Response to Reviewer #1

We sincerely thank Reviewer #1 for their insightful and constructive feedback on our manuscript *Development of a wind-based storm surge model for the German Bight*. The comments contributed to improving the clarity and quality of the manuscript. Below, we provide a detailed point-by-point response to the issues raised and propose how we intend to incorporate the suggestions into the revised manuscript.

Eq.(1): Are the mean wind speed (U_k) and mean zonal/meridional winds calculated from all timesteps of the reanalysis or only those related to the skew surges from the corresponding *hwss*? If all timesteps are considered, then the normalized mean wind (uas_k/U_k, vas_k/U_k) is an averaged normalized wind at this grid cell independently on whether it causes storm surge or not and thus effwind is a projection on average wind direction, not the direction favorable for the generation of surges. In this case it is unclear why it is called "effective wind", effective for what? If only winds associated with skew surge higher than a certain threshold are considered to construct the normalized mean, please specify which timesteps exactly were taken (e.g. at the moment of the skew surge or within 12 hours prior or something else)?

Response: We apologize for the unclear explanation regarding the calculation of the effective wind. To clarify, we perform the calculation of the mean wind speed and mean zonal/meridional winds using only the timesteps associated with skew surges that are greater or equal to the *hwss* training threshold. Moreover, we calculate the specific wind direction through a composite analysis, which is done separately for each hour (up to 12 hours) prior to every *hwss* event. We will adjust the description within the paragraph with a clearer explanation.

L145-153: I'm trying to understand the shape of the individual model (for each grid cell, time lag, hwss). From this passage I would assume it looks like log(skew_surge)=a*(ef_wind^2) +b*(ef_wind)+c. If this is not what was used, please reconsider the description. If this is what was used, then (1) just for the sake of terminology, this is not a multiple linear regression as stated in Line145, but rather a simple quadratic regression (2) I find the sentence "In this way, we ensure the effect of the negative sign" misleading here. Firstly, for high skew surges (hwss) effective winds will be positive anyway. Yes, consideration of only positive coefficients may be helpful later when the skew surges for the whole year are reconstructed and negative effective winds are well possible, as explained in Sect. 2.3.2, but to ensure influence of negative effective winds a special treatment is necessary and positive coefficients alone do not ensure the effect of negative sign. I suggest to refer here directly to the Sect. 2.3.2 for more coherent explanation of this constrain for those interested.

Response: Absolutely, the individual model (for each grid cell, time lag and *hwss*) looks like log(skew_surge) = a*(eff_wind^2)+b*(eff_wind)+c. We apologize for inadvertently using the wrong terminology and will change it to simple quadratic regression. We thank the reviewer for pointing this out.
Correct, to ensure the influence of negative effective winds a special treatment is necessary. We will delete the sentence "In this way, we ensure the effect of the negative sign" and refer to Sect. 2.3.2 instead.

General comment: I wonder how sensitive is the model to the selected training dataset? That is, how much the models (and the estimated skew surge) change when different years are excluded, as it has been done during the validation procedure? If, for example, years Y1 and Y2 are excluded to generate Model1, Y1 and Y3 are excluded to generate Model2, would the reconstructions of skew surges for the year Y1 from these two models be identical? More generally, especially if the model to be used for scenarios, does the size or selection of the training dataset matter?

Response: We thank the reviewer for the insightful comment regarding the sensitivity of the model to the selected training dataset. Yes, the size of the training dataset does indeed matter: If the dataset is small and includes many predictors, the model may overfit, capturing noise rather than the underlying patterns. A larger dataset provides more information, which leads to more stable and reliable coefficient estimates. In turn, this helps the model perform better on unseen data by balancing bias and variance effectively.
To answer the specific question, we conducted the proposed experiment using the year 1976. For Model 1, we excluded 1976 and 1977, while for Model 2, we excluded 1976 and 1975. The reconstructed skew surges for the year 1976 from both models were of the same order of magnitude, but not identical, indicating that the model's output can vary depending on the years excluded from the training dataset.
For scenario applications, we will train the model using all available years from 1959 to 2022, ensuring that the model is based on a comprehensive dataset. However, we would like to emphasize that we perform Leave-One-Out Cross-Validation during the validation procedure to ensure that the final model achieves maximum generalization performance and is as representative as possible.

L20: "Coastal protection institutions" - Meant are those organizations who plan and construct protection structures? Is this an established term?

Response: Yes, we will rephrase that term to ensure clarity.

L22: "continuing rise of sea level" -> continuing rise of mean sea level

Response: We will apply the proposed change.

L24: "as sea level pressure" -> as atmospheric sea level pressure

Response: We will apply the proposed change.

L28: "… these events: The storm surges studied include…" -> … these events. The storm surge studies include …

Response: We will apply the proposed change.

L36: The most widely used and reliable method of such translation is a hydrodynamic model. It is clear, that this study is about simple fast methods beyond classical models, still I think the dynamical models should be mentioned somewhere in the text. Maybe within a short explanation why they are not always the best choice and where the alternatives are needed.

Response: We thank the reviewer for the suggestion. We agree that hydrodynamic models are the most widely used and reliable method for translating atmospheric conditions into storm surge estimates. To address the reviewer's comment, we will include a brief explanation in the text outlining why these models are not always the best choice and why simpler alternatives, such as the method proposed in this study, are useful for particular purposes.

L57: "They find that the external surge…" -> They find that thus considered external surge and…

Response: We will apply the proposed change.

L64: "excluded in the model setup" -> excluded from the model setup

Response: We will apply the proposed change.

L274: "the track consistently follows a northwest to southeast orientation" – I presume here the track refers to the path of grid points with maximum R^2 and not the storm track. Maybe choose another word because "track" has a certain connotation. The position of crosses also doesn't help to identify the storm track itself, maybe only hints to it, so the relation to the next sentence about prevailing northern storm tracks is not obvious.

Response: We thank the reviewer for pointing this out. Indeed, the word "track" could be misleading here. We will write "temporal evolution of grid points with maximum R^2" instead.

L352-356: On what dataset the bias correction (quantile mapping) was trained before it was applied to the omitted years? Would be helpful to specify in the text what was the training and what was the validation datasets.

Response: We thank the reviewer for this question. To clarify, the first step involves reconstructing all skew surges from 1959 – 2022 by training on all years except the year to be predicted. Subsequently, we apply the quantile mapping bias correction to all reconstructed skew surges from 1959 to 2022, without the use of cross-validation. We calculate the transfer function by comparing the observed skew surges with the reconstructed skew surges, and we then apply this transfer function to the reconstructed skew surges. We will ensure to include this clarification in the revised manuscript.