Response to Reviewer#2 Comments

The authors have addressed most of the concerns raised in the initial review, and the manuscript has improved significantly.

Point 1: However, regarding Point 2, while the authors added an extended explanation about the differences between single-site models and their proposed model, much of the text is verbose and does not add meaningful insights. It still fails to clarify what preprocessing or modeling steps make the model sensitive to geographic information. Additionally, the term "regional relationship" is unclear. That said, for a data-focused article, it is not strictly necessary to explain these points in detail. I recommend either deleting the redundant text (around line 140) or providing a simple, clear explanation. The manuscript can then be accepted.

Response:

Thanks for your constructive suggestion! In this version we deleted the redundant text and added a flowchart for the modeling processes of our model, hope this revision could be readable:

2.5 All-site Modeling Framework

Full details of the ASM can be found at Yang et al.(2023). Here, a brief description of its modeling processes is provided. Assuming there are six available TGs within 45°S to 45°N (Fig. 3(a)):

(1) Obtaining predictors (Fig. 3(b)). Four atmospheric data (mslp, u10, v10, and t2m) for each TG station are extracted from the ERA5 dataset through linear interpolation. Changes in sea level pressure and wind are the main factors in generating SSs (Woodworth et al., 2019); adding temperature variations considers the effects of thermal expansion and contraction. Meanwhile, following Yang et al.(2023) and Yang et al. (2024a), another three variables (longitude, latitude, and timestamp) are considered since geographical locations and record lengths of TGs are different. Hence, the predictor matrix for each TG consists of 7 columns: mslp, u10, v10, t2m, longitude, latitude, and time;

(2) All-site modeling (Fig. 3(c)). Predictor matrices and SSs of all six TG stations are stacked into one predictor matrix and one SS matrix. Then, the eXtreme Gradient Boosting Tree (XGBoost) (Chen & Guestrin, 2016) is used to learn the relationship between these two matrices. The XGBoost is a residual machine learning model that generates a new decision tree using SS residuals from the previous tree. Therefore, the new tree will pay more attention to training where the residual errors are significant, making it suitable for modeling SS extremes;

(3) Reconstruction (Fig. 3(d)). SSs can be estimated for any target node along the coastline by inputting the corresponding predictor matrix of that location into the model established in step (2).

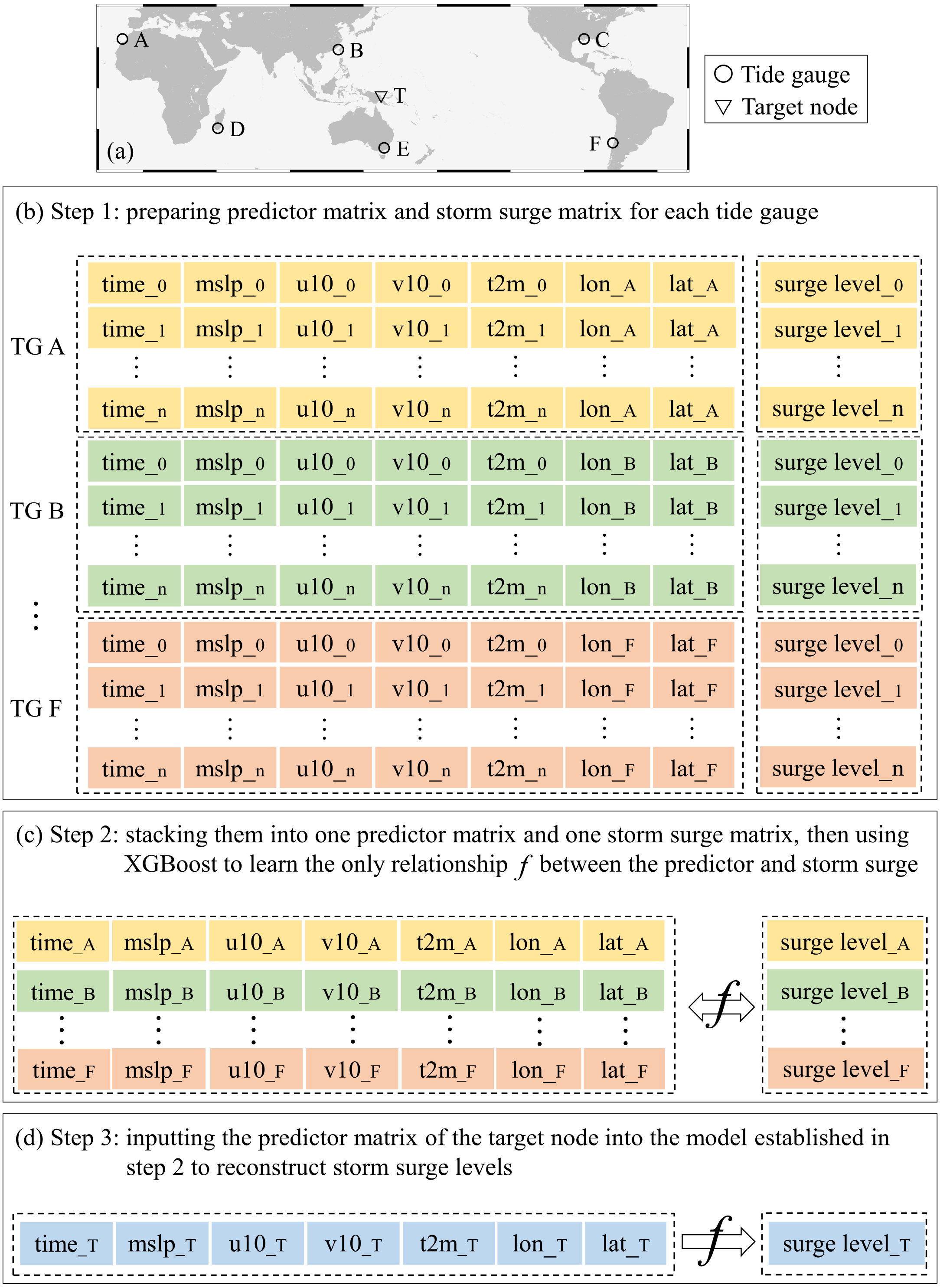


Figure 3: The modeling processes of the ASM framework

Point 2: Additionally, please add spaces before "(m)" in Fig. 4 for consistency.

Response:

Thanks for reminding. The following figure was the adjusted version:

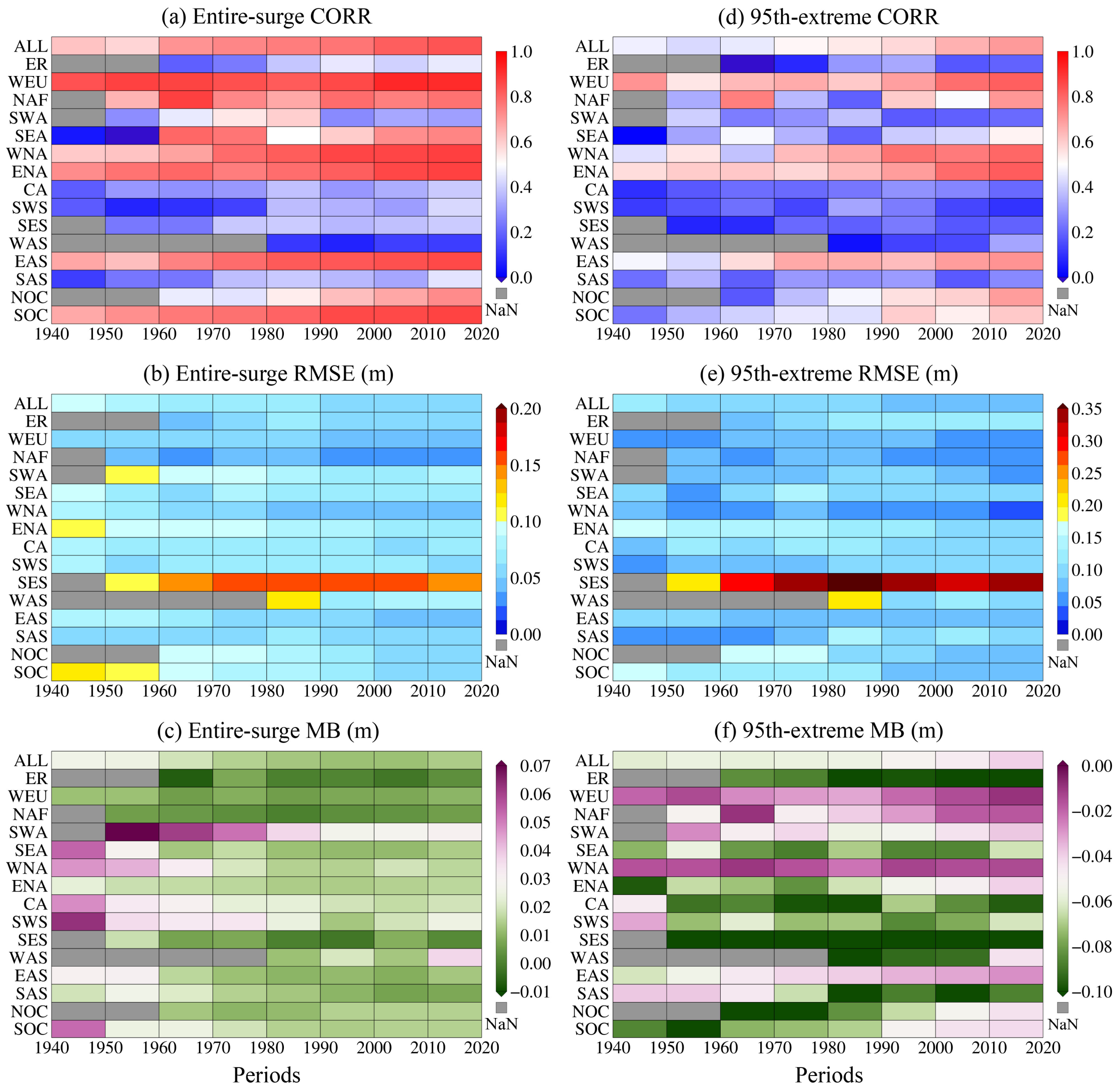


Figure 5: Temporal variations of the ASM model's precision at tide gauges from 1940 to 2020. (a-c) Entire surge evaluation statistics for different regions every 10 years; (d-f) 95th extreme evaluation statistics for different regions every 10 years.

Response to Reviewer#3 Comments

After preparing my comments, I looked at other reviewers' comments and I believe that further clarifications are needed.

Point 1: Given the long period considered, the authors need to critically explain how changes in the observation-based datasets used can affect their results. Are there any spurious changes in predictors to be accounted for? The discussion of this aspect is not detailed and remains speculative. The suggestion is to present some examples not from the training split.

Response:

Thanks for the recommendation. From Yang et al., (2024), the evaluation showed that the precision of the data-driven model is affected by the spatial resolution of TGs.

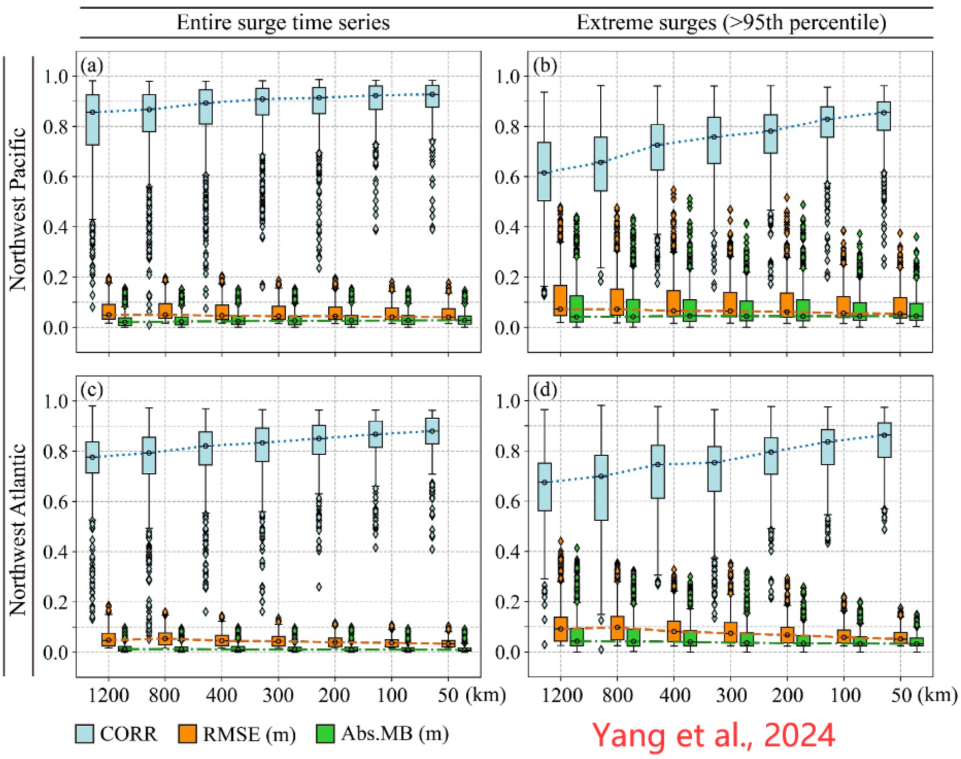
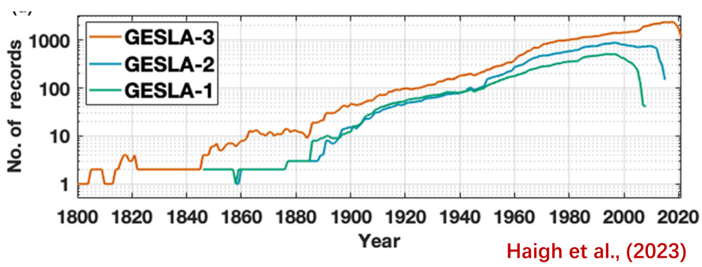


Figure 1. Evaluation statistics of estimated surges along coastlines of the (a, b) northwest Pacific and (c, d) northwest Atlantic Oceans for different virtual permanent tide gauge resolution schemes.

From Haigh et al., (2023), it can be seen that the number of TGs in GESLA-3 was increasing from 1940 to 2020, which means the spatial resolution of TGs was increasing. This can partly explain why "the overall precision (i.e., for ALL TGs) of entire surges and 95th extremes gradually increased from 1940 to 2020." in line 191.



For ERA5 atmospheric data, on the one hand, satellite observations were used in assimilation after the 1970s; on the other hand, the volume of satellite data has increased significantly in recent years (Soci et al., 2024, Figure 2). These developments improve the quality of ERA5 data and hence benefit the data-driven model. As for spurious changes in predictors, it is difficult to evaluate since ERA5 is the best atmospheric product now.

To avoid confusion, in this version, we rewrote the relevant statement in lines 191-195: Results indicate that the overall precision (i.e., for ALL TGs) of entire surges and 95th extremes gradually increased from 1940 to 2020. Possible reasons are as follows: on the one hand, the ASM model is affected by the spatial resolution of TGs (Yang et al., 2024). The increase of TGs in recent decades (Haigh et al., 2023) enhances its precision; on the other hand, the quality of ERA5 reanalysis data improved as increasing satellite data has been assimilated since the 1970s (Soci et al., 2024), which benefits the data-driven model.

Haigh, I. D., Marcos, M., Talke, S. A., Woodworth, P. L., Hunter, J. R., Hague, B. S., Arns, A., Bradshaw, E., and Thompson, P.: GESLA Version 3: A major update to the global higher-frequency sea-level dataset, Geoscience Data Journal, 10, 293–314, https://doi.org/10.1002/gdj3.174, 2023.

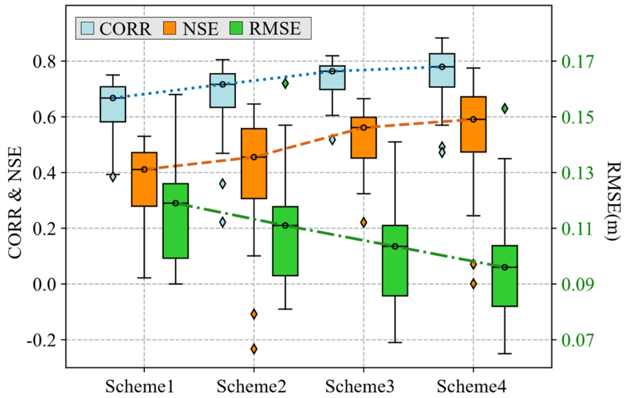
Soci, C., Hersbach, H., Simmons, A., Poli, P., Bell, B., Berrisford, P., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Radu, R., Schepers, D., Villaume, S., Haimberger, L., Woollen, J., Buontempo, C., and Thépaut, J.: The ERA5 global reanalysis from 1940 to 2022, Quart J Royal Meteoro Soc, qj.4803, https://doi.org/10.1002/qj.4803, 2024.

Yang, L., Jin, T., and Jiang, W.: Improving Coastal Storm Surge Monitoring Through Joint Modeling Based on Permanent and Temporary Tide Gauges, Geophysical Research Letters, 51, e2024GL108886, https://doi.org/10.1029/2024GL108886, 2024.

Point 2: The description of the model is sketchy, and the choice of predictors should be explained. Other variables may be relevant, but how this choice was made is not clear to me. Contribution of each variable should also be documented with some tailored experiment. Moreover, results seem to indicate a strong dependence on the availability of long time series (maps of Fig 3). Care should be taken to avoid data leakage.

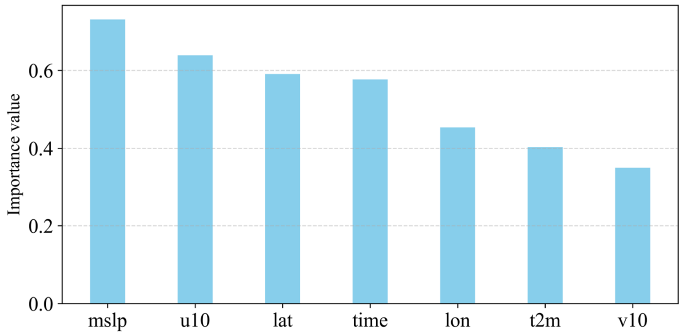
Response:

Thanks for your suggestion. The atmospheric predictors are commonly used by existing studies (Bruneau et al., 2020; Ebel et al., 2024; Tadesse et al., 2020; Tiggeloven et al., 2021). As for the longitude, latitude, and time, we tested them before proposing the all-site modeling framework in 2023. We found that adding these three parameters improved the reconstruction precision:



Scheme 1: atmospheric predictors
Scheme 2: atmospheric predictors + longitude + latitude
Scheme 3: atmospheric predictors + time
Scheme 4: atmospheric predictors + longitude + latitude + time

We also evaluated their contributions through the Permutation Importance analysis:



Bruneau, N., Polton, J., Williams, J., and Holt, J.: Estimation of global coastal sea level extremes using neural networks, Environ. Res. Lett., 15, 074030, https://doi.org/10.1088/1748-9326/ab89d6, 2020.

Ebel, P., Victor, B., Naylor, P., Meoni, G., Serva, F., and Schneider, R.: Implicit Assimilation of Sparse In Situ Data for Dense & Global Storm Surge Forecasting, in: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 471–480, https://doi.org/10.1109/CVPRW63382.2024.00052, 2024.

Tadesse, M., Wahl, T., and Cid, A.: Data-Driven Modeling of Global Storm Surges, Front. Mar. Sci., 7, 260, https://doi.org/10.3389/fmars.2020.00260, 2020.

Tiggeloven, T., Couasnon, A., van Straaten, C., Muis, S., and Ward, P. J.: Exploring deep learning capabilities for surge predictions in coastal areas, Sci Rep, 11, 17224, https://doi.org/10.1038/s41598-021-96674-0, 2021.

.

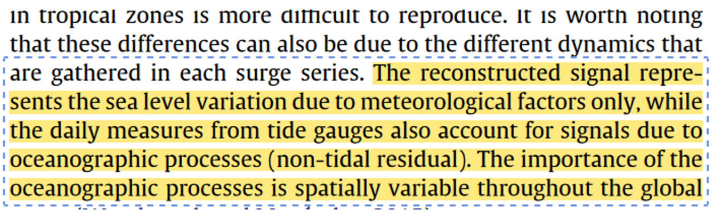
Point 3: The Authors need to make clear what is the advantage of their dataset compared to what is already public (e.g. https://cds.climate.copernicus.eu/datasets/sis-water-level-change-timeseries-cmip6?tab=overview). For this they should include analysis of selected globally distributed examples, such as heavy impact events across their model, observations, and other datasets. This could clarify which other phenomena affect surges, as hinted in the text but not clarified.

Response:

The numerical storm surge data we used was obtained from the website mentioned by the reviewer. The advantage of our dataset was clarified in this paper: 1) longer than the numerical dataset; 2) the precision is better. The extreme events we selected using the 95th percentile are, in a sense, what the reviewer termed "heavy impact events." As for which other phenomena affect surges, experts have discussed and analyzed them in detail. For example, Idier et al., (2019) and Woodworth et al., (2019).

To avoid confusion, in this version we clarified in lines 217-219: At the regional scale (Fig. 6(d-i)), ASM and GTSM perform poorly in areas with no tropical cyclones (ER, SWA, SWS, and SES), indicating that in addition to meteorological factors, oceanographic processes in these regions also contribute to the extremes (Cid et al., 2017; Woodworth et al., 2019).

A similar statement was mentioned by Cid et al., (2017) as well:


in tropical zones is more difficult to reproduce. It is worth noting that these differences can also be due to the different dynamics that are gathered in each surge series. The reconstructed signal represents the sea level variation due to meteorological factors only, while the daily measures from tide gauges also account for signals due to oceanographic processes (non-tidal residual). The importance of the oceanographic processes is spatially variable throughout the global

Cid, A., Camus, P., Castanedo, S., Méndez, F. J., and Medina, R.: Global reconstructed daily surge levels from the 20th Century Reanalysis (1871–2010), Global and Planetary Change, 148, 9–21, https://doi.org/10.1016/j.gloplacha.2016.11.006, 2017.

Idier, D., Bertin, X., Thompson, P., and Pickering, M. D.: Interactions Between Mean Sea Level, Tide, Surge, Waves and Flooding: Mechanisms and Contributions to Sea Level Variations at the Coast, Surv Geophys, 40, 1603–1630, https://doi.org/10.1007/s10712-019-09549-5, 2019.

Woodworth, P. L., Melet, A., Marcos, M., Ray, R. D., Wöppelmann, G., Sasaki, Y. N., Cirano, M., Hibbert, A., Huthnance, J. M., Monserrat, S., and Merrifield, M. A.: Forcing Factors Affecting Sea Level Changes at the Coast, Surv Geophys, 40, 1351–1397, https://doi.org/10.1007/s10712-019-09531-1, 2019.

Point 4: I am also missing references to relevant works, such as Nevo et al 2022 (https://hess.copernicus.org/articles/26/4013/2022/hess-26-4013-2022.html), Nearing et al. 2024 (https://www.nature.com/articles/s41586-024-07145-1), Ebel et al. 2024 (10.1109/CVPRW63382.2024.00052).

Response:

Thanks for reminding us. These papers are excellent works. They were added in line 56: …through multiple linear regression (Cid et al., 2018) or artificial intelligence (Nevo et al., 2022; Bruneau et al., 2020; Ebel et al., 2024; Nearing et al., 2024)

Ebel, P., Victor, B., Naylor, P., Meoni, G., Serva, F., and Schneider, R.: Implicit Assimilation of Sparse In Situ Data for Dense & Global Storm Surge Forecasting, in: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 471–480, https://doi.org/10.1109/CVPRW63382.2024.00052, 2024.

Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., Hassidim, A., Klotz, D., Kratzert, F., Metzger, A., Nevo, S., Pappenberger, F., Prudhomme, C., Shalev, G., Shenzis, S., Tekalign, T. Y., Weitzner, D., and Matias, Y.: Global prediction of extreme floods in ungauged watersheds, Nature, 627, 559–563, https://doi.org/10.1038/s41586-024-07145-1, 2024.

Nevo, S., Morin, E., Gerzi Rosenthal, A., Metzger, A., Barshai, C., Weitzner, D., Voloshin, D., Kratzert, F., Elidan, G., Dror, G., Begelman, G., Nearing, G., Shalev, G., Noga, H., Shavitt, I., Yuklea, L., Royz, M., Giladi, N., Peled Levi, N., Reich, O., Gilon, O., Maor, R., Timnat, S., Shechter, T., Anisimov, V., Gigi, Y., Levin, Y., Moshe, Z., Ben-Haim, Z., Hassidim, A., and Matias, Y.: Flood forecasting with machine learning models in an operational framework, Hydrol. Earth Syst. Sci., 26, 4013–4032, https://doi.org/10.5194/hess-26-4013-2022, 2022.

Point 5: The writing is imprecise in many places, as I reported in the comments by line. Please make sure to support every statement and avoid ambiguities. For example, the recurring phrase "their basic ideas differ" on models just means that you are using transfer learning in the ASM framework, right?

Response:

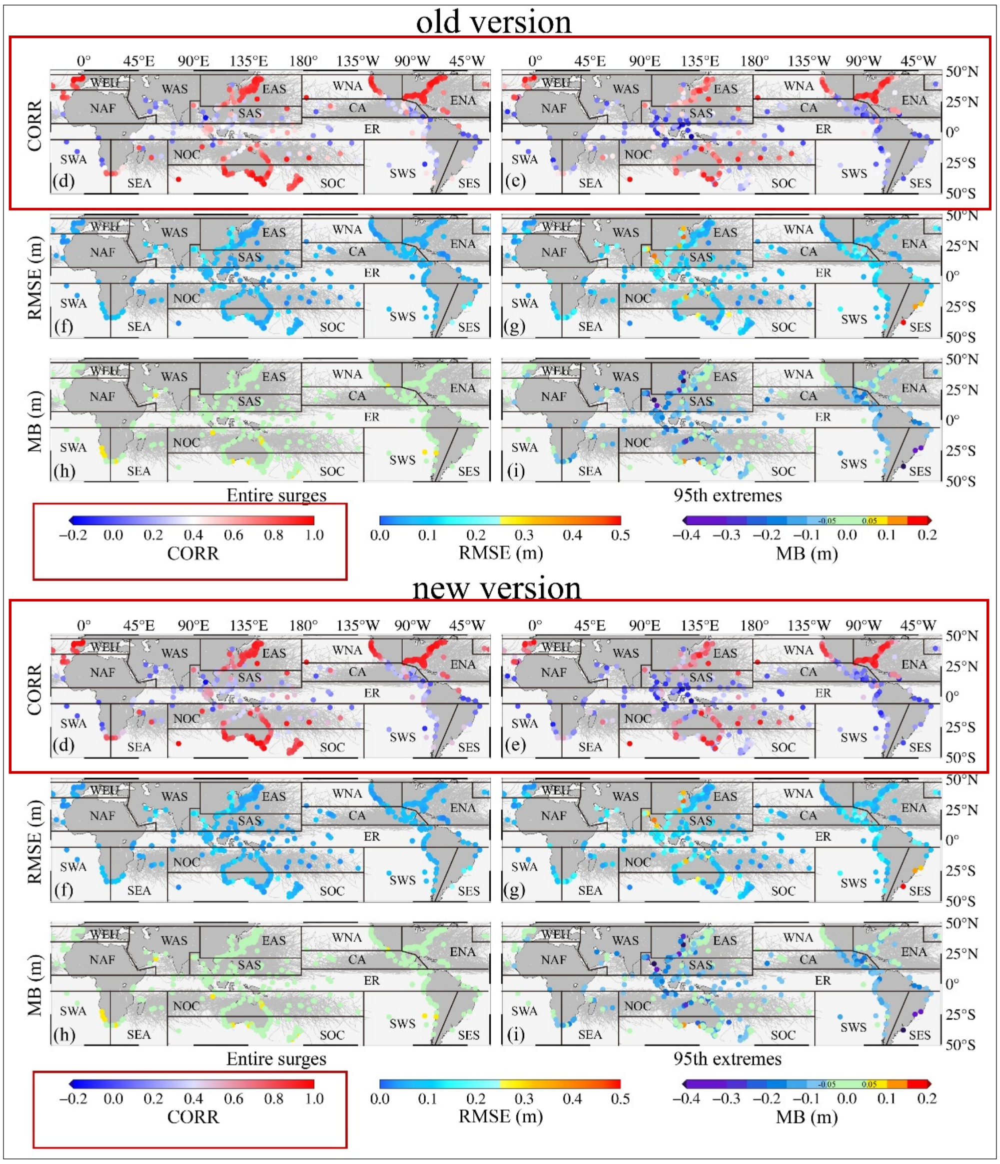
We are sorry for any difficulties in understanding caused by our writing issues. In this version we meticulously revised this manuscript based on the reviewer's suggestions, hoping this version can be more precise and readable.

To avoid ambiguities, in this version, we replaced the phrase "their basic ideas differ" in line 70 with "their modeling processes differ"

Point 6: Colors used in maps are not great. White for correlation is not visible (use a colormap with grey or yellow in the middle), and 0 should be at the middle. Colorbars seem adjusted, e.g. with a single color for near zero values or discontibuity. Please consult the journal specification for these aspects.

Response:

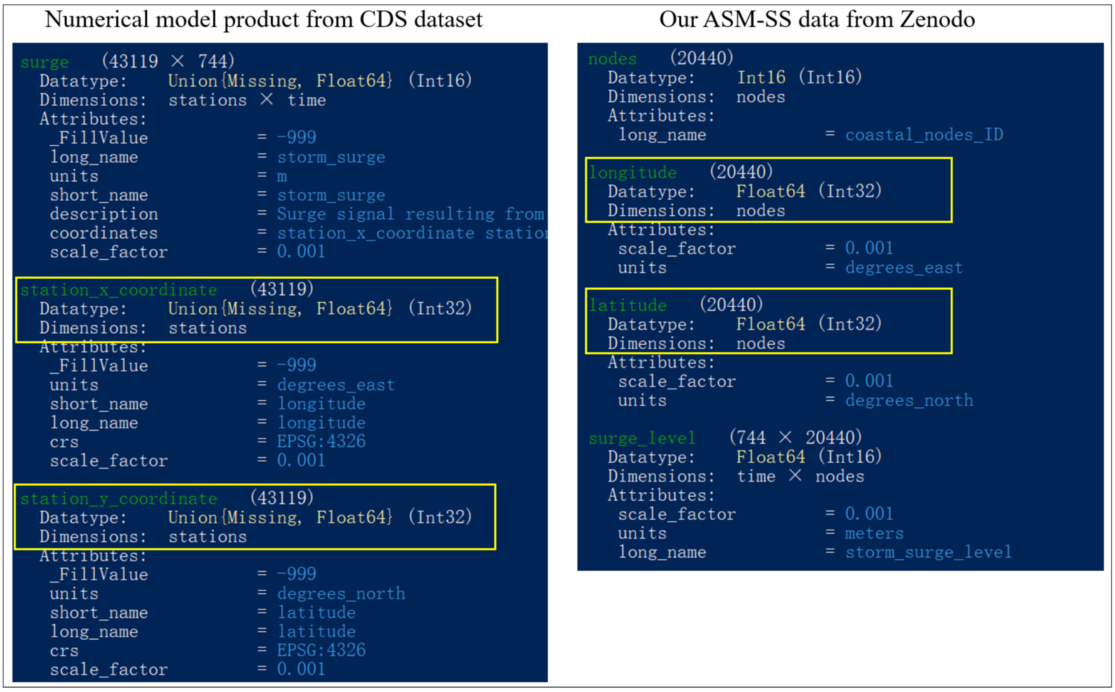
Thanks for your recommendation. In this version, we removed white in the middle of the correlation colorbar in Fig 4, 6, and 8 (differences please see the following figure), hoping this would make maps visible. Additionally, 0 is not at the middle because the data's range is not symmetrically distributed around zero. The case "with a single color for near zero" appears in mean bias (MB), since near-zero MBs (whether positive or negative) represent high precision, there's no need to use different colors to distinguish them.



Point 7: The data dimensions are "time" and "node", the latter to be mapped to static longitudes and latitudes. This is reasonable to save space but complicates access compared e.g. to gridded datasets.

Response:

Thanks for your suggestion. Since the nodes of our model are distributed along the coast, not in a grid, if it is stored in a grid format, there will be many areas without data. In addition, the data storage style (see the following figure) is consistent with the numerical dataset the reviewer mentioned in Point 3.



Specific comments:

Point 8: Line 10: I guess also yours can be considered a model

Response:

Thanks for your suggestion. Since both single-site modeling and all-site modeling belong to the data-driven *model*, in order to avoid confusion, we used "framework" in the paper we proposed the all-site modeling, so we continued to use it here.

Point 9: Line 11: I would not say this is a limitation of models. What is limiting is observational data to constrain them.

Response:

We are sorry for the unclear expression. This is indeed not a limitation of the numerical models themselves, but a limitation of the products released by the numerical models. In this version we rewrote this sentence in line 11 as: such global or quasi-global information could only be provided by global numerical models, while their simulation products span mainly the most recent decades.

Point 10: Line 13: What is a "node" here?

Response:

To avoid misunderstanding, we deleted this term here and changed this sentence in line 14 to: …we generated a high spatial resolution (10 km along the coastline) hourly SS dataset ASM-SS (all-site modeling storm surge) within 45°S to 45°N.

Point 11: Line 14: You should mention the data used for this, which I suspect it's ERA5

Response:

We are sorry for forgetting this information. In this version, this sentence was rewritten in line 13 as: "Using tide gauge records and European Centre for Medium-

Range Weather Forecasts Reanalysis 5 (ERA5) data, we generated a high spatial resolution…"

Point 12: Line 15: Which model?

Response:

We are sorry for the confusion. It was changed in the new version in line 16: …the precision of the ASM-SS model (medians of correlation coefficients, root mean square errors, and mean biases are…

Point 13: Line 57: This statement is misleading. Model performances will anyway depend on features such as bathymetry and grid resolution.

Response:

Data-driven models do not require grid like numerical models to resolve storm surges. As for bathymetry, in theory, this parameter should be considered in the data-driven model. However, to our knowledge, existing data-driven models for storm surge reconstruction rarely consider this parameter.

These two features (bathymetry and grid resolution) are not used in the modeling process, hence the precision of data-driven models is not affected by them.

Point 14: Line 66: This statement makes no sense to me. When constrained by observations, a surrogate model would work better than a free-running physical model. Or do you mean something else?

Response:

We are sorry for any difficulties in understanding. The surrogate model here is defined as "training the data-driven model with the storm surge outputs from numerical models". Therefore, the relationships learned by a surrogate model during the training process are only those expressed by the numerical model. In this case, when we input the same atmospheric data (such as mslp, u10, v10) into the surrogate model and the numerical model to generate hindcast storm surges, theoretically, the precision of the former can at most match but cannot surpass the latter. Hope this explanation could be helpful.

Point 15: Line 76: I don't see how this can be a general prerequisite

Response:

These are some examples:

(1) Sufficiently long and high-spatial resolution: for example, evaluating return levels of extreme storm surge (SS) through the extreme value analysis theory heavily relies on the length of records; the extrapolation of return periods should not be longer than four times the length of available time series (Pugh and Woodworth, 2014, Page 323). For example, at least 50-year SS data are needed to estimate 1 in 200-year SS levels since the estimation uncertainty will increase if the records are too short. In addition, the higher the spatial coverage of SS data, the more information for different places can be obtained.

(2) High-temporal resolution: the SSs caused by tropical cyclones vary quickly,

and cover frequencies from hours to days (WMO, 2011, Figure 1.2). If the time resolution is not enough, it is difficult to provide effective information to study the characteristics of SSs during tropical cyclones.

To avoid confusion, we changed the sentence "High spatiotemporal resolution and sufficiently long SS dataset is the basis for analyzing this disaster" to "High spatiotemporal resolution and sufficiently long SS dataset is important for better analyzing this disaster" in line 77.

Pugh, D. and Woodworth, P.: Sea-level science: understanding tides, surges, tsunamis and mean sea-level changes, Cambridge University Press, Cambridge ; New York, 395 pp., 2014.

WMO: Guide to storm surge forecasting, 2011 ed., World Meteorological Organization, Geneva, Switzerland, 2011.

Point 16: Line 78: I count two at line 76?

Response:

Apologies for this confusion. Since we think "high spatiotemporal resolution" contains two demands (i.e. space and time). In this version we changed the sentence in line 79 from "cannot fulfill all three demands simultaneously" to "cannot fulfill all demands simultaneously".

Point 17: Line 80: ET cyclones go way beyond that, check for example the IBTraCS dataset and expand/clarify

Response:

We are sorry for the inaccurate statement. This article focuses on the areas affected by tropical cyclones. We rewrote this sentence in line 81 from "most destructive tropical and extratropical cyclones" to "most destructive tropical cyclones". Similar inaccurate statements elsewhere in this article have also been deleted.

Point 18: Line 84: I am confused between "support for communities" and "our understanding"

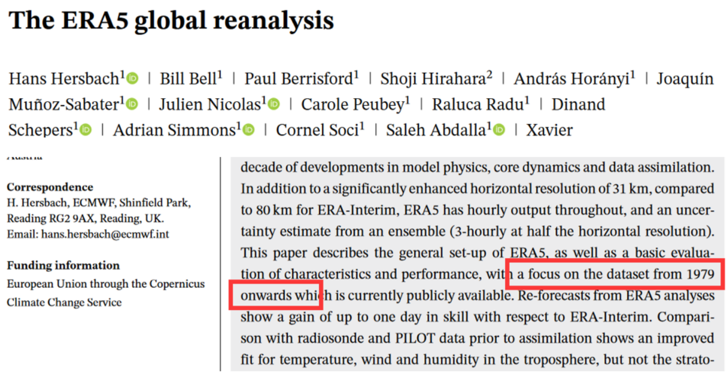
Response:

Apologies for this confusion. In this version we rewrote the sentence in line 85 as: for coastal communities to deepen the understanding of SSs and ESLs

Point 19: Line 88: The paper by Hersbach should be cited as well.

Response:

Thanks for reminding. Hersbach et al., (2020) focus on the dataset from 1979 onwards, the latest paper (Soci et al., 2024) covers the period from 1940-2022. Therefore, we only cited the latest paper.



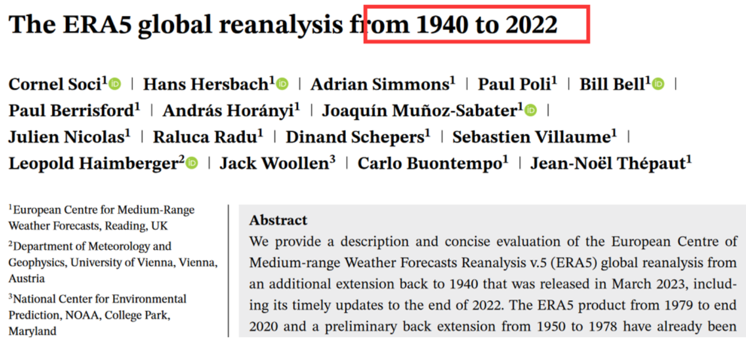
The ERA5 global reanalysis

Hans Hersbach[1] | Bill Bell[1] | Paul Berrisford[1] | Shoji Hirahara[2] | András Horányi[1] | Joaquín Muñoz-Sabater[1] | Julien Nicolas[1] | Carole Peubey[1] | Raluca Radu[1] | Dinand Schepers[1] | Adrian Simmons[1] | Cornel Soci[1] | Saleh Abdalla[1] | Xavier

Correspondence
H. Hersbach, ECMWF, Shinfield Park, Reading RG2 9AX, Reading, UK.
Email: hans.hersbach@ecmwf.int

Funding information
European Union through the Copernicus Climate Change Service

decade of developments in model physics, core dynamics and data assimilation. In addition to a significantly enhanced horizontal resolution of 31 km, compared to 80 km for ERA-Interim, ERA5 has hourly output throughout, and an uncertainty estimate from an ensemble (3-hourly at half the horizontal resolution). This paper describes the general set-up of ERA5, as well as a basic evaluation of characteristics and performance, with a focus on the dataset from 1979 onwards which is currently publicly available. Re-forecasts from ERA5 analyses show a gain of up to one day in skill with respect to ERA-Interim. Comparison with radiosonde and PILOT data prior to assimilation shows an improved fit for temperature, wind and humidity in the troposphere, but not the strato-



The ERA5 global reanalysis from 1940 to 2022

Cornel Soci[1] | Hans Hersbach[1] | Adrian Simmons[1] | Paul Poli[1] | Bill Bell[1] | Paul Berrisford[1] | András Horányi[1] | Joaquín Muñoz-Sabater[1] | Julien Nicolas[1] | Raluca Radu[1] | Dinand Schepers[1] | Sebastien Villaume[1] | Leopold Haimberger[2] | Jack Woollen[3] | Carlo Buontempo[1] | Jean-Noël Thépaut[1]

[1]European Centre for Medium-Range Weather Forecasts, Reading, UK
[2]Department of Meteorology and Geophysics, University of Vienna, Vienna, Austria
[3]National Center for Environmental Prediction, NOAA, College Park, Maryland

Abstract
We provide a description and concise evaluation of the European Centre of Medium-range Weather Forecasts Reanalysis v.5 (ERA5) global reanalysis from an additional extension back to 1940 that was released in March 2023, including its timely updates to the end of 2022. The ERA5 product from 1979 to end 2020 and a preliminary back extension from 1950 to 1978 have already been

Point 20: Line 90: "with...grids" is incorrect. There's just one grid.

Response:

Thanks for reminding. This sentence was changed in line 91 as: …with a 0.25°×0.25° resolution grid.

Point 21: Line 91: please elaborate why other wind variables, such as wind gusts, are not considered. Moreover, u and v are swapped compared to the bracket.

Response:

We are sorry for this mistake. In this version, we rewrote the sentence in line 92 "…10 m northward and eastward wind (u10, v10) …" to "…10 m eastward and northward wind (u10, v10) …"

Existing models all use these two parameters (e.g. Bruneau et al., 2020; Ebel et al., 2024; Tadesse et al., 2020; Tiggeloven et al., 2021), we haven't seen studies using wind gusts or other wind variables.

Bruneau, N., Polton, J., Williams, J., and Holt, J.: Estimation of global coastal sea level extremes using neural networks, Environ. Res. Lett., 15, 074030, https://doi.org/10.1088/1748-9326/ab89d6, 2020.

Ebel, P., Victor, B., Naylor, P., Meoni, G., Serva, F., and Schneider, R.: Implicit Assimilation of Sparse In Situ Data for Dense & Global Storm Surge Forecasting, in: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 471–480, https://doi.org/10.1109/CVPRW63382.2024.00052, 2024.

Tadesse, M., Wahl, T., and Cid, A.: Data-Driven Modeling of Global Storm Surges, Front.

Mar. Sci., 7, 260, https://doi.org/10.3389/fmars.2020.00260, 2020.
Tiggeloven, T., Couasnon, A., van Straaten, C., Muis, S., and Ward, P. J.: Exploring deep learning capabilities for surge predictions in coastal areas, Sci Rep, 11, 17224, https://doi.org/10.1038/s41598-021-96674-0, 2021.

Point 22: Line 105: "Datum" as in geodesy? Or data?

Response:

Datum. Caused by earthquakes or changes in instrument. In this version, we revised it in line 106 as: Datum jumps caused by earthquakes or changes in instrument were adjusted.

Point 23: Line 116: The title does not sound right

Response:

Thanks for reminding. In this version, we changed it in line 117 from "Numerical Model Surge" to "Surge Data Simulated from Numerical Model"

Point 24: Line 123: rather than "participate" I guess they "were not used"

Response:

Thanks for your suggestion. We rewrote it in this version in line 124: they were not used in the training process of the latter.

Point 25:

1) Line 138: please use a clearer symbol than "n"
2) Line 141: I doubt the ASM "believes", rephrase to avoid humanization
3) Line 145: It is unclear what relevance means here
4) Line 151: How the model is trained is unclear to me. What is the training period? How do you define train and validation splits? These details are required.

Response:

Based on the comments from two reviewers, we rewrote Section 2.5, deleted the redundant text, and added a flowchart for the ASM modeling, hope this revision could be clear and readable:

2.5 All-site Modeling Framework

Full details of the ASM can be found at Yang et al.(2023). Here, a brief description of its modeling processes is provided. Assuming there are six available TGs within 45°S to 45°N (Fig. 3(a)):

(1) Obtaining predictors (Fig. 3(b)). Four atmospheric data (mslp, u10, v10, and t2m) for each TG station are extracted from the ERA5 dataset through linear interpolation. Changes in sea level pressure and wind are the main factors in generating SSs (Woodworth et al., 2019); adding temperature variations considers the effects of thermal expansion and contraction. Meanwhile, following Yang et al.(2023) and Yang et al. (2024a), another three variables (longitude, latitude, and timestamp) are considered since geographical locations and record lengths of TGs are different. Hence, the predictor matrix for each TG consists of 7 columns: mslp, u10, v10, t2m, longitude,

latitude, and time;

(2) All-site modeling (Fig. 3(c)). Predictor matrices and SSs of all six TG stations are stacked into one predictor matrix and one SS matrix. Then, the eXtreme Gradient Boosting Tree (XGBoost) (Chen & Guestrin, 2016) is used to learn the relationship between these two matrices. The XGBoost is a residual machine learning model that generates a new decision tree using SS residuals from the previous tree. Therefore, the new tree will pay more attention to training where the residual errors are significant, making it suitable for modeling SS extremes;

(3) Reconstruction (Fig. 3(d)). SSs can be estimated for any target node along the coastline by inputting the corresponding predictor matrix of that location into the model established in step (2).

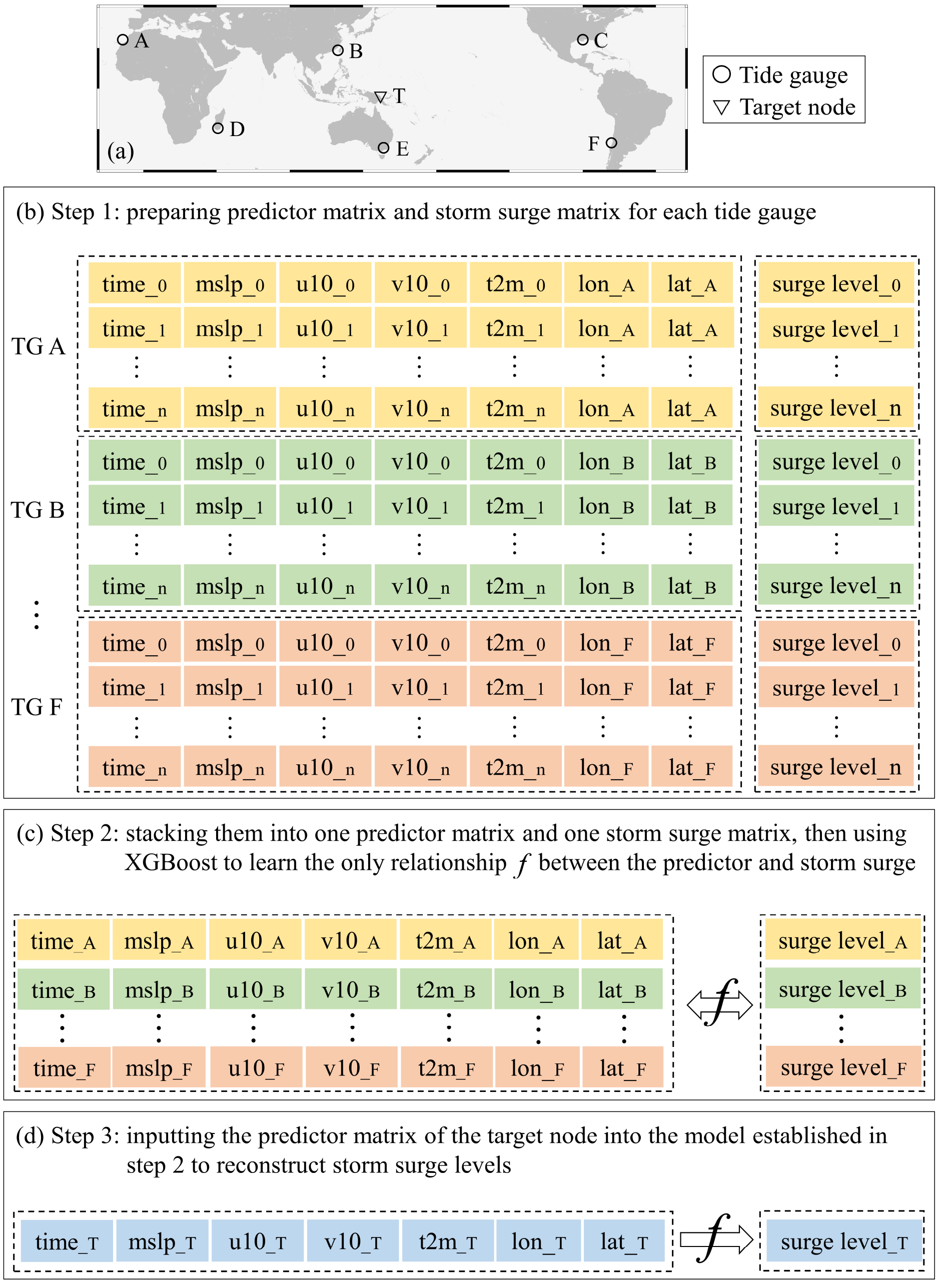


Figure 3: The modeling processes of the ASM framework

Point 26: Line 148: If you refer to seawater, I guess you should rather use skin temperature

Response:

Thanks for your suggestion. This parameter is consistent with Žust et al., (2021). As for whether the skin temperature is more appropriate than 2m temperature, we will discuss it in the future.

poral resolution. In this study, the following forecast fields were subset to the Adriatic basin, represented by a 73 × 57 spatial grid (see Fig. 2): (i) 10 m zonal and meridional winds, (ii) mean sea level pressure, and (iii) air temperature at 2 m. The forecasts were linearly interpolated to hourly time steps to match the SSH temporal resolution. Atmospheric fields

Žust et al., (2021)

Žust, L., Fettich, A., Kristan, M., and Ličer, M.: HIDRA 1.0: deep-learning-based ensembl e sea level forecasting in the northern Adriatic, Geosci. Model Dev., 14, 2057–2074, https://doi.org/10.5194/gmd-14-2057-2021, 2021.

Point 27: Line 174: according to whom the 95th percentile is a good metric?

Response:

The "Peaks Over Threshold" is a commonly used method in the extreme value analysis theory. Following Bruneau et al. (2020), as well as the articles (Tadesse et al., 2021, 2022) from Thomas Wahl's research group, we used the 95th percentile.

Bruneau, N., Polton, J., Williams, J., and Holt, J.: Estimation of global coastal sea level extremes using neural networks, Environ. Res. Lett., 15, 074030, https://doi.org/10.1088/1748-9326/ab89d6, 2020.

Tadesse, M. G. and Wahl, T.: A database of global storm surge reconstructions, Sci Data, 8, 125, https://doi.org/10.1038/s41597-021-00906-x, 2021.

Tadesse, M. G., Wahl, T., Rashid, M. M., Dangendorf, S., Rodríguez-Enríquez, A., and Talke, S. A.: Long-term trends in storm surge climate derived from an ensemble of global surge reconstructions, Sci Rep, 12, 13307, https://doi.org/10.1038/s41598-022-17099-x, 2022.

Point 28: Line 182: I don't think you ever mentioned the data source for cyclone tracks

Response:

Thanks for reminding. We added the source in Acknowledgments in line 305: The tropical cyclone paths shown in Fig. 4, 6, and 8 are from Gahtan et al. (2024).

Gahtan, J., Knapp, K. R., Schreck, C. J. I., Diamond, H. J., Kossin, J. P., and Kruk, M. C.: International Best Track Archive for Climate Stewardship (IBTrACS) Project, Version 4.01 [data set], https://doi.org/10.25921/82ty-9e16, 2024.

Point 29: Line 216: The information is a bit lost in the text, please consider adding a summary table

Response:

Thanks for your constructive suggestion. In this version, we added three summary tables for Figures 4, 6, and 8.

Table 1: The median of evaluation statistics for different regions in Fig. 4.

	Median of CORRs		Median of RMSEs (m)		Median of MBs (m)	
	Entire surges	95th extremes	Entire surges	95th extremes	Entire surges	95th extremes
ALL	0.78	0.59	0.063	0.094	0.014	-0.052
ER	0.39	0.20	0.054	0.120	0.002	-0.106
WEU	0.87	0.68	0.050	0.069	0.011	-0.020
NAF	0.76	0.55	0.036	0.059	0.004	-0.031
SWA	0.25	0.19	0.074	0.080	0.029	-0.046
SEA	0.61	0.43	0.070	0.105	0.012	-0.087
WNA	0.83	0.69	0.044	0.055	0.018	-0.019
ENA	0.84	0.75	0.073	0.117	0.016	-0.053
CA	0.30	0.19	0.072	0.116	0.027	-0.098
SWS	0.29	0.25	0.061	0.098	0.017	-0.088
SES	0.29	0.15	0.141	0.312	0.011	-0.303
WAS	0.09	0.17	0.091	0.131	0.016	-0.083
EAS	0.81	0.62	0.054	0.077	0.013	-0.037
SAS	0.34	0.24	0.060	0.107	0.012	-0.099
NOC	0.62	0.54	0.068	0.101	0.013	-0.056
SOC	0.83	0.53	0.064	0.093	0.017	-0.046

Table 2: The median of evaluation statistics for different regions in Fig. 6.

	Median of CORRs		Median of RMSEs (m)		Median of MBs (m)	
	ASM-GESLA	GTSM-GESLA	ASM-GESLA	GTSM-GESLA	ASM-GESLA	GTSM-GESLA
ALL	0.63	0.55	0.093	0.106	-0.050	-0.045
ER	0.20	0.10	0.122	0.075	-0.106	-0.026
WEU	0.72	0.62	0.066	0.071	-0.019	-0.001
NAF	0.53	0.46	0.057	0.044	-0.025	0.003
SWA	0.20	0.30	0.081	0.087	-0.048	-0.063
SEA	0.44	0.35	0.103	0.076	-0.087	-0.058
WNA	0.72	0.58	0.054	0.085	-0.019	-0.061
ENA	0.77	0.72	0.112	0.138	-0.052	-0.072
CA	0.21	0.19	0.116	0.122	-0.107	-0.106
SWS	0.25	0.14	0.098	0.105	-0.088	-0.086
SES	0.21	0.24	0.340	0.155	-0.329	-0.123
WAS	0.17	0.22	0.131	0.077	-0.083	-0.056
EAS	0.66	0.59	0.071	0.096	-0.036	-0.041
SAS	0.27	0.29	0.107	0.092	-0.099	-0.045
NOC	0.58	0.48	0.095	0.113	-0.057	0.017
SOC	0.57	0.47	0.088	0.102	-0.047	-0.010

Table 3: The median of evaluation statistics for different regions in Fig. 8.

	Median of CORRs		Median of RMSEs (m)		Median of MBs (m)	
	Entire surges	95th extremes	Entire surges	95th extremes	Entire surges	95th extremes
ALL	0.32	0.23	0.084	0.138	-0.056	-0.126
ER	0.19	0.09	0.085	0.127	-0.076	-0.123
WEU	0.89	0.70	0.058	0.090	-0.039	-0.073
NAF	0.28	0.10	0.060	0.118	-0.022	-0.112
SWA	0.37	0.12	0.068	0.060	0.030	-0.050
SEA	0.30	0.36	0.114	0.176	-0.105	-0.172
WNA	0.70	0.70	0.055	0.080	-0.025	-0.023
ENA	0.48	0.52	0.073	0.144	-0.009	-0.091
CA	0.38	0.28	0.063	0.098	-0.037	-0.083
SWS	0.30	0.09	0.043	0.060	-0.008	-0.044
SES	0.42	0.11	0.118	0.204	-0.059	-0.180
WAS	0.28	0.12	0.090	0.174	-0.040	-0.167
EAS	0.47	0.40	0.132	0.225	-0.065	-0.212
SAS	0.29	0.25	0.100	0.148	-0.083	-0.143
NOC	0.22	0.22	0.100	0.159	-0.074	-0.149
SOC	0.82	0.39	0.095	0.154	-0.068	-0.140

Point 30: Line 219: Indeed, please provide explanations.

Response:

Thanks for reminding. A similar conclusion was mentioned by Cid et al., (2017):

> in tropical zones is more difficult to reproduce. It is worth noting that these differences can also be due to the different dynamics that are gathered in each surge series. The reconstructed signal represents the sea level variation due to meteorological factors only, while the daily measures from tide gauges also account for signals due to oceanographic processes (non-tidal residual). The importance of the oceanographic processes is spatially variable throughout the global

Cid et al., 2017

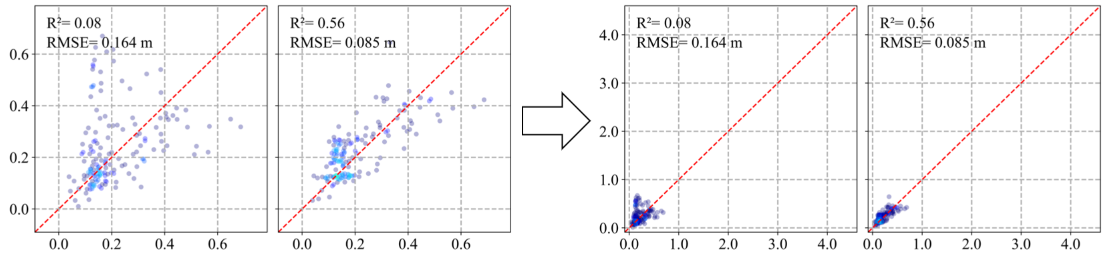
We rewrote the sentence "indicating that there may be additional contributions from other physical factors to the extremes" as "indicating that in addition to meteorological factors, oceanographic processes in these regions also contribute to the extremes (Cid et al., 2017; Woodworth et al., 2019)." in lines 218-219.

Cid, A., Camus, P., Castanedo, S., Méndez, F. J., and Medina, R.: Global reconstructed daily surge levels from the 20th Century Reanalysis (1871–2010), Global and Planetary Change, 148, 9–21, https://doi.org/10.1016/j.gloplacha.2016.11.006, 2017.

Point 31: Fig 6 Is it possible to normalize the data somehow and facilitate comparison across regions?

Response:

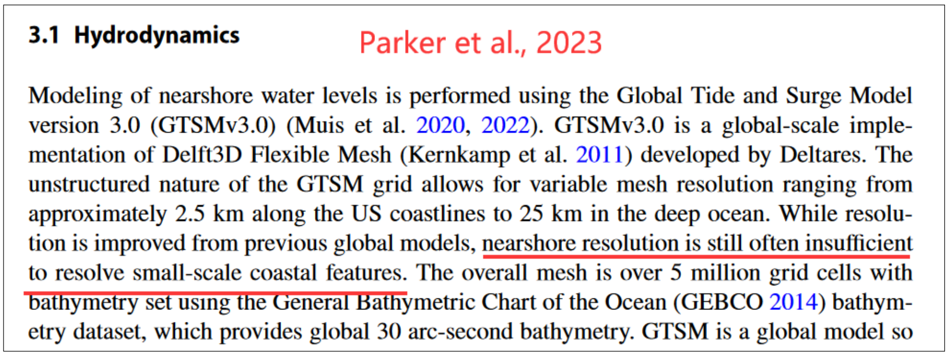
Thanks for reminding. However, we haven't found a more suitable way yet. The storm surge Amax range varies greatly in different regions (some places are over 4.5 meters, some are less than 0.7 meters). If we unify the y-axis, the readability of the area with a small Amax range would be affected. For example, the Western Asia (WAS):



Point 32: Line 231: Spatial resolution of your method does not seem so superior compared to GTSM, moreover I am concerned about the use of TG data for both training and evaluation (if this is not done carefully)

Response:

"Meanwhile, the grid with a resolution of several kilometers affects the effective simulation of small-scale physical factors", this sentence means that if the resolution of the grid used to **resolve** storm surges is, for example, 2.5km, the physical process smaller than this spatial scale would not be effectively simulated by the numerical model. This was mentioned by Parker et al. (2023) as well.



3.1 Hydrodynamics Parker et al., 2023

Modeling of nearshore water levels is performed using the Global Tide and Surge Model version 3.0 (GTSMv3.0) (Muis et al. 2020, 2022). GTSMv3.0 is a global-scale implementation of Delft3D Flexible Mesh (Kernkamp et al. 2011) developed by Deltares. The unstructured nature of the GTSM grid allows for variable mesh resolution ranging from approximately 2.5 km along the US coastlines to 25 km in the deep ocean. While resolution is improved from previous global models, nearshore resolution is still often insufficient to resolve small-scale coastal features. The overall mesh is over 5 million grid cells with bathymetry set using the General Bathymetric Chart of the Ocean (GEBCO 2014) bathymetry dataset, which provides global 30 arc-second bathymetry. GTSM is a global model so

For the data-driven model in this article, on the one hand, it does not require the grid; on the other hand, the training process is based on the observations from tide gauges directly. Tide gauges are currently the most precise coastal sea-level monitoring method, which can be considered to capture the impact of more small-scale physical factors than numerical models.

As for TG data preprocessing. First, the team produced the GESLA-3 dataset has processed it very carefully, this dataset was used in a wide range of ocean research (details see Haigh et al., 2023). Second, we also checked all TGs one by one through visual inspection before they were used for training and evaluation. This was a huge amount of work.

Parker, K., Erikson, L., Thomas, J., Nederhoff, K., Barnard, P., and Muis, S.: Relative contributions of water-level components to extreme water levels along the US Southeast Atlantic Coast from a regional-scale water-level hindcast, Nat Hazards, https://doi.org/10.1007/s11069-023-05939-6, 2023.

Haigh, I. D., Marcos, M., Talke, S. A., Woodworth, P. L., Hunter, J. R., Hague, B. S., Arns, A., Bradshaw, E., and Thompson, P.: GESLA Version 3: A major update to the global higher-frequency sea-level dataset, Geoscience Data Journal, 10, 293–314, https://doi.org/10.1002/gdj3.174, 2023.

Point 33: Line 260: Both variables should be available in ERA5: why not adding those?

Response:

There are currently no numerical/data-driven storm surge model products that consider the effects of rainfall and runoff, which may not be a simple linear addition relationship. How to add both variables into data-driven models and how they affect model precision need to be analyzed in future work.

Point 34: Line 268: 10 km or rather 0.1 degrees, hence variable with latitude?

Response:

"10 km along the coastline" means "there is a node every 10 kilometers along the coastline".