Dear Authors,

Thank you for submitting the revised manuscript. I appreciate your detailed point-by-point responses. However, I must note that several of my comments from the previous round have not been adequately addressed. To facilitate further improvements, I will focus on the most critical aspects that require attention.

Dear reviewer, thank you for spending the time to review the manuscript and providing feedback. We have considered your suggestions and used them to further improve the manuscript. The main changes include the following:

- Better explanation of what drives the water levels differences between the model configurations G1 and N1;
- Better interpretation of the meaning of the RMSE values and comparison against other large-scale studies;
- Restructuring of the discussion sections and rewriting some other parts of the manuscript to better highlight the added value of our work;
- Updating of references to better reflect the latest state-of-the-art modelling approach for large-scale storm surge and flood assessments.

Below we provide response to your comments and include the changes that have been made in the manuscript.

3. Modeling Results:

I acknowledge the progress made in addressing some aspects of the modeling results, particularly the temporal component (G1 to G2), which is clear and logical. However, I remain very concerned about the significant water level changes (over 10 cm) observed in the Gulf of Mexico due to changes in input bathymetry (G1 to N1). Neither the revised manuscript nor the rebuttal provides a sufficient justification for these findings. Phrases such as "may help explain" are speculative and do not constitute a robust explanation.

Dear reviewer, thank you for your acknowledgment of the improvement in the temporal component results. Regarding the differences between G1 and N1, we have conducted further analyses based on your feedback, to better understand the source of these differences. As you suggested, to investigate the mechanism behind the water level changes we have examined the tidal and storm surge components separately.

Tidal contribution

To analyse the tidal contribution, we have ran tide-only simulations for both model configurations without meteorological forcing. Figure A6 shows the differences in maximum tide-only water levels between G1 and N1, indicating no significant differences occur in the Gulf of Mexico, with variations only along the coast. Figure A7 further confirms this at the observation stations, where most locations exhibit minimal differences between the two configurations. The only exceptions are stations 5 and 9. At station 5, the difference comes from its location in an estuary, while at station 9, the difference is caused by the interpolation of bathymetry from N1 onto a higher-resolution grid, allowing better representation of the barrier islands and affecting the tidal range (see Figure A11).

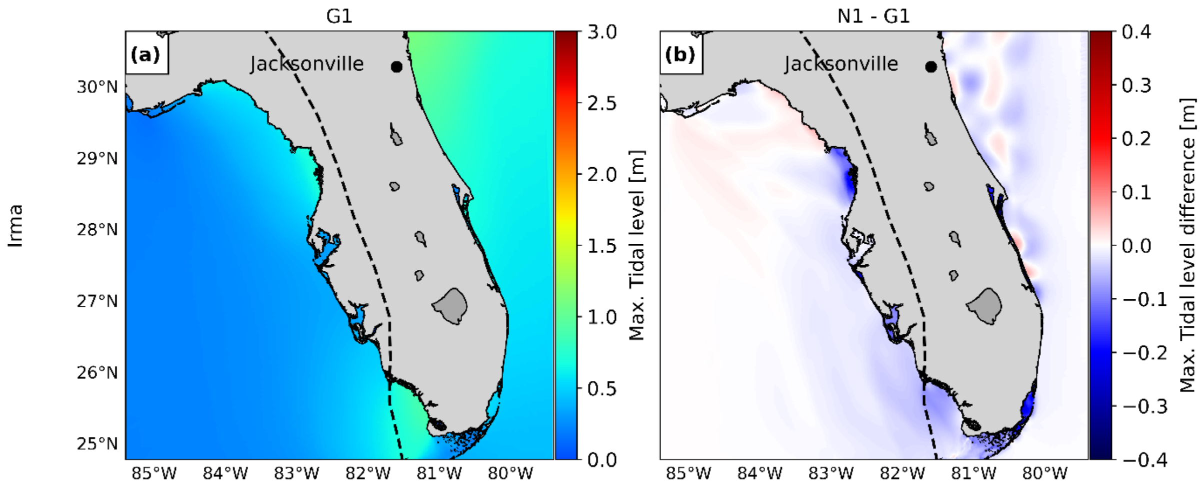

Figure A6. Maximum water levels for the tide only simulation of G1 (panel a). Difference between the maximum water level for the tide only simulations of N1 and G1 (panel b).

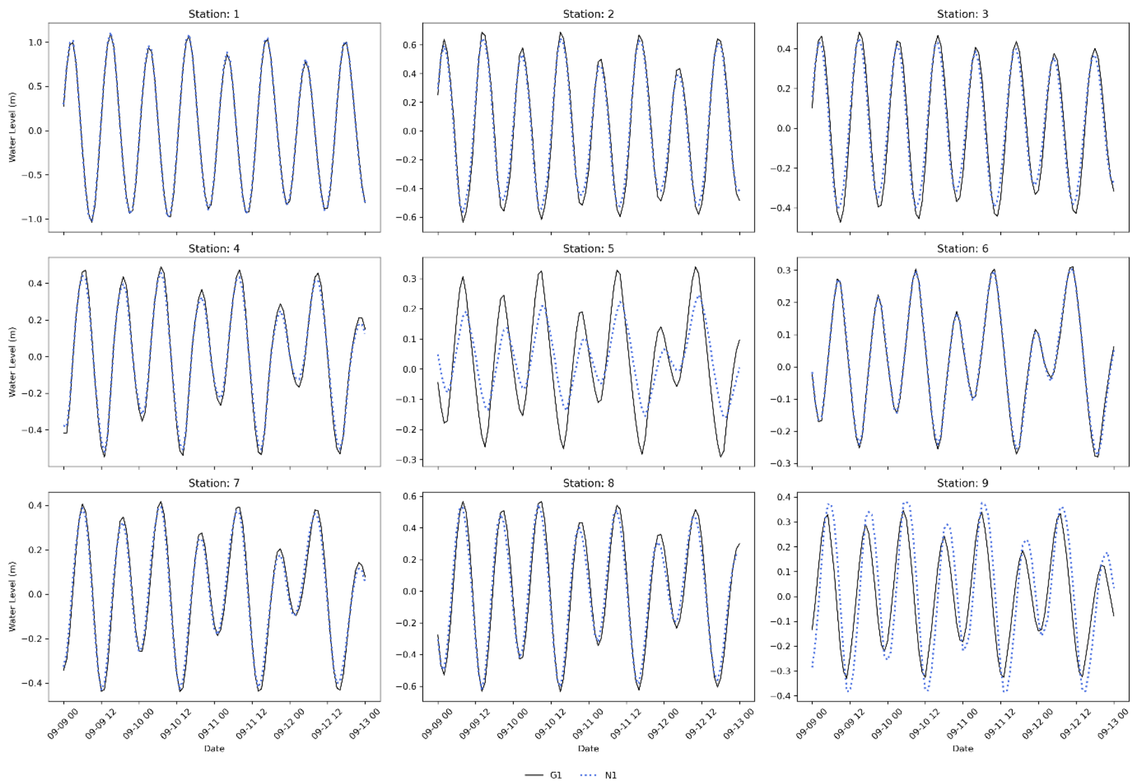

Figure A7. Water levels for the tide only simulations for the case study Irma model configurations G1 and N1, for the nine locations depicted in Fig. 3.

Surge contribution

To analyse the storm surge contribution, we have ran surge-only simulations with meteorological forcing but without tidal forcing. Figure A8 shows that differences between G1 and N1 occur primarily due to the storm surge propagations. The timeseries of surge-only simulations (Figure A9) and spatial differences between G1 and N1 per timestep (Figure A10) show that the differences between G1 and N1 do not persist throughout the event, but are most pronounced at the peak of TC Irma.

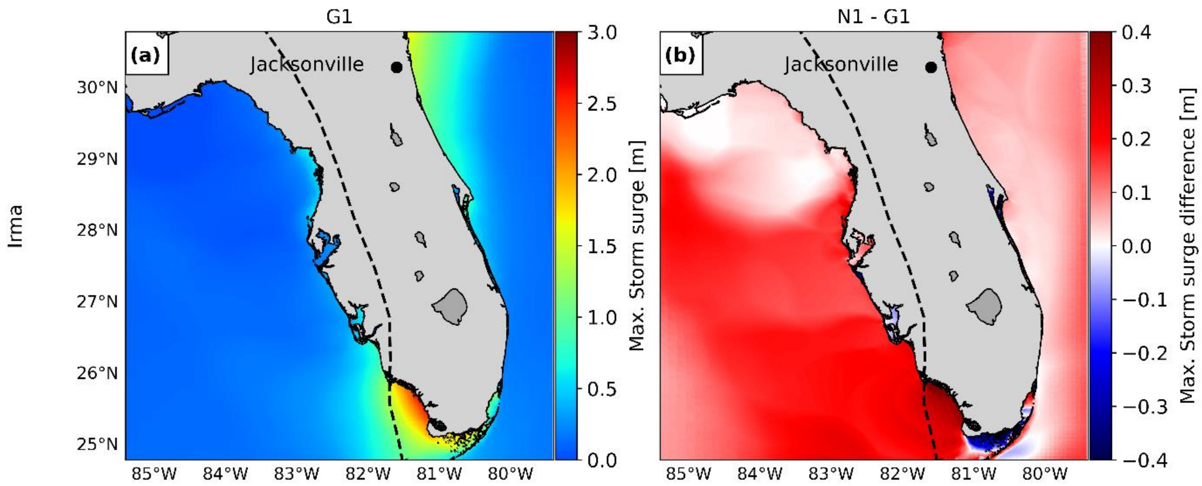

Figure A8. Maximum water levels for the storm surge only simulation of G1 (panel a). Difference between the maximum water level for the tide only simulations of N1 and G1 (panel b).

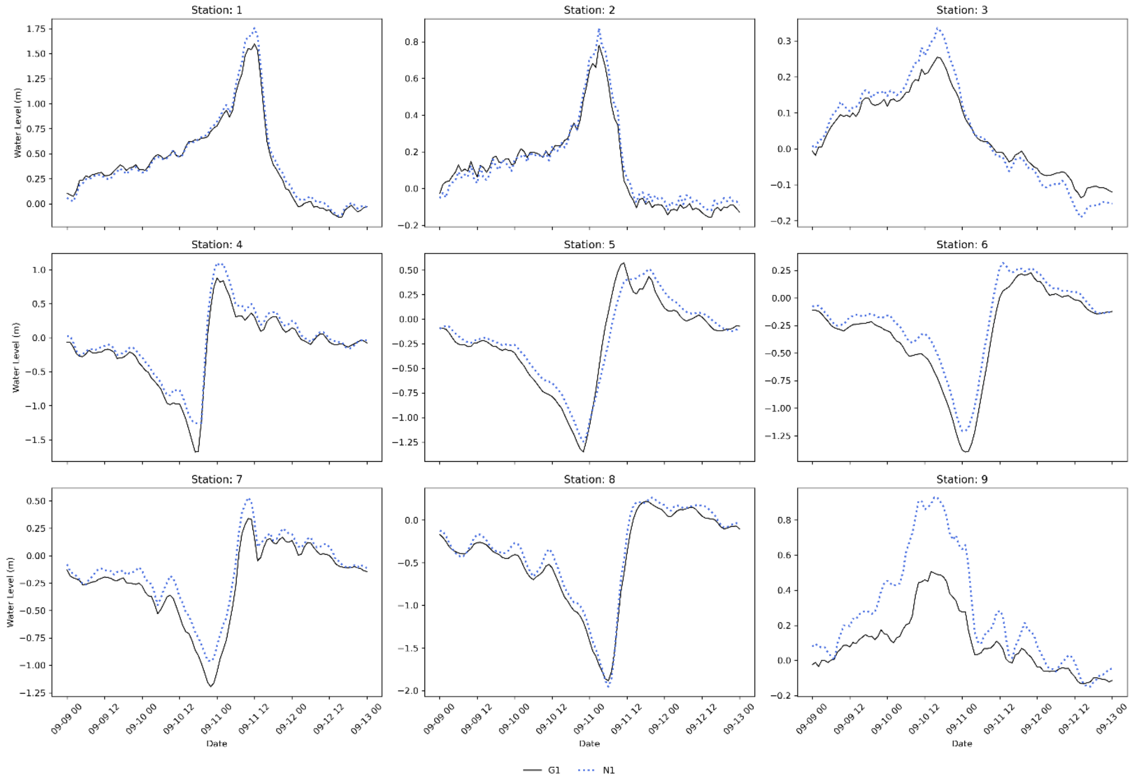

Figure A9. Water levels for the storm surge only simulations for the case study Irma model configurations G1 and N1, for the nine locations depicted in Fig. 3.

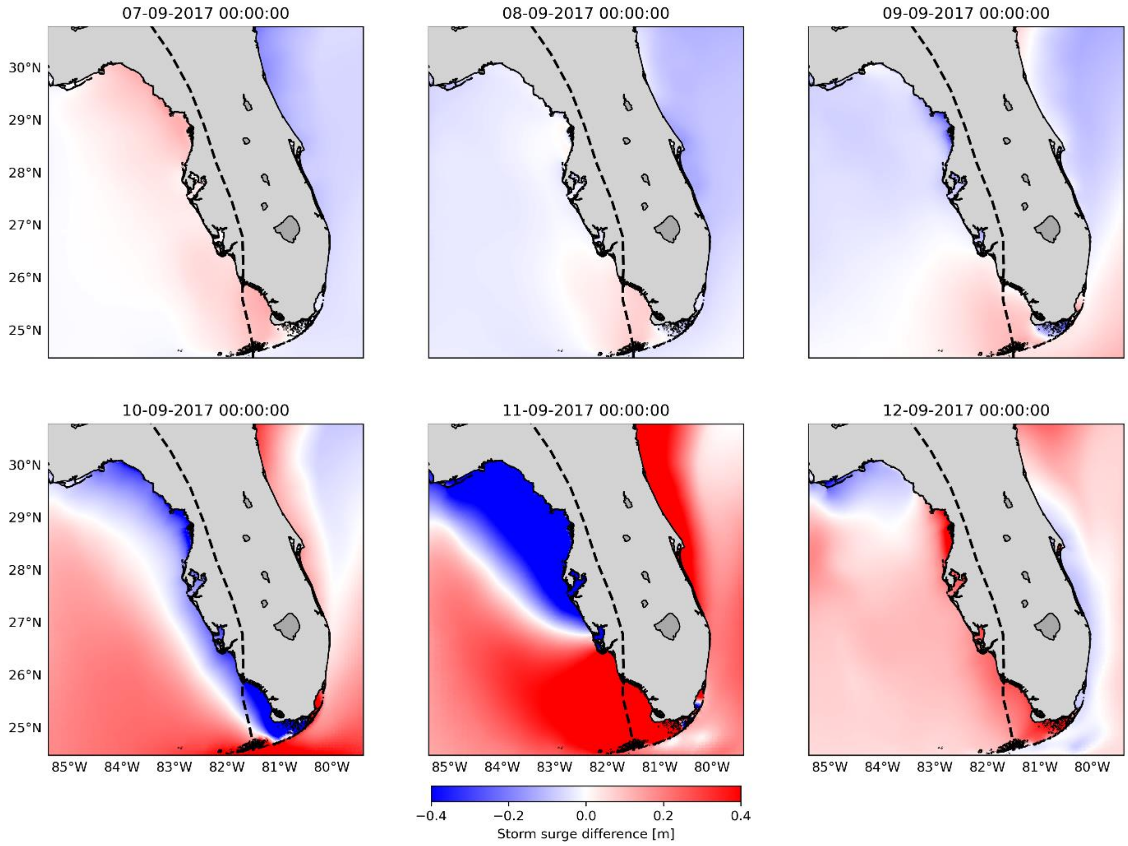

Figure A10. Difference in water levels for the storm surge only simulations of N1 and G1 for different timesteps, before TC Irma makes landfall (07-09-2017 until 09-09-2017), during the peak (between 10-09-2017 and 11-09-2017) and after the peak (12-09-2017).

Two main factors influence storm surge propagation: (1) meteorological forcing and (2) bathymetry.

(1) Regarding the meteorological forcing, we have looked at the differences due to the interpolation of ERA5 data into the grid of G1 and N1. Figures R1, R2 and R3, show that pressure and wind fields remain almost identical between the two configurations, indicating that atmospheric forcing is not responsible for the observed discrepancies.

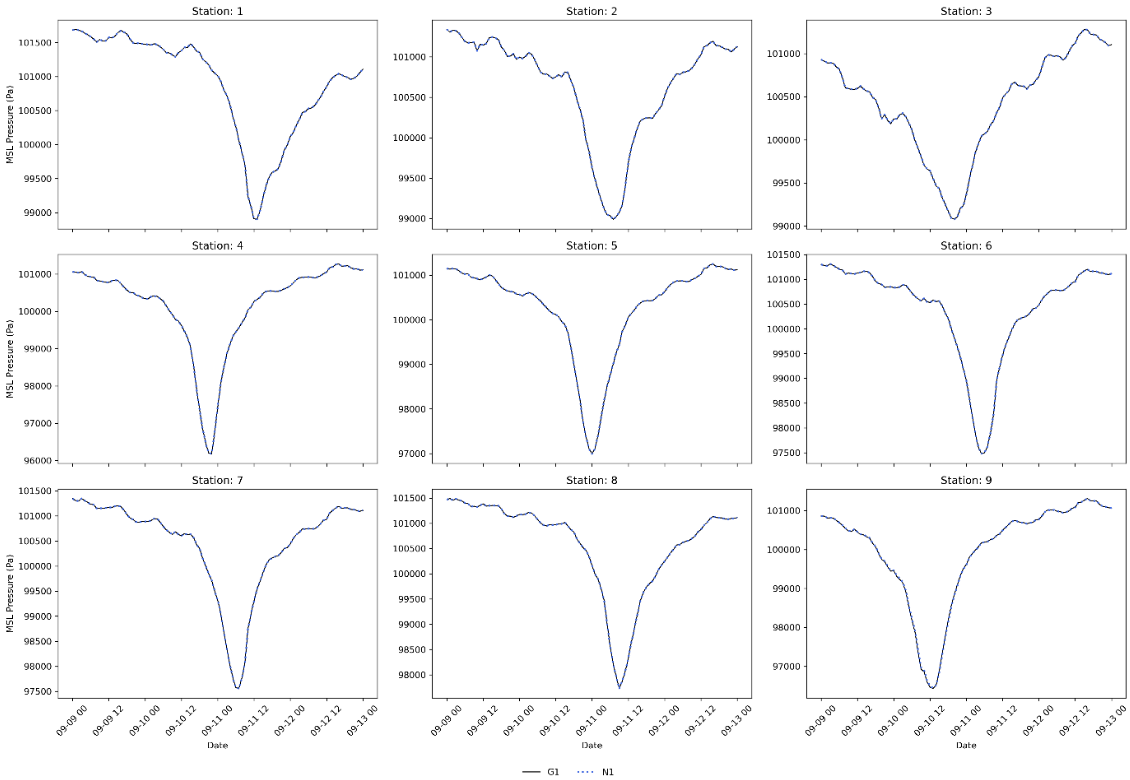

Figure R1. MSL pressure for the storm surge only simulations for the case study Irma model configurations G1 and N1, for the nine locations depicted in Fig. 3.

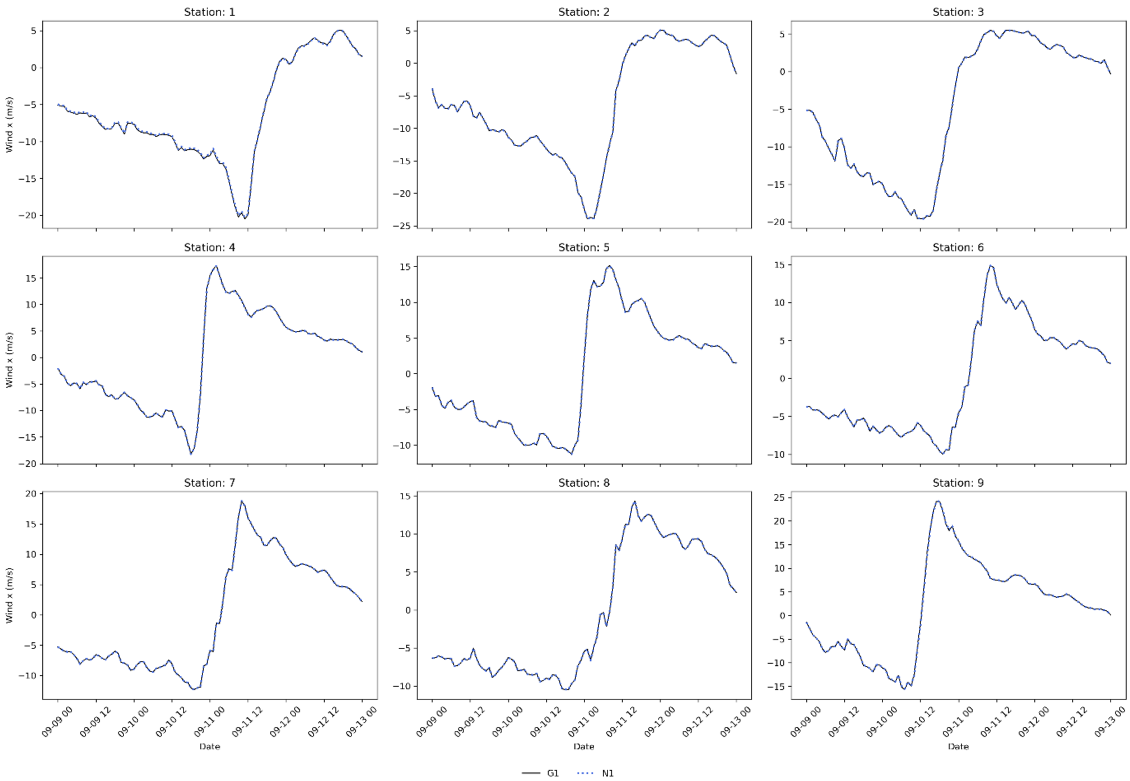

Figure R2. X component of the wind vector for the storm surge only simulations for the case study Irma model configurations G1 and N1, for the nine locations depicted in Fig. 3.

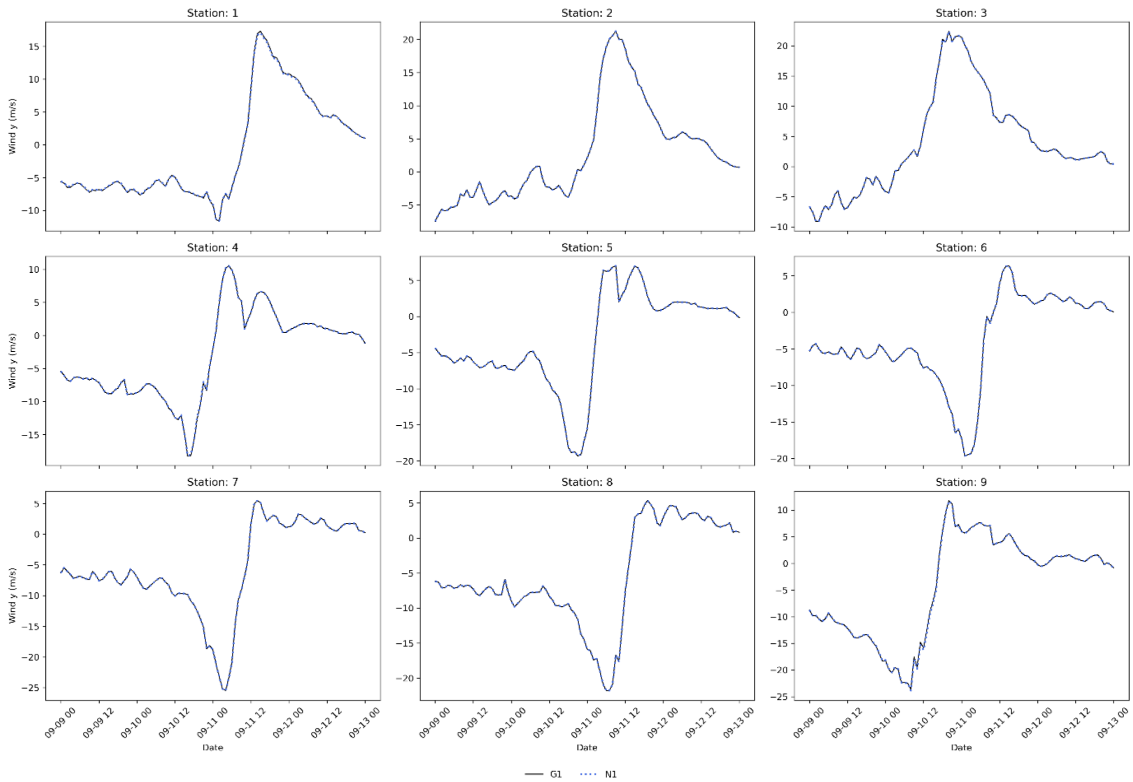

Figure R3. Y component of the wind vector for the storm surge only simulations for the case study Irma model configurations G1 and N1, for the nine locations depicted in Fig. 3.

(2) When analysing the bathymetry, significant differences arise from the interpolation of GEBCO 2019 onto the grids of G1 and N1. The sharp changes in bathymetry along the coasts of Florida (see Figure A11, left panel), lead to interpolation differences that can exceed 100 m in depth at the same location. Despite both configurations using the same source of bathymetry, GEBCO 2019, differences in resolution result in substantial different depth representations.

In the barrier islands south of Florida, for instance, the coarser G1 grid does not fully resolve the barrier island topography (Figure A11, left panel), while the finer N1 grid captures these features more accurately (Figure A11, middle panel). These differences in bathymetry have a direct impact on storm surge propagation. Figure A10 illustrates that in the finer-resolution N1 configuration, water can pass through the barrier islands more effectively than in G1. At timestep 10-09-2017, when there is a negative surge north of the barrier island, G1 produces higher water levels because the water remains trapped. Conversely, during the peak of TC Irma on the 11-09-2017, the water levels in G1 are lower than N1 because less water can travel northward. The increased northward surge in N1 propagates further into the Gulf of Mexico, leading to the observed differences between the two configurations.

This effect of the barrier island can also be observed in Figure 4. At station 9, located on the barrier island, N1 produces a higher storm surge peak than G1, resulting in simulated peak water levels closer to the observations. Similarly, at station 4, situated north of the barrier island, N1's peak water levels resemble better the observed peak than G1.

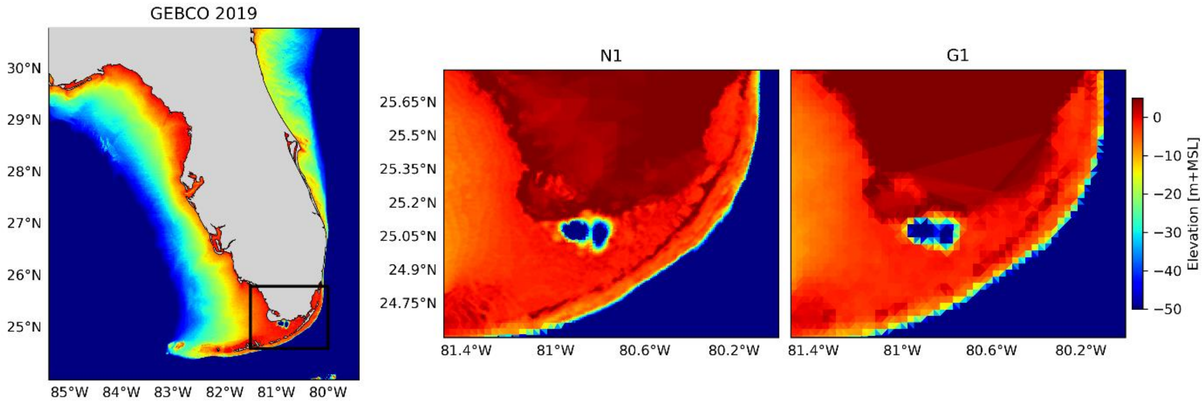

Figure A11. Left: GEBCO2019 for the study area, black rectangle shows the barrier island region from the middle and right panels. Middle: Bathymetry in the barrier island interpolated to the grid of the model configuration N1. Right: Bathymetry in the barrier island interpolated to the grid of the model configuration G1.

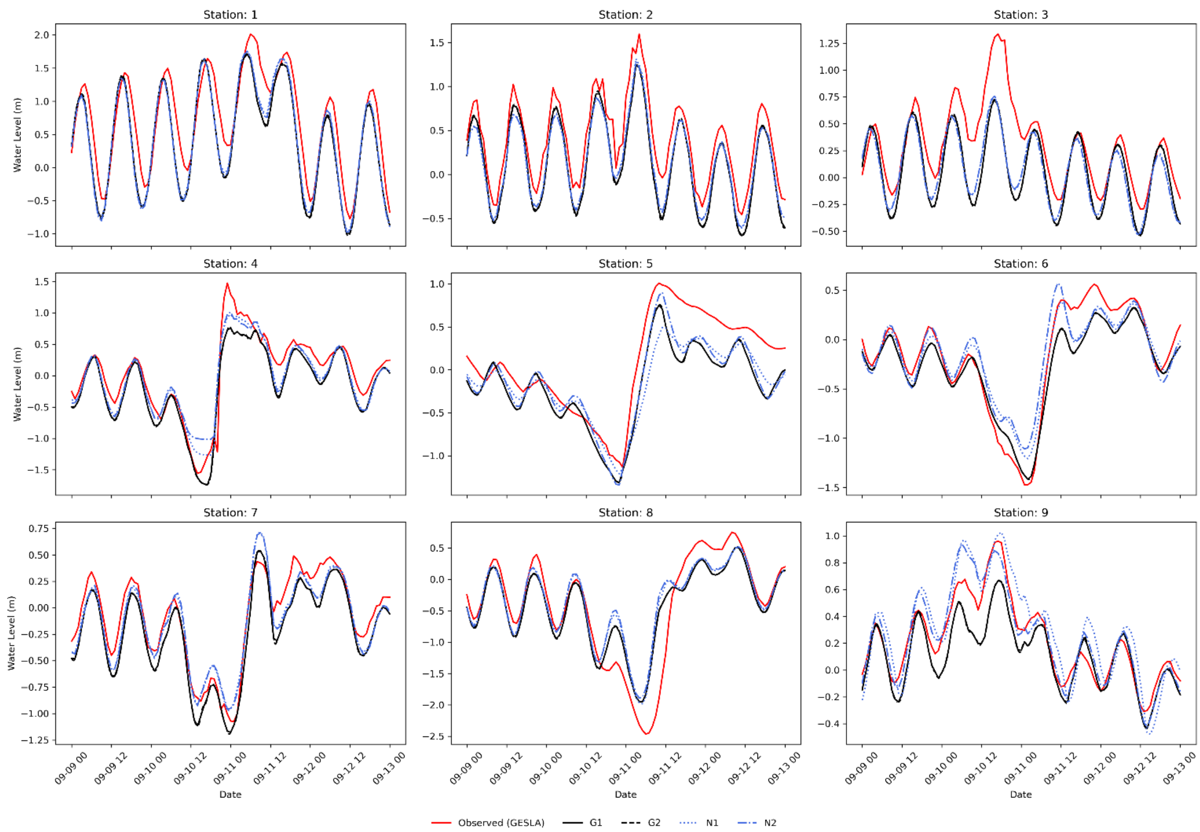

Figure 4. Validation of water levels for the case study Irma, for the nine tide gauge stations depicted in Fig. 3.

We have updated the manuscript to clarify the key aspects discussed above:

(Lines 271 – 287): *"For TC Irma (Fig. 8 panel c), the nesting of a local model at high-resolution with GEBCO2019 results in maximum water levels that are up to 0.3 m higher than G1 in the southwest of Florida. These differences between N1 and G1 gradually increase over time and are maximum at the peak of TC Irma (Fig. A10). While higher grid resolution affects tidal propagation mainly along the coast of Florida (Fig. A6 and Figure A7), storm surge propagation is more sensitive to the used bathymetry (Fig. A8 and Figure A9). High resolution is needed in areas with steep bathymetry. In contrast to the coarser grid of G1, N1 better resolves complex topographic features around the barrier islands (Fig. A11), allowing water to flow more freely through these barriers. At timestep 10-09-2017 in Figure A10, when there is a negative surge north of the barrier island, G1 produces higher water levels because water remains trapped in the north. Conversely, during the peak of TC Irma, on the 11-09-2017, the water levels in G1 are lower than N1 because less water is able to travel northwards. The increased northward surge of N1 propagates further into the Gulf of Mexico, leading to higher water levels that also propagate further into the Gulf of Mexico (see Figure A10). Water levels for nine tide gauge stations along the coast indicate that while G1 underestimates the peak of TC Irma in most locations (Fig. A2, all stations but station 7), N1 simulates on average higher peaks, resulting sometimes in overestimations (Fig. A2, station 9). The improved resolution of topographic features in the barrier island region allows stations nearby (Fig. A2, stations 4 and 9) to better capture the event's peak compared to G1. Additionally, the performance of N1 is slightly better than G1 for six tide gauge stations (stations 1-6), as reflected in Table A1, which shows lower RMSE values. However, for stations 7-9, G1 shows slightly higher RMSE and Pearson's correlation."*

Furthermore, the characterization of the Hurricane Irma results as "good" is questionable given an RMSE of 40 cm. I strongly recommend that you address these points with additional analysis and provide a clearer, evidence-based justification in the manuscript.

We acknowledge the reviewer's concern regarding the RMSE of 40 cm for one station of Hurricane Irma. However, we would like to clarify the overall model performance. The Pearson's correlation for TC Irma across all model configurations is around 0.9, demonstrating that the model captures the hydrographs of the event well. Additionally, the average RMSE across all stations is 0.28 m for model configuration G1, which aligns with previous large-scale studies that have reported RMSE values of storm events ranging from 0.24 m to 0.31 m (Gori et al., 2023; Marsooli & Lin, 2018; Vogt et al., 2024). It is important to note that these studies also present RMSE averaged over all stations, rather than individual stations.

The specific case of the Fernandina Beach station (station 1 in the manuscript), where the RMSE reaches 0.41 m, is consistent with the results from Leijnse et al. (2021), who simulated water levels using a nested local model for Jacksonville within a regional model of Florida. While their study did not explicitly report RMSE values, visual comparisons suggest similar performance of their SFINCS and Delft 3D models, as illustrated in the figure below.

A key factor influencing the model's performance is its ability to predict the peak water levels of the storm. As observed in Figure 4, the model tends to underestimate water levels around the peak of the event. This is primarily due to the use of ERA5 forcing data, which do not resolve tropical cyclones with the same level of detail as higher-resolution meteorological datasets. Additionally, the model does not account for wave contributions, which play a significant role in southern Florida. For instance, at Station 3, significant wave heights during TC Irma have been modeled to exceed 10 m (Xian et al., 2018), and the absence of wave contributions in our water levels can explain those differences. Furthermore, the location of certain stations affects the validations results. Many of the stations with higher RMSE values are situated behind barrier islands or semi-enclosed bays, areas where GTSM is known to have limitations in resolving water levels (Muis et al., 2019).

Overall, while some stations show higher RMSE values, our results are within the range of similar large-scale studies. Given the high correlation of approximately 0.9, the model successfully captures the event's overall dynamics, and differences in the peaks could be further improved by users of MOSAIC when updating the meteorological forcing or including wave dynamics. However, validation of stations located in places with difficult topography will remain challenging.

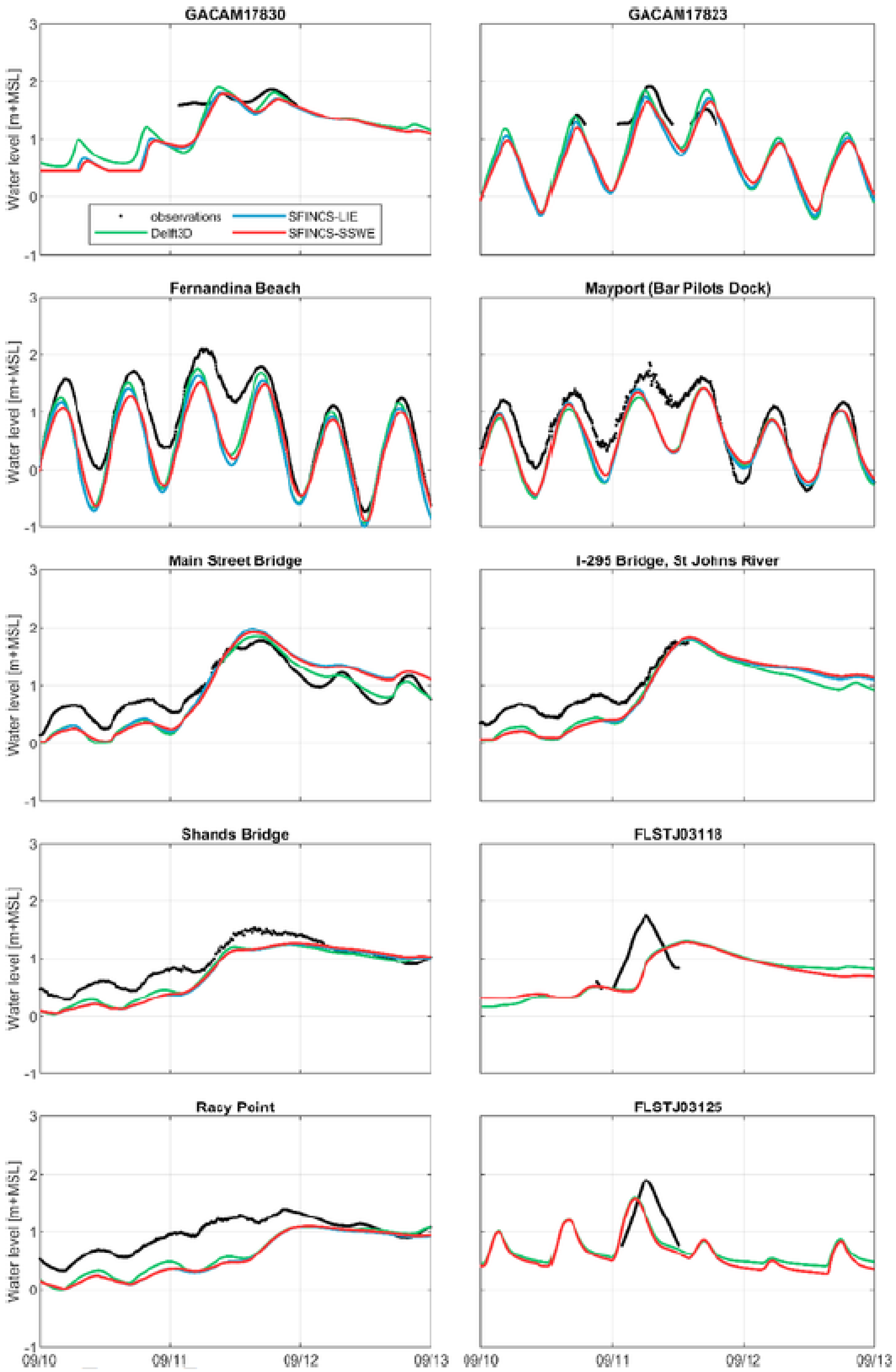

Figure from Leijnse et al., 2021. The hydrograph of Fernandina Beach modelled by Leijnse et al., 2021 with Delft3D and SFINCS shows very similar performance to that we present in Figure 4 above (station 1).

We have updated the manuscript to clarify the validation of our results:

(Lines 143 – 150): "*Additionally, TC Irma has an average RMSE of 0.28 m with a standard deviation of 0.09 m. ETC Xynthia has a RMSE of 0.22 m with a standard deviation of 0.08 m. The stations performing less well are those located in enclosed harbours or behind the barrier islands. The RMSE values of GTSM for both storms show results comparable to other large-scale studies that have used hydrodynamic models to simulate storm tides of storm events. Marsooli and Lin (2018) and Gori et al. (2023), for example, used the ADvanced CIRCulation model (ADCIRC) to simulate storm tides with an average RMSE over stations of 0.31 and 0.29 m, respectively. Vogt et al. (2024) used the GeoCLaw solver and reported an average RMSE of 0.24 m over 213 tide gauge stations, but with a Pearson's correlation of 0.5, showing less good agreement with observed storm tides than the MOSAIC model setup presented in this study.*"

5. References:

I have noticed a recurring issue with references that needs to be addressed. While I appreciate that some of my suggested references were incorporated, the broader review of citations appears incomplete. In several cases, the citations remain incorrect or overly reliant on self-publication. For example, the description of van Ormondt et al. (2020) as solely addressing a tsunami is inaccurate. As noted in the abstract, the paper discusses three case studies: tides in the North Sea, storm surge and wave modeling under tropical cyclone conditions, and tsunami simulation. This misrepresentation suggests insufficient engagement with the cited literature. Such oversights undermine the scientific rigor of the manuscript. A thorough review of all references is essential to ensure accurate representation and a meaningful engagement with the broader scientific community. Without these improvements, it is difficult to support the manuscript in its current form.

Thank you for your feedback. We appreciate your observation regarding the citations and we have carefully reviewed and updated them. On the other hand, flood modelling is a very active research field with numerous regional studies and it is impossible (nor the aim of this paper) to provide a complete overview, so we have attempted to include the most important key papers of the coastal modelling community that are relevant to our work. In particular, we have incorporated additional findings and references from other research groups throughout the introduction, methods, results and discussion sections to ensure a more comprehensive and accurate representation of the current state of the art.

Below we include a list of updated references:

Alfieri, L., Bisselink, B., Dottori, F., Naumann, G., de Roo, A., Salamon, P., Wyser, K., Feyen, L., 2017. Global projections of river flood risk in a warmer world. Earths Future 5, 171–182. https://doi.org/10.1002/2016EF000485

Barnard, P.L., Befus, K.M., Danielson, J.J., Engelstad, A.C., Erikson, L.H., Foxgrover, A.C., Hayden, M.K., Hoover, D.J., Leijnse, T.W.B., Massey, C., McCall, R., Nadal-Caraballo, N.C., Nederhoff, K., O'Neill, A.C., Parker, K.A., Shirzaei, M., Ohenhen, L.O., Swarzenski, P.W., Thomas, J.A., van Ormondt, M., Vitousek, S., Vos, K., Wood, N.J., Jones, J.M., Jones, J.L., 2025. Projections of multiple climate-related coastal hazards for the US Southeast Atlantic. Nat. Clim. Change 15, 101–109. https://doi.org/10.1038/s41558-024-02180-2

Barnard, P.L., Erikson, L.H., Foxgrover, A.C., Hart, J.A.F., Limber, P., O'Neill, A.C., van Ormondt, M., Vitousek, S., Wood, N., Hayden, M.K., Jones, J.M., 2019. Dynamic flood modeling essential to assess the coastal impacts of climate change. Sci. Rep. 9, 4309. https://doi.org/10.1038/s41598-019-40742-z

Camus, P., Mendez, F.J., Medina, R., 2011. A hybrid efficient method to downscale wave climate to coastal areas. Coast. Eng. 58, 851–862. https://doi.org/10.1016/j.coastaleng.2011.05.007

Gori, A., Lin, N., Schenkel, B., Chavas, D., 2023. North Atlantic Tropical Cyclone Size and Storm Surge Reconstructions From 1950-Present. J. Geophys. Res. Atmospheres 128, e2022JD037312. https://doi.org/10.1029/2022JD037312
Gori, A., Lin, N., Smith, J., 2020. Assessing Compound Flooding From Landfalling Tropical Cyclones on the North Carolina Coast. Water Resour. Res. 56, e2019WR026788. https://doi.org/10.1029/2019WR026788
Hirabayashi, Y., Alifu, H., Yamazaki, D., Imada, Y., Shiogama, H., Kimura, Y., 2021. Anthropogenic climate change has changed frequency of past flood during 2010-2013. Prog. Earth Planet. Sci. 8, 36. https://doi.org/10.1186/s40645-021-00431-w
Hodges, K., Cobb, A., Vidale, P.L., 2017. How Well Are Tropical Cyclones Represented in Reanalysis Datasets? https://doi.org/10.1175/JCLI-D-16-0557.1
Islam, M.R., Lee, C.-Y., Mandli, K.T., Takagi, H., 2021. A new tropical cyclone surge index incorporating the effects of coastal geometry, bathymetry and storm information. Sci. Rep. 11, 16747. https://doi.org/10.1038/s41598-021-95825-7
Marsooli, R., Lin, N., 2018. Numerical Modeling of Historical Storm Tides and Waves and Their Interactions Along the U.S. East and Gulf Coasts. J. Geophys. Res. Oceans 123, 3844–3874.
Murakami, H., 2014. Tropical cyclones in reanalysis data sets. Geophys. Res. Lett. 41, 2133–2141. https://doi.org/10.1002/2014GL059519
Nederhoff, K., Leijnse, T.W.B., Parker, K., Thomas, J., O'Neill, A., van Ormondt, M., McCall, R., Erikson, L., Barnard, P.L., Foxgrover, A., Klessens, W., Nadal-Caraballo, N.C., Massey, T.C., 2024. Tropical or extratropical cyclones: what drives the compound flood hazard, impact, and risk for the United States Southeast Atlantic coast? Nat. Hazards 120, 8779–8825. https://doi.org/10.1007/s11069-024-06552-x
NOAA, 2014. Continuously Updated Digital Elevation Model (CUDEM) - Ninth Arc-Second Resolution Bathymetric-Topographic Tiles files [WWW Document]. URL https://coast.noaa.gov/htdata/raster2/elevation/NCEI_ninth_Topobathy_2014_8483/ (accessed 2.3.25).
NOAA National Geophysical Data Center, 2001. U.S. Coastal Relief Model Vol.3 - Florida and East Gulf of Mexico [WWW Document]. https://doi.org/10.7289/V5W66HPP
Slager, K., Burzel, A., Bos, E., De Bruikn, K., Wagenaar, D.J., Winsemius, H.C., 2016. User Manual Delft-FIAT version 1.
Thomas, S.R., Nicolau, S., Martínez-Alvarado, O., Drew, D.J., Bloomfield, H.C., 2021. How well do atmospheric reanalyses reproduce observed winds in coastal regions of Mexico? Meteorol. Appl. 28, e2023. https://doi.org/10.1002/met.2023
Vogt, T., Treu, S., Mengel, M., Frieler, K., Otto, C., 2024. Modeling surge dynamics improves coastal flood estimates in a global set of tropical cyclones. Commun. Earth Environ. 5, 1–19. https://doi.org/10.1038/s43247-024-01707-x
Wang, X., Verlaan, M., Veenstra, J., Lin, H.X., 2022b. Data-assimilation-based parameter estimation of bathymetry and bottom friction coefficient to improve coastal accuracy in a global tide model. Ocean Sci. 18, 881–904. https://doi.org/10.5194/os-18-881-2022
Wing, O.E.J., Bates, P.D., Quinn, N.D., Savage, J.T.S., Uhe, P.F., Cooper, A., Collings, T.P., Addor, N., Lord, N.S., Hatchard, S., Hoch, J.M., Bates, J., Probyn, I., Himsworth, S., Rodríguez González, J., Brine, M.P., Wilkinson, H., Sampson, C.C., Smith, A.M., Neal, J.C., Haigh, I.D., 2024. A 30 m Global Flood Inundation Model for Any

Climate Scenario. Water Resour. Res. 60, e2023WR036460. https://doi.org/10.1029/2023WR036460

I hope this feedback helps guide your revisions. I look forward to seeing the manuscript further refined and strengthened.