the influence of the initial distribution of the reactants. Reviewer #1 made a similar comment. We repeat our AC to Reviewer #1 here for an easier reading:
The reviewer is right that the initial distribution of Rf and Rs could have had an effect on the final patterns of their mixing product (Mp). In order to resolve this issue, we have now subtracted the accumulated mixing concentration after 10a (when the initial concentrations of TDS did not influenced final salinity patterns any longer) from the final (20a) simulation concentration. The general mixing patterns look very similar to the previous results, as presented in the new Figure 5 (see below). However, slight changes in normalized concentrations are visible (Fig. 5). Therefore, we decided to follow the reviewers recommendations and update Figure 5 and also re-calculate the cluster analysis (Fig. 6) which also shows slight differences. Moreover we will modify the corresponding text about the cluster analysis. However, the overall changes were only minor and the general picture and conclusion driven are still valid.
The methodology section will be extended by:

“The decision to assess SD TDS and RP_c based on the final 10 years of the simulation period was taken in order to circumvent the potential influence of the initial distribution of TDS, R_s and R_f. Hence, the first 10a of the simulation period serve as model spin-up.”

Update to Figure 5:

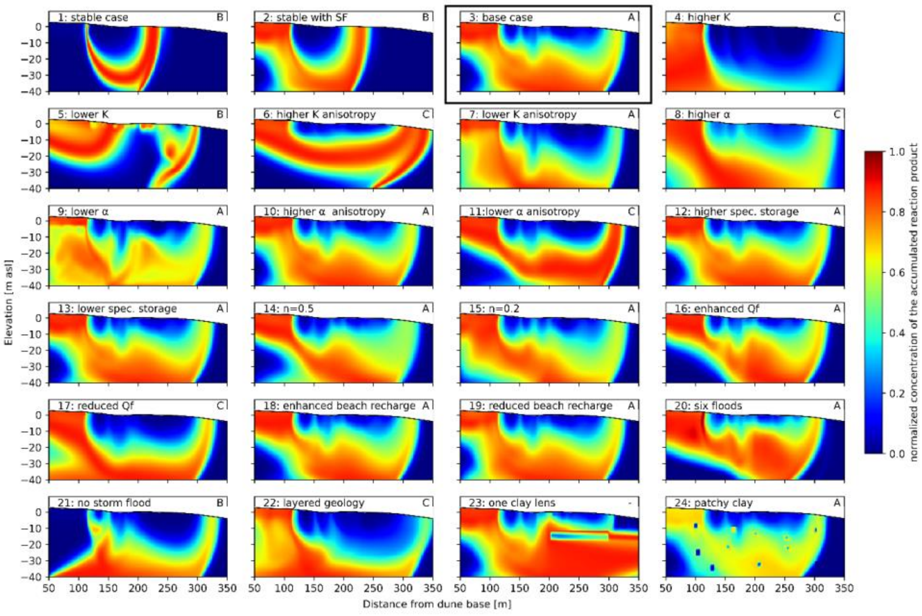

Figure 5: Normalized concentration of the accumulated reaction product (Mp) indicating the reaction potential for all 24 model cases after 20 years of simulation time. The base case is framed in black. The letter in the upper right corner refers to the cluster group (cf. Tab. 1, Fig. 6).

Update to Figure 6:

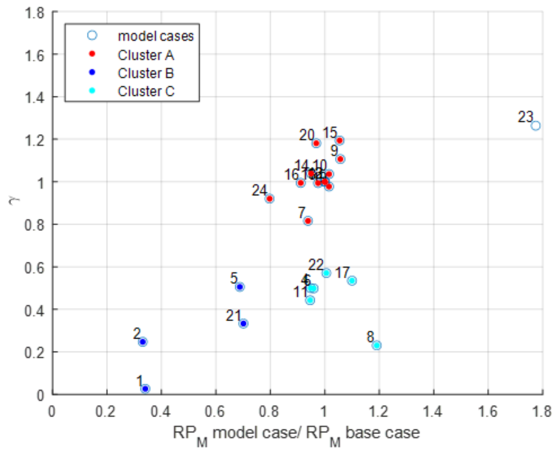

Figure 6: Cluster analysis (k-means,3) of all model cases based on the models' variation in salinity (γ) and the sum of reaction product of each model (RP_M). Individual model values were normalized to the base case (located at 1,1 in the diagram).

The section about the cluster analysis in the result section will be modified to:

"A k-means analysis based on γ and RP_M of each model case normalized to the same statistics of the base case, resulted in the clustering of the model cases into three main groups (Fig. 6, Table 1). Cluster A (red circles) had a γ (+/- 20%) and RP_M (+/- 20%) similar to the base case (located at the coordinates 1,1 in the diagram in Fig. 6). Cluster B included the less dynamic and more stable cases with a lower γ, reduced by 40-98%, and lower RP_M, reduced by 30-70%, compared to the base case. Cluster B contained the cases with either changing topography only (and no storm floods, case 21) or only storm floods (and no changing topography, case 2) or neither (case 1) and the low K case (case 5). Cluster C was characterized by a lower γ reduced by 40-80%, while keeping a RP_M (+/- 20%) similar to the base case. One outlier, case 23 with one clay layer, showed a significantly higher γ and higher RP_M compared to the base case."

Line 180 I think it would help the reader to describe the difference between the base case and the stable case before discussing the results.

AC: We will add a few sentence to describe the difference between base case and stable case at the beginning of the results section.

"The model proposed by Greskowiak and Massmann (2021) which incorporates a transient beach morphology and three storm floods was employed as *base case* in the present study. Boundary conditions and aquifer parameters were then varied to identify their individual influence on flow and salt transport as well as mixing-controlled reaction potential. In the *stable case* the average beach topography from the base case and no storm floods were considered."

Line 193-195 I am confused by Figure 5 case 3. How is it normalized? If it is normalized to the base case then there should be no variation in the base case figure. A description of the normalization should be included in the methods.

AC: We understand the necessity for clarification of the normalization procedure that yields the concentration illustrated in Figure 5. The concentrations of the accumulated reactant are normalized to the maximum reactant concentration (M_p) across all model versions. We will provide a more comprehensive description of the normalisation in the methodology section.

"The model results were evaluated according to (1) the flow regime visualized as flow lines (Fig. 2, Fig. 3); (2) the TDS distribution shown as snapshots at the end of the simulation (Fig. 3) as well as the standard deviation of the TDS concentration (SD) in each cell over the last 10a of simulation time period (Fig.. 4); and (3) the reaction potential (RP_c = sum of accumulated mixing products in each cell (Mp_C) over the last 10a of simulation period) normalized to the absolute maximum Mp_C concentration across all model versions (Fig. 5). The decision to assess SD TDS and RP_c based on the final 10a of the simulation period was taken to circumvent the potential influence of the initial distribution of TDS, R_s and R_f. Hence, the first 10a of the simulation period serve as model spin-up."

Technical Corrections

Line 60 led should be lead

AC: Will be corrected as suggested.

Lines 67-69 This sentence does not make sense. Move citation to end of sentence.

AC: The sentence will be modified as suggested:

“Greskowiak et al. (2023) concluded that redox zone dynamics in the STE are strongly affected by beach morphodynamics. While some redox reactions take place in the USPs and storm flood affected area, mixing controlled reactions driven by mixing of two solutes in different end members occur in the fringes of the USP and at the SW interface (Heiss et al., 2017).”

Table 1 Case 2 Description What does SF stand for? I think storm flood but this is not defined anywhere.

AC: We will add the acronym for storm floods (SF) to the text.

“Furthermore, cases with a stable beach morphology with and without storm floods (SF) as well as with three different permeability distributions (Fig. 1, c-f) were tested.”

Line 261 reached should be reach and mowing should be moving

AC: We will change “reached” to “reach” and “mowing” to “moving”.

Line 289 Where is the black cross?

AC: The reviewer is right there is no black cross. We will remove it and modify the caption of Figure 6

“Figure 6: Cluster analysis (k-means,3) of all model cases based on the models’ variation in salinity (γ) and the sum of reaction product of each model (RP_M). Individual model values were normalized to the base case (located at 1,1 in the diagram).”

Figure 6 Some of the points are hard to see. It would be helpful if the scenarios were labeled so that each number could be read.

AC: We understand the need for a better recognition of model cases related to the cluster analysis. Reviewer #1 made a similar comment, so we decided that we will add the Cluster groups (A,B,C) to the respective subfigure in Figures 3 to 5 which are also mentioned in Table 1.

Line 308 lead should be led

AC: We will change as suggested.

Line 322 Remove “related to”

AC: We will remove as suggested.

Line 331 Rephrase “impact significantly on the shape”

AC: We will rephrase as suggested to:

“We found that in addition to morphodynamics and storm floods, the horizontal and vertical hydraulic conductivity, and longitudinal and transverse dispersivity significantly influence the shape, location, stability and extent of the USP, FDT and mixing controlled reactions.”

Line 339 Rephrase “made that simulations ended”

AC: We will rephrase the sentence to:

“Changing the values of hydraulic conductivity or dispersivity and their respective anisotropies caused the model cases to end up in different clusters.”

References:

Greskowiak, J., Massmann, G., 2021. The impact of morphodynamics and storm floods on pore water flow and transport in the subterranean estuary. Hydrol. Process. 35, 1–5. https://doi.org/10.1002/hyp.14050

Greskowiak, J., Seibert, S.L., Post, V.E.A., Massmann, G., 2023. Redox-zoning in high-energy subterranean estuaries as a function of storm floods , temperatures , seasonal groundwater recharge and morphodynamics. Estuar. Coast. Shelf Sci. 290, 108418. https://doi.org/10.1016/j.ecss.2023.108418

Heiss, J.W., Post, V.E.A., Laattoe, T., Russoniello, C.J., Michael, H.A., 2017. Physical Controls on Biogeochemical Processes in Intertidal Zones of Beach Aquifers. Water Resour. Res. 53, 9225–9244. https://doi.org/10.1002/2017WR021110