Reviewer #2

G1 *This paper applies a storm surge model (with tide) forced with an atmospheric model, which is used with different resolutions. The scope is to assess the impact of different spatial and temporal resolutions on the accuracy of the ocean model results. The paper has good English and is easily readable. However, the purpose of the paper is weak; the findings are based only on one case, and the comparison with the observations is poor. Therefore, I recommend a substantial revision before a new submission, and I had to reject it in the current form.*

Response: We thank Reviewer 2 for their constructive feedback and the opportunity to improve the manuscript. We acknowledge the concerns on the original submission regarding the limited scope of the original study and the need for more extensive validation.

In response, we have significantly expanded the validation effort by incorporating additional sea surface elevation stations from the coastal and offshore North Sea, in addition to those within the estuary, as well as wind observations over both land and sea. These new validation highlight that higher-resolution atmospheric forcing improves wind field accuracy throughout the domain and even more over land and estuary, which, in turn, enhances the simulation of surge in the hydrodynamic results. Notably, we show that finer resolution strengthens the representation of momentum transfer at the sea surface, leading to more realistic surge propagation from the open sea into the estuary. This highlights the importance of capturing both local and remote dynamics—insights we believe are applicable beyond the specific case studied here.

These new validations are discussed in detail, and we have expanded our interpretation of the results to reflect their broader implications.

In light of Reviewer 2's comment regarding the concern on a single case study, and in line with Reviewer 1's suggestion, we have revised the manuscript's title to better reflect its focused scope:

The modified title now reads: *"Assessing the sensitivity of storm surge simulation to the atmospheric forcing resolutions across an estuary-sea continuum"*

As also noted by Reviewer 1, this study is among the first to explore the combined influence of atmospheric forcing resolution—both spatial and temporal—on storm surge dynamics in an estuarine context. While the analysis focuses on a single, well-documented storm and estuarine system, we believe the findings offer valuable insights that can inform future modeling efforts in similar environments.

To better position our work within the existing literature, we have clarified the study's objectives and expanded the background discussion. For instance, we now reference Weaver et al. (2016), who found that spatial resolution has a greater impact than temporal resolution in storm surge simulations for the Indian River Lagoon, and Lorenz and Gräwe (2023), who highlighted limitations of coarse reanalysis datasets such as ERA5 in resolving coastal wind speeds over the Western Baltic Sea. These studies help contextualize our findings and reinforce the broader relevance of resolution choices in coastal modeling.

We hope that these substantial revisions covering validation, context, and interpretation, address Reviewer 2's concerns and demonstrate the scientific contribution of this study to storm surge modeling in estuarine and complex coastal environments.

G1 *About the atmospheric model: it is not clear the increase in the model performances with the resolution. For example, the 5km seems better than the 2km (see the gravity waves in the North),* *but also the correlation. However, to evaluate the wind quality of a model, many stations must be used (not only one) and scatterometer data, if available. Finally, the authors should discuss the computational time with different resolutions, which is not linear with the resolution increase;*

Response: We thank the reviewer for noting this helpful point. We clarify the atmospheric model performance in the revised manuscript. We have added 16 additional weather stations to validate the atmospheric forcing in the Scheldt and North Sea. These additional stations are sourced from the Global Historical Climatology Network hourly (GHCNh) dataset, which compiles observations from fixed land-based stations operated by various meteorological agencies. The locations of the stations used are shown in Figure 2a-b, with blue markers indicating stations near the Scheldt and green markers for those in the North Sea region. This has been clarified in the text by adding details about it in the methods section 2.4.

The updated section 2.4 now reads as: *We use various error metrics to assess the quality of our model results compared to observations from the different stations (see locations in Fig. 1). Let o and m represent the observed and modeled values, respectively, for a time series consisting of N data points. The mean of the observed time series is denoted by $\overline{o}$, and the mean of the modeled time series is denoted by $\overline{m}$. The performance metrics used to evaluate the model are summarized in Table 2. Observations include several types of data: wind speed, atmospheric pressure and water elevation. Water elevation data were sourced from Rijkswaterstaat data, providing continuous 10 min observation data along the Scheldt estuary. Wind speed and atmospheric pressure observation were acquired from from two sources: the Flemish Banks Monitoring Network and located in the Belgian Continental Shelf, Meetnet Vlaamse Banken, with resolution of 10 min and from the Global Historical Climatology Network hourly (GHCNh) dataset maintained by NOAA. The GHCNh includes observations from fixed land-based stations operated by various meteorological agencies (Fig. 2). These datasets were used to validate the atmospheric forcing. In addition to the metrics listed in Table 2, we assess the performance of the hydrodynamic model to simulate the peak in the sea level by computing the mean absolute error (MAE) for the highest water elevation value during the storm surge. Sea level or water elevation is defined as the barotropic sea level, including both tide and surge components.*

Table 2: Model performance metrics.

Metrics	Description	Formula
Mean Absolute Error (MAE)	Averaged absolute difference between observation and model estimate at data point	$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} \lvert m_i - o_i \rvert$
Root Mean Square Error (RMSE)	Weight of the mismatch between model and data by the variability in data and variability	$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(m_i - o_i)^2}$
Standard Deviation (σ)	Dispersion of data (model or observation) around the mean	$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}$
Correlation Coefficient (R)	Quantifies the strength and direction of the linear relationship between modeled and observed values	$R = \frac{\sum_{i=1}^{N}(m_i - \overline{m})(o_i - \overline{o})}{\sqrt{\sum_{i=1}^{N}(m_i - \overline{m})^2 \sum_{i=1}^{N}(o_i - \overline{o})^2}}$
Bias	*A systematic deviation from the observed value*	$bias = \frac{1}{N}\sum_{i=1}^{N} m_i - o_i$
Normalized RMSE (NRMSE)	*RMSE normalized by the standard deviation of observations; useful for comparing datasets with different scales.*	$NRMSE = \frac{RMSE}{\sigma_o}$

The result section has been updated to add the new validation stations and to clarify the performance of the atmospheric model.

The updated paragraph about this latter reads as: *The high-resolution atmospheric model with*

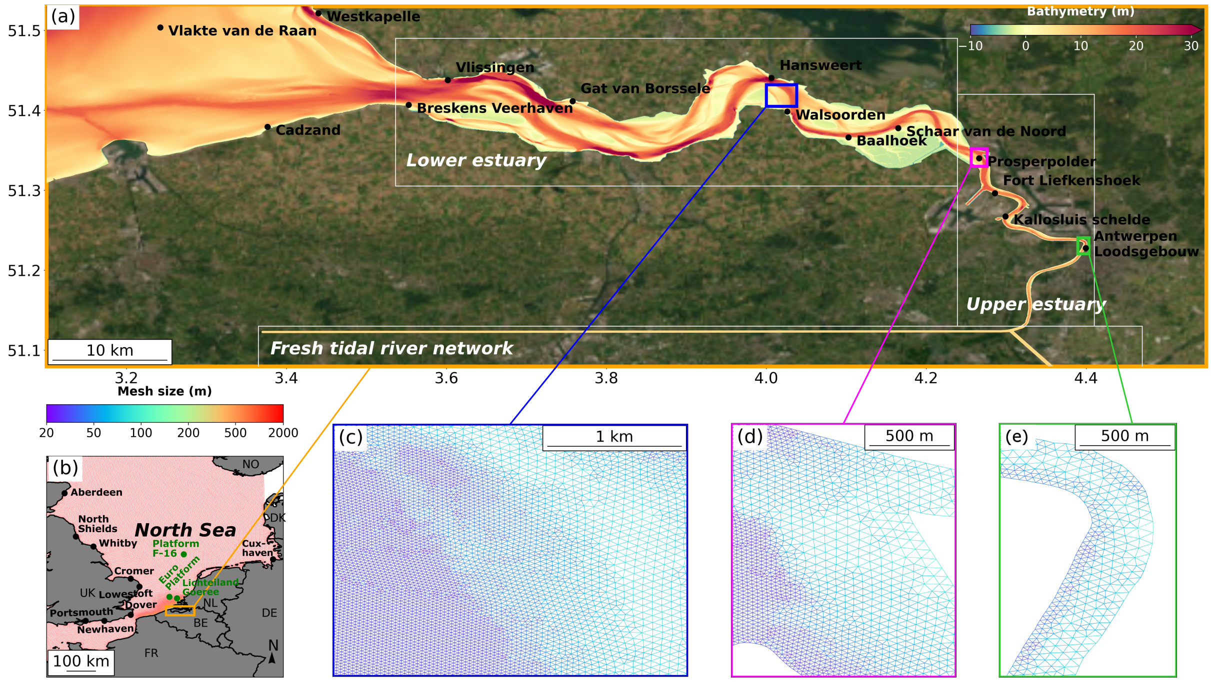

Figure 1: *Model computation domain covering the Scheldt and North Sea region. The upper panel (a) displays the bathymetry of the Scheldt estuary,with Esri World Imagery as basemap (ESRI, 2025). The white rectangles outline the tidal zones of the estuary and the fresh tidal river network is represented by artificial linear channels. Lower panels show the unstructured mesh configuration with (b) large view across the North Sea with black and green dots mark validation and (c-d) close-up views at three distinct locations on the Scheldt estuary. The dots in (a-b) mark validation points for the hydrodynamic model, colors in (b) differentiate coastal and offshore validation stations.*

spatial resolutions of 2 and 5 km (Fig. 2b-c and d-e) captures finer detail of the wind speed variability, particularly over continental regions, where the complex topography induces localized wind patterns. The best performance is over land with the atmospheric model at 5 km and 15 min resolution (Fig. 4). These regions benefit most from high resolution due to the influence of land features on wind flow. In contrast, wind speeds are generally higher over the open sea but tend to be more homogeneously distributed. As the spatial resolution coarsens to 10-30 km (Fig. 2d-f), fine-scale variations are increasingly unresolved with larger grid cells losing the details of localized phenomena. This illustrate the ability of higher resolutions to resolve mesoscale features, particularly in coastal and estuarine environments.

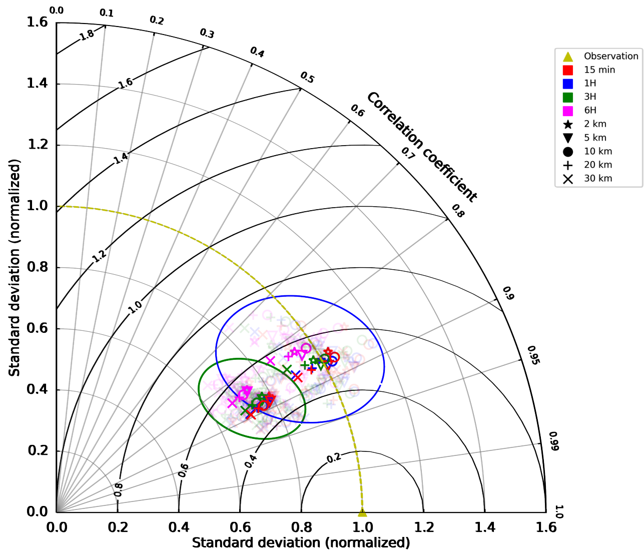

Figure 4: *Normalized Taylor diagram of the MAR simulation for wind speed, comparing different spatial resolutions, indicated by marker type, and temporal resolutions, indicated by color. Semi-transparent markers show individual validation points; bold markers represent regional centroids for weather stations in the Scheldt (blue ellipse) and the North Sea (green ellipse) regions as defined in Figure 2(a–b).*

G2 *About the ocean model: The results should be presented better; a Taylor diagram would better show the difference between the resolutions;*

Response: Thank you for the suggestion. We considered using a Taylor diagram to compare model resolutions in the Scheldt and open North Sea.

The Taylor diagram and its results analysis for the ocean read as: *The validation across different temporal and spatial resolutions reveals two distinct clusters corresponding to the Scheldt estuary and the offshore North Sea stations (Fig. 4). To ensure clarity and maintain readability, only those validation points (i.e. stations in the Scheldt shown in Fig. 1.a) and offshore North Sea stations represented by green markers in Fig. 1.b) were included in the analysis, while other coastal stations, where no specific calibration was applied, were excluded. For both regions, the position of centroid of the model performance in the Taylor diagram improves as the temporal resolution of the atmospheric forcing becomes finer, and the best results are obtained with 15 min temporal*

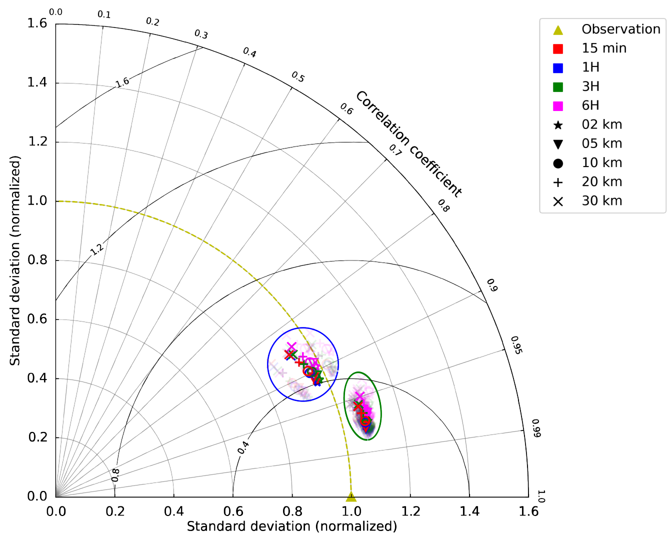

Figure 4: *Normalized Taylor diagram of sea surface elevation simulated by SLIM when forced by different spatial resolutions, indicated by marker type, and temporal resolutions, indicated by color. Semi-transparent markers show individual validation points; bold markers represent regional centroids for stations near the Scheldt estuary (blue ellipse) and the off-shore North Sea (green ellipse) regions as defined in Figure 1(a–b).*

resolution. The Scheldt cluster appears more compact and lies closer to the observational reference point, indicating that the model performs more accurately and consistently in this confined estuarine environment. Within this group, simulations at 5 km spatial resolution show the best alignment with observations, mirroring the behavior observed in the atmospheric forcing validation over the Scheldt (Fig. 4). Concerning the spatial resolution, the best results over the North Sea are obtained with the 2 km.

G3 *The results are just presented, without deep discussion and interpretation.*

Response: We appreciate this comment and have addressed it by substantially revising the discussion section. We now provide a more in-depth interpretation of the results, highlighting the physical mechanisms behind the observed model behavior and discussing broader implications and limitations.

In particular, we analyze how higher spatial and temporal resolution in atmospheric forcing improves wind field representation and surge modeling, especially within estuarine systems. While 5 km resolution performs well over land, only the 2 km resolution captures the localized wind structures needed to accurately simulate surge dynamics in the estuary.

We also discuss the influence of using ERA5 to force the regional MAR model, acknowledging that ERA5 tends to underestimate wind speeds, particularly during extreme events. This bias propagates into MAR and contributes to the underestimation of surge heights, even with high-resolution downscaling. These insights help explain the limits of resolution alone in correcting model performance and point to the need for improved or bias-corrected forcing datasets.

Finally, we separated the discussion and conclusion to improve clarity, better emphasize the main findings, and highlight their broader relevance for coastal and estuarine modeling.

We added the following paragraphs to the discussion section. They are articulated within the discussion section to provide a more comprehensive interpretation of the results and their implications.: *The improved performance of the model in estuarine regions can be linked to enhanced representation of wind fields at higher spatial and temporal resolutions. High-resolution atmospheric forcing significantly improves the representation of wind speed both over the sea and land (Fig. 4). Notably, wind speed simulations at 5 km spatial and 15 min temporal resolution yield the best agreement with observations over land. However, when evaluating the hydrodynamic response—particularly the surge within the estuary—the best performance is achieved with 2 km spatial resolution combined with 15 min temporal forcing. This distinction suggests that, while 5 km resolution may sufficiently capture atmospheric dynamics over land, the more refined 2 km resolution provide the best representation of the surge dynamics within the estuary. Nevertheless, given that 5 km resolution delivers close performances with lower computational cost, it may be a practical alternative in some cases. A key factor may be the improved representation of the coastline and topographic feature in the atmospheric model at higher spatial resolution (Fig. 2), which plays an important role in the atmospheric dynamics over land. Another additional explanation lies in the higher wind intensity over the sea, at high spatial and temporal resolution it enhances momentum transfer from the open sea into the estuary (Fig. 2). This improved representation of wind forcing enables more accurate simulation of surge propagation and amplification within the complex estuarine geometry.*

The benefit of better representing the atmospheric data in storm surge modeling is not specific to estuaries and can also include larger region. For example, in the Western Baltic Sea, Lorenz and Gräwe (2023) showed that the coarse resolution of atmospheric models such as ERA5 is also insufficient to resolve wind speeds correctly, leading to underestimated peak sea levels. The determination of appropriate spatial resolution for atmospheric forcing is ultimately dependent on the specific processes and scales being modeled. For coastal storm surge dynamics, which typically unfold over several hours and tens of kilometers, atmospheric forcing at such spatial scales may suffice (Agulles et al., 2024; Meyer and Gaslikova, 2023; Muller et al., 2014; Ridder et al., 2018). However, our results indicate that these standard resolutions may be inadequate for modeling storm surges in complex topographic and smaller scale regions such as estuaries. In those settings, finer-scale forcing is required to capture the local dynamics that drive the surge intensity. Furthermore, as demonstrated in our study, even some offshore areas benefit from higher-resolution atmospheric data, which yield more realistic representation of wind and hence improved surge model results.

Specific comments:

1. *r112: Describe better the terms in the equations of the model;*

   Response: The manuscript has been revised to include a brief description of the term in the equation and their physical significance.

   The additional sentence reads: *Eq. (1) and (2) represents mass and momentum conservation, respectively. The right-hand side of Eq. (2) includes terms representing the Coriolis force, gravitational acceleration, bottom friction, horizontal diffusion, wind stress, and the atmospheric pressure gradient.*

2. *r175: Does the model have the tidal potential terms? For such a large domain, they would be necessary. If not, show the good reproduction of tide;*

Response: The current model setup does not include explicit tidal potential terms. However, to accurately represent tidal dynamics, we apply boundary forcing derived from the TPXO9.5 global tidal model, which includes 21 tidal constituents, among them the 8 principal constituents (M2, S2, N2, K2, K1, O1, P1, and Q1). This tidal signal is combined with depth-averaged currents and sea surface elevation from the CMEMS reanalysis to reconstruct realistic boundary conditions that capture both tidal and residual ocean dynamics.

In light of your comment, we have extended the validation section to explicitly assess the hydrodynamic model's ability to reproduce tidal behavior. We now present a comparison of the modeled and observed water levels at several locations, with particular attention to the M2 tidal constituent. While the model slightly underestimates tidal height and low tide by approximately 0.1 m in some locations, this deviation remains small relative to the local tidal range (around 6 m). Overall, the dominant tidal signals are well represented.

In Section 3.2, we now describe how the model performs across different parts of the domain. Within the Scheldt estuary, where the calibration was focused, the model achieves high accuracy (e.g., RMSE values between 0.221 and 0.297 m and correlation coefficients consistently above 99%). At other stations locations where calibration was limited, performance is slightly lower (e.g., RMSE up to 0.968 m at Dover), but still within reasonable bounds, and the model maintains a good representation of both tidal amplitude and phase.

The update of section 3.2 about the ocean model validation now reads:

The time series analysis obtained from the model and tide gauges show that the model accurately reproduces the tidal cycles at the different stations (Fig. 4). There is a slight overestimation of the tidal height and low tide by about 0.1 m, this overestimation is also visible when we compare the tidal constituent M2 between the observation and the model (Fig. 4). However, these deviations are relatively minor as compared to the tidal range of about 6 m, and overall, the stations show an good representation of the tidal dynamics. The tidal components in the model are well captured, particularly for the dominant M2 constituent. Error metrics further indicate the model's robustness (Table 4). The RMSE values in the Scheldt range from 0.221 m at Westkapelle to 0.297 m at Antwerpen Loodsgebouw, corresponding to areas with tidal amplitudes of approximately 4 m. Offshore stations show slightly higher RMSEs—up to 0.968 m at Dover, likely due to the focus of calibration on the Scheldt and coarser mesh resolution far from the estuary. Nevertheless, the model still performs adequately offshore, showing good agreement in both amplitude and phase of the M2 tidal constituent. Overall, the bias and MAE values remain close to zero, and correlation coefficients—consistently above 99% in the Scheldt and over 85% elsewhere, which demonstrate a strong linear relationship between modeled and observed water levels. This high correlation highlights the model's capability in accurately capturing water elevation dynamics, particularly within the estuary, which was our primary focus.

3. *r199: Why averaging? It would be more physically correct to take the exact time;*

   Response: Averaging was chosen based on practices used in a similar study assessing the impact of temporal resolution on hydrodynamic modeling (Cucco et al., 2019). This method is also consistent with the approach used in reanalysis products such as ERA5, where coarser temporal resolutions (e.g., monthly) are derived by averaging higher-frequency data (e.g., hourly) (Hersbach et al., 2020).

   In our case, we chose to average the atmospheric fields to better represent the overall trends in wind and pressure during extreme events. While this approach smooths peak values, we found it

Table 4: Error metrics and principal tidal constituent M2 amplitude and phase for various stations along the Scheldt estuary (top part of table) and other locations in the domain (lower part of table), from November 25 to December 20, 2013.

Station	**RMSE [m]**	**Correlation [-]**	**MAE [m]**	**Bias [cm]**	**M2$_{\mathrm{obs}}$ [m]**	**M2$_{\mathrm{mod}}$ [m]**	**M2$_{\mathrm{obs}}$ [°]**	**M2$_{\mathrm{mod}}$ [°]**
Vlakte van de Raan	0.223	0.985	0.174	0.049	1.421	1.582	18.03	17.40
Cadzand	0.218	0.987	0.166	0.019	1.605	1.726	21.08	21.93
Westkapelle	0.213	0.986	0.163	0.000	1.469	1.590	25.51	24.70
Breskens Veerhaven	0.214	0.988	0.161	-0.003	1.671	1.773	29.70	29.48
Vlissingen	0.219	0.988	0.166	-0.004	1.673	1.795	31.52	32.19
Gat van Borssele	0.226	0.989	0.171	-0.078	1.756	1.895	38.13	37.80
Hansweert	0.238	0.990	0.180	-1.730	1.929	2.085	52.16	51.86
Walsoorden	0.242	0.990	0.185	-1.871	1.966	2.137	54.81	54.04
Baalhoek	0.245	0.990	0.187	-2.169	2.021	2.188	58.55	57.69
Schaar van de Noord	0.245	0.991	0.187	-2.316	2.048	2.220	60.95	60.43
Prosperpolder	0.255	0.991	0.194	-2.562	2.086	2.270	64.70	64.35
Fort Liefkenshoek	0.249	0.991	0.188	-2.868	2.154	2.295	67.21	65.86
Kallosluis Schelde	0.262	0.990	0.199	-9.805	2.191	2.339	69.13	67.29
Antwerpen Loodsgebouw	0.297	0.987	0.232	-12.059	2.218	2.361	75.65	70.91
Platform F16-A	0.158	0.949	0.118	0.019	0.502	0.560	180.27	183.49
Euro Platform	0.208	0.956	0.160	0.010	0.690	0.826	26.46	28.16
Lichteiland Goeree	0.227	0.963	0.175	0.003	0.818	0.991	40.55	41.71
Cuxhaven	0.540	0.899	0.442	-0.368	1.366	1.194	313.55	335.81
Cromer	0.684	0.844	0.569	1.708	1.513	1.479	158.56	188.84
Dover	0.968	0.866	0.803	-1.040	2.226	2.482	303.47	331.30
Lowestoft	0.403	0.813	0.323	0.581	0.665	0.676	227.88	263.35
Newhaven	0.924	0.877	0.816	-1.071	2.230	2.487	293.58	322.18
North Shields	0.635	0.880	0.537	-0.384	1.573	1.671	59.51	87.27
Portsmouth	0.663	0.859	0.594	-0.932	1.398	1.666	299.50	328.75
Whitby	0.671	0.873	0.569	-0.141	1.633	1.697	75.09	103.82

preferable for ensuring that lower-resolution scenarios retain a meaningful representation of storm evolution. For example, because the peak of a storm does not always occur at the same time across different temporal resolutions, averaging allows us to capture the general progression of the event without omitting key features. Since the center trajectory path of Xaver Storm crosses the North Sea region in a time frame of 12 h (Fig. 2), even with the coarsest temporal resolution of 6 h, the storm evolution is captured, which might not be the case if we take the exact time.

We acknowledge that this method may not be the most physically accurate for all applications. However, in the context of this study, focused on how temporal resolution influences modeled extremes we believe that averaging provides a more consistent and interpretable input, and aligns with current practices. A clarification has been added to the manuscript to justify this choice and highlight its implications.

The added explanation reads as: *This approach was chosen to preserve the overall structure and trends in wind and pressure during extreme events, following practices used in previous hydrodynamic modeling study (Cucco et al., 2019) and consistent with the averaging methods used in reanalysis products such as ERA5, where coarser temporal resolutions (e.g., monthly) are derived by averaging higher-frequency data (Hersbach et al., 2020). This method ensures that lower-resolution forcings still represent the general evolution of storm events, particularly where peak timing varies across resolutions. This allows for a more consistent comparison between temporal resolutions and avoids the risk of omitting key features due to temporal aliasing.*

4. *You use the terms "sea level" and "water elevation". Are they the same quantity? Barotropic sea level with surge and tide? Specify the definition the first time;*

Response: Thank you for this suggestion, your guess is correct. We have clarified the definition of ”sea level” and ”water elevation” in the text.

The definition has been added at the end of section 2.4. The added sentence reads as: *Sea level or water elevation is defined as the barotropic sea level, including both tide and surge components.*

5. *r222: Qualitative results. For me, the 5km is the best. Check in more in-situ stations and, possibly, during different events;*

   Response: 16 additional in situ stations were included for validation of the atmospheric data, they are depicted in Figure 2(a-b). We acknowledge that the 5 km resolution outperformed others in certain aspects and now discuss this in greater detail. Due to the computational cost of running the model at very high resolution, analyzing multiple events was not feasible; this limitation has been clarified in the revised manuscript.

   The adapted version of Figure 2 can be seen after.

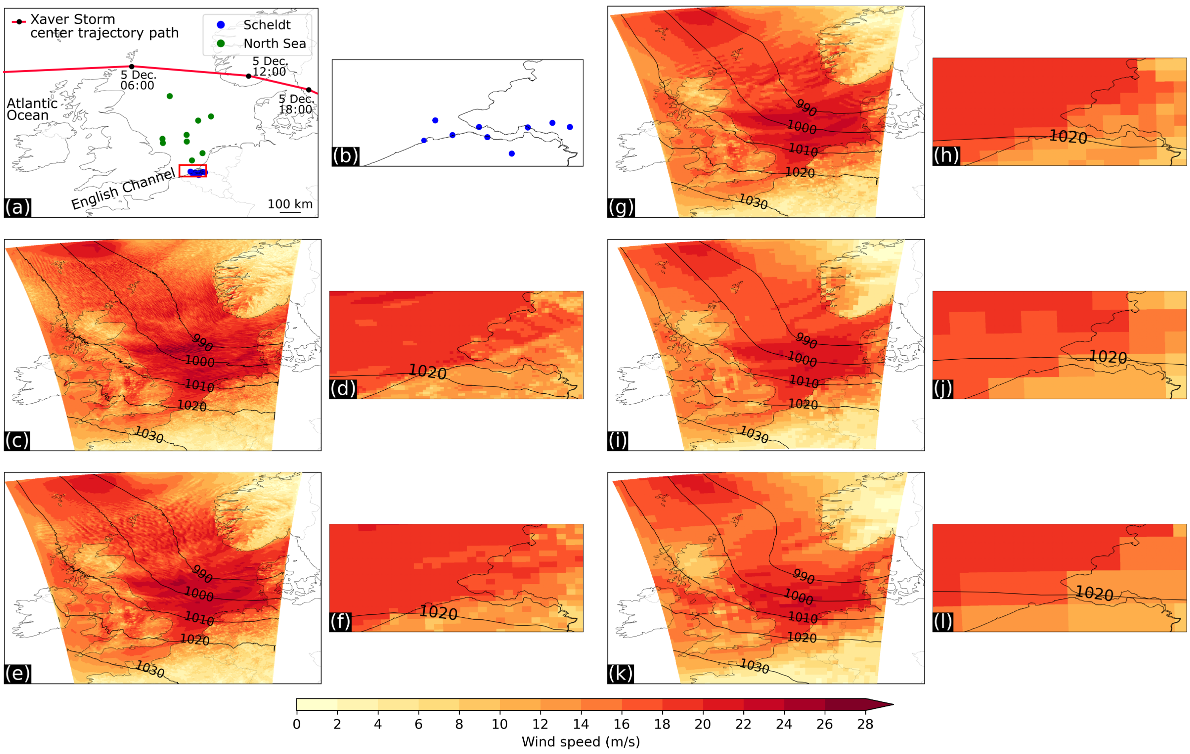

Figure 2: *(a-b) Storm Xaver’s central low-pressure trajectory over the North Sea and surrounding areas, showing its path on December 5, 2013, with time stamps in UTC. The storm path is indicated by a red curve; blue and green dots represent atmospheric data validation stations for the Scheldt estuary and North Sea, respectively. Panels (c–l) display snapshots of wind velocity magnitude at 10 m above the sea surface and sea level pressure contours at 12:00 UTC on December 5, 2013, as simulated by the MAR atmospheric model at varying spatial resolutions. These resolutions range from (c-d) 2 km, (e-f) 5 km, (g-h) 10 km, (i-j) 20 km, to (k-l) 30 km, each resolution includes an inset focusing on the Scheldt estuary and adjacent coastal areas.*

The section formerly in r219-225 now reads as:

The high-resolution atmospheric model with spatial resolutions of 2 and 5 km (Fig. 2.b-c and d-e) captures finer detail of the wind speed variability, particularly over continental regions, where the complex topography induces localized wind patterns. The best performance is performed over land with the atmospheric model at 5 km and 15 min resolution (Fig. 4). These regions benefit most from high resolution due to the influence of land features on wind flow. In contrast, wind speeds are generally higher over the open sea but tend to be more homogeneously distributed. As the spatial resolution coarsens to 10-30 km (Fig. 2.d-f), fine-scale variations are increasingly unresolved with larger grid cells losing the details of localized phenomena. This illustrate the ability of higher resolutions to resolve mesoscale features, particularly in coastal and estuarine environments.

The additional Figure can be seen after, since the figure was not in the original submission its numbering corresponding to the one in the new submission:

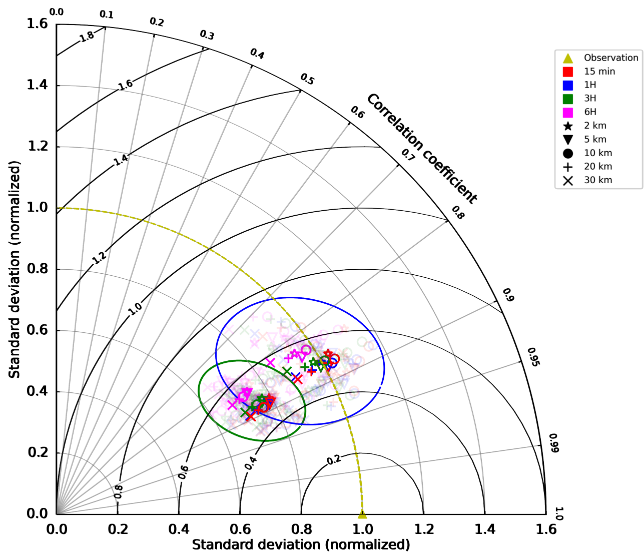

Figure 4: *Normalized Taylor diagram of the MAR simulation for wind speed, comparing different spatial resolutions, indicated by marker type, and temporal resolutions, indicated by color. Semi-transparent markers show individual validation points; bold markers represent regional centroids for weather stations in the Scheldt (blue ellipse) and the North Sea (green ellipse) regions as defined in Figure 2(a–b).*

And the section commenting this result reads as: *The normalized Taylor diagram, which includes additional observation stations, reveals consistent patterns in the performance of MAR wind speed simulations across different spatial and temporal resolutions (Figure 4). Simulations with finer spatial resolutions (2 km and 5 km) and shorter temporal intervals (15 minutes and 1 hour) tend to cluster closer to the observation point, indicating better agreement with observed wind speed in terms of variability and correlation. Among these, the 5 km resolution consistently outperforms all other configurations. These high-resolution setups exhibit stronger correlations and a better representation of wind speed variability. In contrast, simulations using coarser spatial resolutions (20 km and 30 km) and longer temporal outputs (3 h and 6 h) are located farther from the ob-*

servation point, reflecting a systematic decrease in performance associated with lower correlation and underestimated variability. Clear regional contrasts are also apparent: simulations over the Scheldt region generally align more closely with observations, while those over the North Sea show comparatively weaker agreement. This discrepancy between these two areas is largely due to the fact that the wind in North Sea is measured at 30-50 m a.s.l (on oil platforms) and not at 10 m a.s.l. like MAR. This distribution underscores the influence of spatial and temporal resolution on model accuracy across differing geographic settings.

Concerning the possibility to consider different events this was adressed in the discussion method as a limitation of the model as such:

Finally, due to the high computational cost of modeling at such high resolutions ranging from 20 m to 2 km for the hydrodynamic model and from 2 km for the atmospheric model, this study was limited to a single, extreme storm surge event. Generating high-resolution atmospheric forcing alone required several hours to days per simulation. Therefore, we focused on a single event, extreme storm surge one of the highest recent event occurring in the region Rucińska (2019), as a representative case. Including multiple events would improve robustness, but reproducing this level of detail across a broader dataset is currently not feasible. Future research should explore strategies such as adaptive meshing or region-specific refinement to extend the analysis to a wider range of events while maintaining computational efficiency.

6. *r241: Knowing the height of the pressure sensor would allow for a better correction;*

   Response: We thank the reviewer for this remark. The height of the pressure sensor is not specified in the dataset we used. However, we have added a sentence in the text to clarify this limitation and its potential impact on the results.

   The addded sentence reads: *As the exact height of the pressure sensor at ZDI is not available, a precise altitude correction could not be applied. Instead, the mean bias is removed, and the centered RMSE (CRMSE) is computed to better isolate model performance from the elevation-induced offset.*

7. *Table 3: RRMSE is not defined;*

   Response: Thank you for pointing this out. Although RRMSE was previously defined as RMSE divided by the standard deviation of observations, we agree that it should also be explicitly stated in the section presenting the error metrics. While we referred to it as RRMSE, this is more commonly known as NRMSE. The definition has now been added for clarity in Table 2.

8. *Figure 2: use a discrete colorbar;*

   Response: Figure 2 has been updated as suggested, adjusting the colorbar to a discrete version. Please note that the figure now contains additional details following suggestions from Reviewer 1, which we believe enhance its clarity and completeness.

   The revised version of Figure 2 can be found hereafter.

9. *Figure 4: reduce the x-axis, use different line types. It would be nice to see also the residual part (with a harmonic analysis);*

   Response: The x-axis has been shortened as suggested to a shorter time period, and different line types are now used to better distinguish the simulations. We have also included the residual part

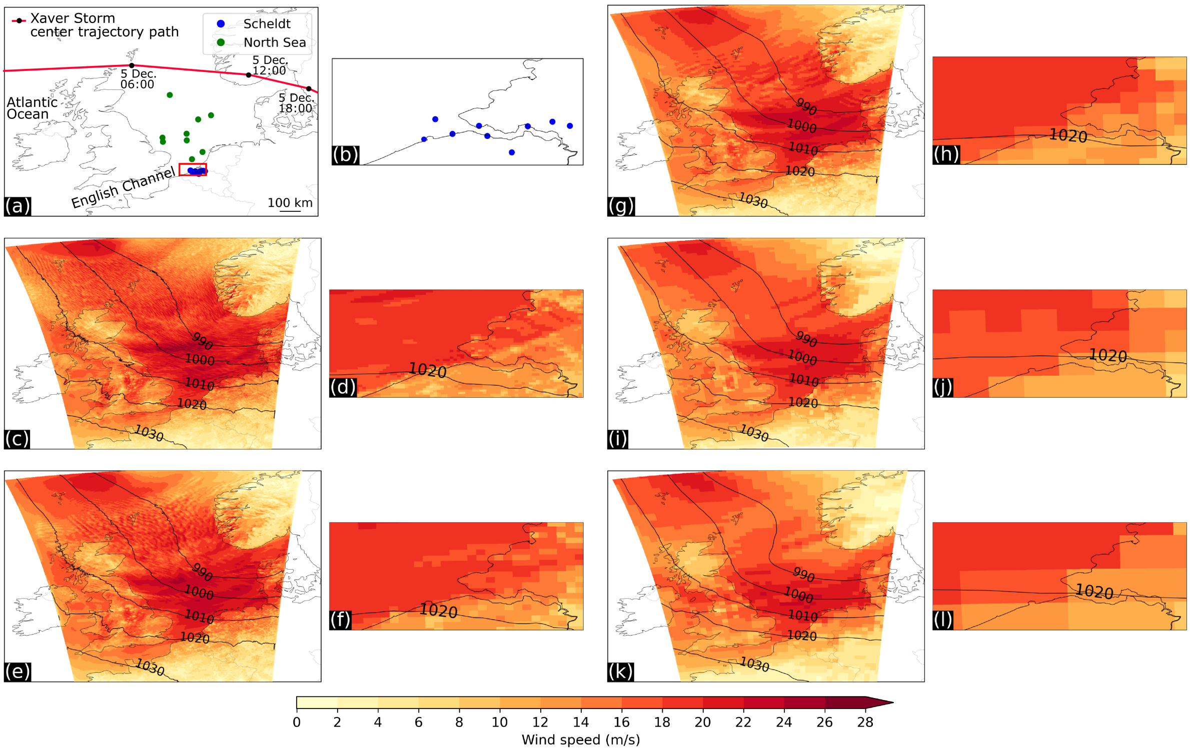

Figure 2: *(a-b) Storm Xaver's central low-pressure trajectory over the North Sea and surrounding areas, showing its path on December 5, 2013, with time stamps in UTC. The storm path is indicated by a red curve; blue and green dots represent atmospheric data validation stations for the Scheldt estuary and North Sea, respectively. Panels (c–l) display snapshots of wind velocity magnitude at 10 m above the sea surface and sea level pressure contours at 12:00 UTC on December 5, 2013, as simulated by the MAR atmospheric model at varying spatial resolutions. These resolutions range from (c-d) 2 km, (e-f) 5 km, (g-h) 10 km, (i-j) 20 km, to (k-l) 30 km, each resolution includes an inset focusing on the Scheldt estuary and adjacent coastal areas.*

derived through harmonic analysis using the package UTide. 5 panels for tidal gauges have been added.

The new version of Figure 4 can be found hereafter.

10. *Use Taylor diagrams both for the atmosphere and ocean models;*

    Response: To further assess the model performance, Taylor diagrams were used to synthesize the standard deviation, correlation coefficient, and centered root mean square error (CRMSE) of both atmospheric and oceanic variables. Figure 4 shows the performance of different atmospheric model configurations with different spatial and temporal resolution, while Figure 4 summarizes the skill of the ocean model in reproducing observed sea surface height and surge levels. The use of Taylor diagrams complements the individual metrics reported in Table 2 and reinforces the conclusions regarding the impact of spatial and temporal resolution on model performance.

11. *r323: other studies, but you cite only one;*

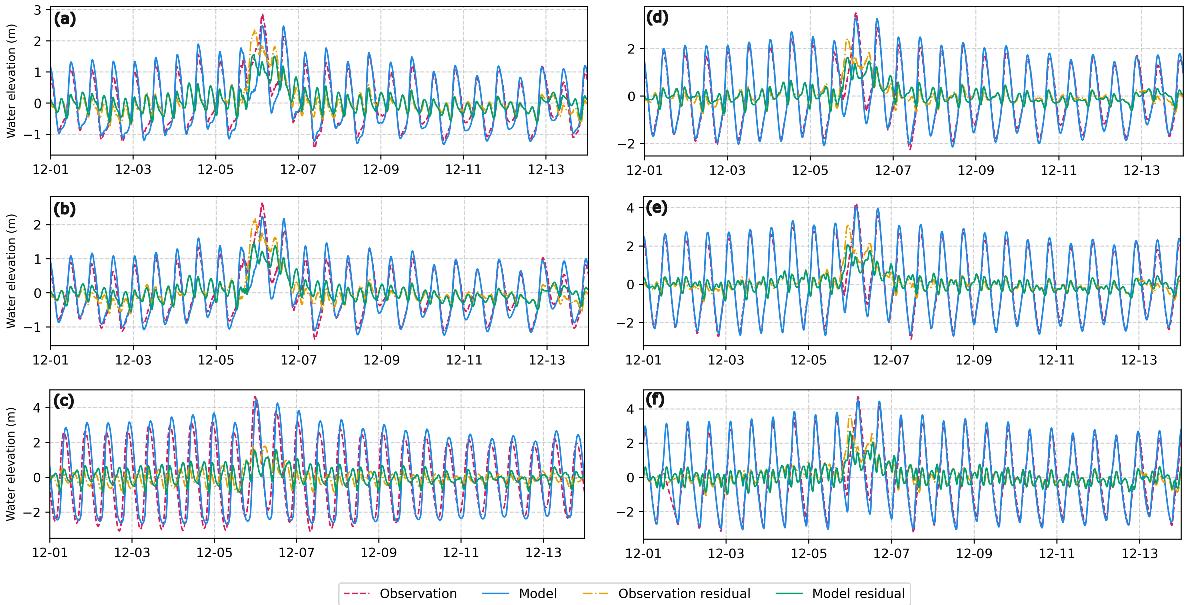

Figure 4: Time series of sea surface elevation from December 1st to December 15th, 2013, at six locations, on the offshore and coastal North Sea stations (a) Lichteiland Goeree, (b) Euro Platform, (c) Dover; and Scheldt estuary stations: (d) Vlakte van de Raan, (e) Hansweert, and (f) Antwerpen Loodsgebouw. Results are shown as simulated by SLIM using MAR atmospheric forcing with 2 km spatial and 15 min temporal resolution (blue), compared against observed sea surface elevation (red). Detided (residual) elevations are also shown for both observations (orange) and model outputs (green).

Response: We thank the reviewer for pointing this out. In the revised manuscript, we added more references to support the statement.

The magnitude of the modeled storm surges is highly sensitive to the resolution of atmospheric forcing, a finding that aligns with previous studies (Weaver et al., 2016; Agulles et al., 2024). While Agulles et al. (2024) emphasized the role of temporal resolution, particularly at larger spatial scales, our results indicate that spatial resolution has a stronger influence when modeling estuarine environments at finer scales. This contrasts with Agulles et al.'s findings, where temporal resolution played a more significant role than spatial resolution at scales of 1.5–25 km and temporal resolutions ranging from hourly to daily. This divergence in results may be because of the difference in scales of interest, Agulles et al. (2024) focused primarily on larger-scale coastal dynamics using atmospheric data at broader spatial and temporal scales, on the continental scale along the European coastline and on several decades, while our study concentrated on the finer-scale dynamics of estuarine regions and smaller temporal scale. Similar conclusions on the importance of spatial resolution were drawn by Weaver et al. (2016), who showed that increasing the spatial resolution of meteorological forcing improved the simulation of storm surge during a cold front event in the Indian River Lagoon, Florida. This highlights that the relative importance of spatial versus temporal resolution can vary depending on the scale and characteristics of the system under study. Furthermore, our use of a different atmospheric data source from them may have contributed to these contrasting results. While Agulles et al. (2024) found optimal results using the ERA5 Reanalysis Dataset at an hourly resolution with a 31 km spatial scale, our study suggests that such *spatial resolutions lead to the under performance of the surge model in complex environment such as estuaries, where finer spatial scales yield more accurate storm surge modeling. The benefit of better representing the atmospheric data in storm surge modeling is not specific to estuaries and can also include larger region. For example, in the Western Baltic Sea, Lorenz and Gräwe (2023) showed that the coarse resolution of atmospheric models such as ERA5 is also insufficient to resolve wind speeds correctly, leading to underestimated peak sea levels. The determination of appropriate spatial resolution for atmospheric forcing is ultimately dependent on the specific processes and scales being modeled. For coastal storm surge dynamics, which typically unfold over several hours and tens of kilometers, atmospheric forcing at such spatial scales may suffice (Agulles et al., 2024; Meyer and Gaslikova, 2023; Muller et al., 2014; Ridder et al., 2018). However, our results indicate that these standard resolutions may be inadequate for modeling storm surges in complex topographic and smaller scale regions such as estuaries. In those settings, finer-scale forcing is required to capture the local dynamics that drive the surge intensity. Furthermore, as demonstrated in our study, even some offshore areas benefit from higher-resolution atmospheric data, which yield more realistic representation of wind and hence improved surge model results.*

12. *r367: The non-linear interaction between tides and surge was well studied; cite more papers and show it in the results if present.*

    Response: Two additional references (Ferrarin et al., 2012; Wahle et al., 2022) have been included and the discussion has been expanded to better contextualize the results and highlight their broader implications concerning the non-linear interaction between tides and surges in the model result.

    The paragraph now reads as: *Consistency in the timing of peak storm surges across various spatial and temporal resolutions is a another finding of this study. We found that the predicted timing of peak surge events remained relatively stable, with deviations typically within ±20 minutes, regardless the resolution of the atmospheric model. This stability suggests that factors beyond the spatial or temporal resolution of atmospheric forcing may be influencing the timing of peak surges. One mechanism in action in the North Sea is the modulation of surge timing by the tidal cycle. Surge maxima in the region where studied to occur predominantly on rising tide (Horsburgh and Wilson, 2007). During the Xaver Storm, the peak surge event coincided with high tide, strongly implying that this synchronization played a role in shaping the observed timing. This alignment likely contributed to the consistent timing predictions across simulations. Crucially, such tide-surge interactions are not linear; they involve complex, non-linear processes where the combined effects of tide and surge differ substantially from the sum of their individual contributions. These dynamics have been extensively documented in the literature (Ferrarin et al., 2012; Al Azad et al., 2018; Thomas et al., 2019; Wahle et al., 2022). In our results, the tight clustering of surge peak timings across all model configurations, regardless of their forcing resolution, supports the presence of these non-linear interactions.*

References

Miguel Agulles, Marta Marcos, Angel Amores, and Tim Toomey. Storm surge modelling along European coastlines: The effect of the spatio-temporal resolution of the atmospheric forcing. *Ocean Modelling*, 192:102432, December 2024. ISSN 1463-5003. doi: 10.1016/j.ocemod.2024.102432.

A. S. M. Alauddin Al Azad, Kazi Samsunnahar Mita, Md Wasif Zaman, Marin Akter, Tansir Zaman Asik, Anisul Haque, Mohammad Asad Hussain, and Md Munsur Rahman. Impact of Tidal Phase on

Inundation and Thrust Force Due to Storm Surge. *Journal of Marine Science and Engineering*, 6(4): 110, December 2018. ISSN 2077-1312. doi: 10.3390/jmse6040110.

Andrea Cucco, Giovanni Quattrocchi, and Stefano Zecchetto. The role of temporal resolution in modeling the wind induced sea surface transport in coastal seas. *Journal of Marine Systems*, 193:46–58, May 2019. ISSN 0924-7963. doi: 10.1016/j.jmarsys.2019.01.004.

Christian Ferrarin, Aron Roland, Marco Bajo, Georg Umgiesser, Andrea Cucco, Silvio Davolio, A. Buzzi, Piero Malguzzi, and O. Drofa. Tide-surge-wave modelling and forecasting in the Mediterranean Sea with focus on the Italian coast. *Ocean Modelling*, 61, January 2012. doi: 10.1016/j.ocemod.2012.10.003.

Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Saleh Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Gionata Biavati, Jean Bidlot, Massimo Bonavita, Giovanna De Chiara, Per Dahlgren, Dick Dee, Michail Diamantakis, Rossana Dragani, Johannes Flemming, Richard Forbes, Manuel Fuentes, Alan Geer, Leo Haimberger, Sean Healy, Robin J. Hogan, Elías Hólm, Marta Janisková, Sarah Keeley, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Gabor Radnoti, Patricia de Rosnay, Iryna Rozum, Freja Vamborg, Sebastien Villaume, and Jean-Noël Thépaut. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146 (730):1999–2049, 2020. ISSN 1477-870X. doi: 10.1002/qj.3803.

K. J. Horsburgh and C. Wilson. Tide-surge interaction and its role in the distribution of surge residuals in the North Sea. *Journal of Geophysical Research: Oceans*, 112(C8):2006JC004033, August 2007. ISSN 0148-0227. doi: 10.1029/2006JC004033.

Marvin Lorenz and Ulf Gräwe. Uncertainties and discrepancies in the representation of recent storm surges in a non-tidal semi-enclosed basin: A hindcast ensemble for the Baltic Sea. *Ocean Science*, 19 (6):1753–1771, December 2023. ISSN 1812-0784. doi: 10.5194/os-19-1753-2023.

Elke Magda Inge Meyer and Lidia Gaslikova. Investigation of historical severe storms and storm tides in the German Bight with century reanalysis data. *Natural Hazards and Earth System Sciences*, 24(2): 481–499, September 2023. doi: 10.5194/egusphere-2023-2068.

Héloïse Muller, Lucia Pineau-Guillou, Déborah Idier, and Fabrice Ardhuin. Atmospheric storm surge modeling methodology along the French (Atlantic and English Channel) coast. *Ocean Dynamics*, 64 (11):1671–1692, November 2014. ISSN 1616-7228. doi: 10.1007/s10236-014-0771-0.

Nina Ridder, Hylke de Vries, Sybren Drijfhout, Henk van den Brink, Erik van Meijgaard, and Hans de Vries. Extreme storm surge modelling in the North Sea. *Ocean Dynamics*, 68(2):255–272, February 2018. ISSN 1616-7228. doi: 10.1007/s10236-018-1133-0.

Dorota Rucińska. Describing Storm Xaver in disaster terms. *International Journal of Disaster Risk Reduction*, 34:147–153, March 2019. ISSN 2212-4209. doi: 10.1016/j.ijdrr.2018.11.012.

Ajimon Thomas, JC Dietrich, TG Asher, M Bell, BO Blanton, JH Copeland, AT Cox, CN Dawson, JG Fleming, and RA Luettich. Influence of storm timing and forward speed on tides and storm surge during Hurricane Matthew. *Ocean Modelling*, 137:1–19, May 2019. ISSN 1463-5003. doi: 10.1016/j.ocemod.2019.03.004.

Kathrin Wahle, B. Emil Vassilev Stanev, and Joanna Staneva. Detecting anomalous sea-level states in North Sea tide gauge data using of autoassociative Neural Network, June 2022.

Robert J. Weaver, Peyman Taeb, Steven Lazarus, Michael Splitt, Bryan P. Holman, and Jeffrey Colvin. Sensitivity of modeled estuarine circulation to spatial and temporal resolution of input meteorological forcing of a cold frontal passage. *Estuarine, Coastal and Shelf Science*, 183:28–40, December 2016. ISSN 02727714. doi: 10.1016/j.ecss.2016.10.014.