Ocean Sci., 20, 1513–1526, 2024
https://doi.org/10.5194/os-20-1513-2024

Ocean Science

Assessing storm surge model performance: what error indicators can measure the model's skill?

Rodrigo Campos-Caba[1,2], **Jacopo Alessandri**[1,2], **Paula Camus**[3], **Andrea Mazzino**[4,5], **Francesco Ferrari**[4,5], **Ivan Federico**[6], **Michalis Vousdoukas**[7,8], **Massimo Tondello**[9], and **Lorenzo Mentaschi**[1,2,6]

[1]Department of Physics and Astronomy (DIFA), University of Bologna, 40127 Bologna, Italy
[2]Interdepartmental Research Centre for Environmental Sciences (CIRSA), University of Bologna, 48123 Ravenna, Italy
[3]Departamento de Ciencias y Técnicas del Agua y del Medio Ambiente, University of Cantabria, 39005 Santander, Spain
[4]Department of Civil, Chemical and Environmental Engineering, University of Genoa, 16145 Genoa, Italy
[5]Istituto Nazionale di Fisica Nucleare, Sezione di Genova, Via Dodecaneso 33, 16146 Genoa, Italy
[6]CMCC Foundation – Euro-Mediterranean Center on Climate Change, 73100 Lecce, Italy
[7]Department of Marine Sciences, University of the Aegean, 81100 Mitilene, Greece
[8]MV Coastal and Climate Research Ltd, 3046 Limassol, Cyprus
[9]HS Marine SrL, 35027 Noventa Padovana, Italy

Correspondence: Rodrigo Campos-Caba (rodrigo.camposcaba@unibo.it)

Received: 13 May 2024 – Discussion started: 22 May 2024
Revised: 4 September 2024 – Accepted: 5 October 2024 – Published: 25 November 2024

Abstract. A well-validated storm surge numerical model is crucial, offering precise coastal hazard information and serving as a basis for extensive databases and advanced data-driven algorithms. However, selecting the best model setup based solely on common error indicators like the root-mean-square error (RMSE) or Pearson correlation does not always yield optimal results. To illustrate this, we conducted 34-year high-resolution simulations for storm surge under barotropic (BT) and baroclinic (BC) configurations using atmospheric data from ERA5 and a high-resolution downscaling of the Climate Forecast System Reanalysis (CFSR) developed by the University of Genoa (UniGe). We combined forcing and configurations to produce three datasets: (1) BT-ERA5, (2) BC-ERA5, and (3) BC-UniGe. The model performance was assessed against nearshore station data using various statistical metrics. While RMSE and Pearson correlation suggest BT-ERA5, i.e., the coarsest and simplest setup, is the best model (followed by BC-ERA5), we demonstrate that these indicators are not always reliable for performance assessment. The most sophisticated model (BC-UniGe) shows worse values of RMSE or Pearson correlation due to the so-called "double penalty" effect. Here we propose new skill indicators that assess the ability of the model to reproduce the distribution of the observations. This, combined with an analysis of values above the 99th percentile, identifies BC-UniGe as the best model, while ERA5 simulations tend to underestimate the extremes. Although the study focuses on the accurate representation of storm surge by the numerical model, the analysis and proposed metrics can be applied to any problem involving the comparison between time series of simulation and observation.

1 Introduction

In coastal areas, accurately depicting storm surge is paramount for effective risk assessment, preparedness, and mitigation strategies, as they can lead to coastal erosion, inundation, and infrastructure damage and threaten important cultural heritage sites (Reimann et al., 2018; Vousdoukas et al., 2022). Storm surges arise from the interaction between the atmosphere and the sea. Essentially, the atmosphere exerts forces on the waterbody, causing sea levels to rise due to low-atmospheric-pressure systems and strong wind fields (Pirazzoli and Tomasin, 2022). The atmospheric pressure effect, known as the inverse barometer effect or static amplification, typically contributes 10 % to 15 % of the total storm surge magnitude (World Meteorological Organization,

Published by Copernicus Publications on behalf of the European Geosciences Union.

2011). The second and more significant part of the storm surge, called dynamic amplification or wind setup, arises from tangential wind stress associated with the weather system's wind field acting on the ocean surface (Chaumillon et al., 2017).

Numerical simulations play a pivotal role in unraveling the complexities of physical phenomena such as storm surges (Park et al., 2022). They offer invaluable insights into various processes and greatly contribute to building extensive databases for further analysis and comprehension. Concerning storm surge, this refers to a complex oceanographic phenomenon that demands accurate oceanic and atmospheric data for precise representation. Due to diverse orographic configurations, atmospheric models often exhibit significant errors, necessitating the utilization of local-scale models with high resolution (Umgiesser et al., 2021). Additionally, the intricate coastal and bathymetric features and interactions pose challenges for existing hydrodynamical models to fully capture the relevant dynamics, partly due to their low resolution (Mentaschi et al., 2015; Toomey et al., 2022).

On the other hand, the utilization of unstructured-grid models enables a more accurate portrayal of coastal dynamics, considering the intricacies of bathymetry and shoreline configurations (Federico et al., 2017). This approach offers the advantage of employing a higher resolution at the coastlines while maintaining a more modest resolution in deeper waters (Ferrarin et al., 2019). Unstructured meshes offer flexibility in resolving basin geometry, allowing for local refinement of computational domains to simulate regional dynamics on a global mesh with coarse resolution. This flexibility is particularly valuable for coastal applications, where computational domains encompass complex coastlines and varying scales, ranging from basin size to details of river estuaries or riverbeds (Danilov, 2013). Over recent years, unstructured-grid models have increasingly emerged as alternatives to regular grids for large-scale simulations (e.g., Mentaschi et al., 2020; Muis et al., 2016; Vousdoukas et al., 2018; Fernández-Montblanc et al., 2020; Saillour et al., 2021; Wang et al., 2022; Zhang et al., 2023; Mentaschi et al., 2023), with established circulation unstructured models such as the Advanced Circulation Model for Shelves, Coastal Seas, and Estuaries (ADCIRC, Luettich et al., 1992; Pringle et al., 2021); the Finite-Volume Coastal Ocean Model (FVCOM, Chen et al., 2003); the Semi-implicit Cross-scale Hydroscience Integrated System Model (SCHISM, Zhang and Baptista, 2008; Zhang et al., 2016); the System of HydrodYnamic Finite Element Modules (SHYFEM, Umgiesser et al., 2004; Bellafiore and Umgiesser, 2010; Micaletto et al., 2022); TELEMAC (Hervouet and Bates, 2000); and Delft3D-FM (Deltares: Delft, 2024) being available.

In this study, we developed numerical simulations of storm surge in the northern Adriatic Sea with two main objectives: first we want to generate long-term databases of storm surge with a focus on accurately representing extreme values, and second we want to analyze the ability of different metrics to capture the skill of the model. The northern Adriatic Sea is a semi-enclosed body of water characterized by intricate bathymetry. The region's coastline exhibits distinct features, with the western coastline being relatively smooth and sandy, while the eastern coastline is fragmented and rocky, dotted with numerous islands. Both bathymetry and the configuration of the coastline significantly influence the physical processes occurring along the coast (Bellafiore and Umgiesser, 2010). The semi-enclosed nature of the Adriatic Sea predisposes it to experiencing intense storm surge events, leading to anomalous increases in sea level. These events are typically driven by local low-pressure system cyclogenesis and the associated strong winds, which are influenced by the region's orographic features (Umgiesser et al., 2021).

The application of numerical tools to study storm surge in the northern Adriatic Sea has garnered significant attention over the years, primarily due to its status as a high-risk area with unique cultural and environmental heritage and significant economic activities (Ferrarin et al., 2020). Previous efforts in this field have included predictive models projecting future storm scenarios (Yu et al., 1998), long-term numerical simulations (Lionello et al., 2010), analyses of storm events, use of various atmospheric forcings (De Vries et al., 1995; Zampato et al., 2006; Međugorac et al., 2018), investigations into seiches influence and data assimilation impacts (Bajo et al., 2019), and storm surge ensemble prediction systems for lagoons (Alessandri et al., 2023).

In this study, the numerical simulations are based on a long-term ocean circulation downscaling carried out with the SHYFEM model, which is an unstructured-grid finite-element hydrodynamic open-source code that solves the Navier–Stokes equations with hydrostatic and Boussinesq approximations (Umgiesser et al., 2004; Micaletto et al., 2022). The model has been already implemented in operational (Federico et al., 2017) and relocatable (Trotta et al., 2016) forecasting frameworks and for storm surge events (Park et al., 2022; Alessandri et al., 2023). The choice of SHYFEM is driven by its flexibility in handling complex bathymetry and irregular coastlines through its unstructured-grid framework, allowing for higher resolution in critical areas. Additionally, its successful implementation in operational and relocatable forecasting frameworks and storm surge events confirms its reliability for this study. The simulations consider different setups to explore the influence of different atmospheric forcings and model configurations on the model's skill. Regarding model configurations, both barotropic and baroclinic simulations were conducted to compare potential differences between these two widely used approaches, as covered in the literature for the proper representation of storm surge (e.g., Weisberg and Zheng, 2008; Staneva et al., 2016; Hetzel et al., 2017; Ye et al., 2020; Muñoz et al., 2022). Furthermore, we focus on the use of different metrics and their ability to provide reliable indications of the model's performance, which is an essential aspect in assessing model skill and to select the best model configuration. In addition to classical metrics, such as the Pearson correlation coefficient and root-mean-square error (RMSE), two customized versions of the mean absolute deviation (MAD) are introduced. These tailored metrics incorporate observed and simulated percentiles, ranging from 0 % to 100 %, to ensure accurate representation of extreme values during the performance evaluation.

The paper is organized as follows. Materials and methods are described in Sect. 2, including the description of the two atmospheric databases considered for the simulations, the model setup, and the procedures to carry out the performance evaluation. Section 3 shows the main results of the comparisons between observed and simulated storm surge. The paper continues with a discussion of the results in Sect. 4. Finally, the conclusion in Sect. 5 summarizes the key points of the study.

2 Materials and methods

2.1 Atmospheric forcing

In this study, we utilized two distinct atmospheric databases to force the circulation model, incorporating mean sea level pressure and wind fields. The first database is ERA5, the fifth generation of reanalysis data generated by the European Centre for Medium-Range Weather Forecasts (ECMWF). ERA5 builds upon the Integrated Forecasting System (IFS) Cy41r2, which became operational in 2016, providing hourly output with a horizontal resolution of $0.25° \times 0.25°$ for atmospheric variables (Hersbach et al., 2020). ERA5 is relatively high resolution and accurate for a global reanalysis, although it is known to be affected by negative biases at high percentiles, particularly when compared with measured wind speed (Pineau-Guillou et al., 2018; Vannucchi et al., 2021; Benetazzo et al., 2022; Gumuscu et al., 2023).

Since ERA5 is relatively coarse for local studies and exhibits significant underestimation of extremes, we employed an alternative approach using a high-resolution (3.3 km) atmospheric downscaling developed by the University of Genoa (UniGe). Wind forcing was derived from 10 m wind fields via the Weather Research and Forecast (WRF-ARW) model v3.8.1, allowing for improved representation of small-scale forcings and physics. The computational domains comprised a 10 km resolution grid covering the Mediterranean, northern Africa, and southern Europe (A10) and a 3.3 km grid over the Tyrrhenian Sea basin and northern Adriatic Sea basin (A3) nested within A10. Initial conditions were obtained from the Climate Forecast System Reanalysis (CFSR) data, which are known their for reliability but occasionally underestimate extreme events (Saha et al., 2010). WRF simulations were conducted for 24 h with hourly outputs, employing established physical parameterization schemes to ensure accuracy across various atmospheric conditions. For further details, readers are referred to Mentaschi et al. (2015).

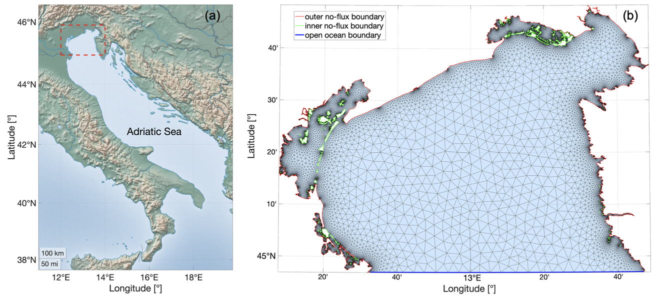

Figure 1. **(a)** Location of study area, marked with a dashed red line. **(b)** Unstructured grid for the study area, in which the blue line represents the location of the open boundary condition, the red line the coastline, and the green lines the coastline formed by islands.

2.2 Model setup

The SHYFEM model utilizes staggered finite elements in an unstructured Arakawa B horizontal grid, with the vertices of the triangle elements referred to as nodes. Vectors (velocity) are calculated at the center of each element, while scalars (temperature, salinity, and water levels) are determined at nodes (Federico et al., 2017). The unstructured grid for the simulations in this study was generated using the OceanMesh2D tool (Roberts et al., 2019) with a horizontal resolution of 3 km on the open-ocean boundary and 50 m in the coastline (Fig. 1.a). The General Bathymetric Chart of the Oceans (GEBCO) dataset (Weatherall et al., 2015) was used, incorporating a high-resolution coastline from the European Environmental Agency. However, due to identified overestimations in water depth in the Venice and Marano lagoons from GEBCO bathymetry, adjustments were made based on the contributions from Fagherazzi et al. (2007), Lovato et al. (2010), and Zaggia et al. (2017) for the Venice lagoon and Petti et al. (2019) and Bosa et al. (2021) the for the Marano lagoon.

Sea level residuals, current velocity, temperature, and salinity from the Copernicus Mediterranean Sea Physics reanalysis (Escudier et al., 2021) were considered as initial and open-ocean boundary conditions. Tides with hourly resolution from the Finite Element Solution (FES) 2014 (Lyard et al., 2021) were also included to account for the total sea level in the simulations. Specifically, the constituents included for the tide reconstruction are SA, SSA, O1, P1, S1, K1, N2, M2, MKS2, S2, R2, K2, M3, M4, and MS4, which were selected based on preliminary harmonic analysis applied to sea level observation data in the locations specified in Sect. 2.2.

Two model configurations were considered: (a) barotropic (BT) and (b) baroclinic (BC), employing 33 vertical levels with a layer thickness of 1 m up to 10 m depth and then 2 m up to a maximum depth of 60 m (BC). To determine vertical viscosities and diffusivities, we utilize a k-ε turbulence scheme derived from the General Ocean Turbulence Model (GOTM) (Burchard and Petersen, 1999). For wind stress at the air–sea interface, a constant wind drag coefficient of 2.5×10^{-3} was employed, following the works from Orlić et al. (1994) and Zampato et al. (2007). The bottom stress is determined through the quadratic formulation:

$$\tau_{xz}^{z_N} = \frac{C_B}{H_N^2} |U_N| U_N \quad \tau_{yz}^{z_N} = \frac{C_B}{H_N^2} |U_N| V_N, \tag{1}$$

where $\tau_{xz}^{z_N}$ and $\tau_{yz}^{z_N}$ are the turbulent shear stresses at the bottom interface of the deepest layer, H_N is bottom-layer thickness, and U_N and V_N are the zonal and meridional transports of the bottom layer. C_B is the bottom drag coefficient, which is defined as follows:

$$C_B = \left(\frac{0.4}{\ln\left(\frac{\lambda_B + 0.5 H_N}{\lambda_B}\right)} \right)^2, \tag{2}$$

where λ_B is the bottom roughness length expressed in meters, which in this study remains constant at 0.01 m. For further details, readers are referred to Maicu et al. (2021).

The simulation period extends from 1987 to 2020 with hourly output. Three combinations of atmospheric forcing and configuration are considered here: (1) barotropic forced by ERA5 (BT-ERA5), (2) baroclinic forced by ERA5 (BC-ERA5), and (3) baroclinic forced by UniGe (BC-UniGe).

2.3 Model performance evaluation

The model output was compared with observations from tide gauges located in the northern Adriatic Sea. The observational data were acquired from the Italian National Institute for Environmental Protection and Research (ISPRA), the Civil Protection of the Friuli-Venezia Giulia Region, and Raicich (2023). Table 1 summarizes the locations considered and the available time spans for comparison that match with the simulation time span. Fig. 2 shows the locations considered for comparison between measured and simulated storm surge, together with the bathymetry used for the simulations.

Both the model output and the observations were processed as follows to enable their intercomparability. To start, both measurement and simulation were centered with a zero mean and then detrended. This approach mitigates possible effects of unmodulated land motion (Chepurin et al., 2014) and ensures that extreme values across the years can be considered as homogeneous and can be compared despite relative sea level changes (Ferrarin et al., 2022). Harmonic analysis was performed for each calendar year on the detrended sea levels using the T-Tide MATLAB package (Pawlowicz et al., 2022), and the non-tidal residual was obtained as the arithmetic difference between sea level and tides (Tiggeloven et al., 2021). Performing yearly harmonic analysis reduces timing errors that could cause tidal energy to seep into the non-tidal residual (Merrifield et al., 2013).

Finally, to obtain the pure storm surge (hereafter also called "surges"), a low-pass filter is applied to the non-tidal residual, following the work of Park et al. (2022). In this study, we consider a cut-off period of 13 h for the filter based on the mixed semidiurnal tidal regime around the northern Adriatic Sea (Lionello et al., 2021).

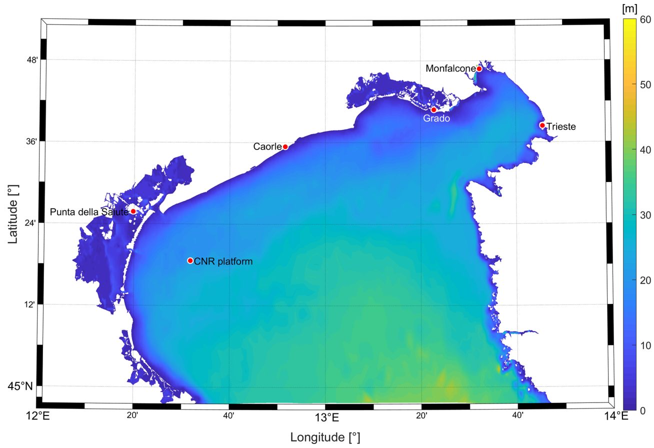

Figure 2. Tide gauge locations and bathymetry (depth values on positive).

The performance evaluation of the simulations relies on the computation of statistical metrics of hourly data, which encompass the entire dataset, as well as values exceeding the 99th percentile from the cumulative distribution of measured data at each location. The following metrics are considered.

We first consider the Pearson correlation:

$$\rho = \frac{1}{N-1} \sum_{i=1}^{N} \left(\frac{S_i - \mu_S}{\sigma_S} \right) \left(\frac{O_i - \mu_O}{\sigma_O} \right), \tag{3}$$

where S_i and O_i are the ith simulated and observed data, respectively; N is the sample size; and μ and σ are the mean and standard deviations of S and O, respectively. A value closer to one identifies a better performance.

Second, we consider the root-mean-square error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (S_i - O_i)^2}, \tag{4}$$

where a value closer to zero indicates a better performance.

Third, we consider bias, defined as follows:

$$\mathrm{Bias} = \overline{S} - \overline{O}, \tag{5}$$

where $\overline{S}$ and $\overline{O}$ are the average simulation and observation values, respectively. A value closer to zero identifies a better performance, negative values indicate underestimation, and positive values indicate overestimation from the simulations. Given that both observed and simulated data were detrended and had their mean removed, bias was solely applied to the analysis of values exceeding the 99th percentile.

Fourth, we consider the slope of the linear fit between observations and the simulation:

$$S = mO + b, \tag{6}$$

Table 1. Locations considered for validation, including available start and end dates matching the simulation time span.

Location	Long [°]	Lat [°]	Start date	End date
ISMAR-CNR research platform "Aqua Alta" (hereafter CNR platform)	12.53	45.31	1 Jan 1987	31 Dec 2020
Punta della Salute	12.33	45.43	1 Jan 1987	31 Dec 2020
Caorle	12.86	45.59	1 Jan 2000	31 Dec 2020
Grado	13.38	45.68	1 Jan 1991	31 Dec 2020
Monfalcone	13.54	45.78	1 Jan 2008	31 Dec 2020
Trieste	13.76	45.64	1 Jan 1987	31 Dec 2020

where the slope is given by the coefficient m. A value closer to one indicates a better performance.

Fifth, we consider mean absolute deviation (MAD):

$$\mathrm{MAD} = \overline{|S - O|}, \tag{7}$$

where a value closer to one indicates a better performance.

Additionally, with the aim of considering the representation of extremes by the simulations, we introduce two new metrics based on customized versions of the mean absolute deviation.

The first new metric is the MAD of the percentiles (MADp):

$$\mathrm{MADp} = \overline{|S_{\mathrm{prc}} - O_{\mathrm{prc}}|}, \tag{8}$$

where S_{prc} and O_{prc} are the simulation and observation percentile values, respectively, considered from 0 % to 100 % every 1 %. The MADp metric provides a comprehensive assessment of simulation model performance by comparing percentile values derived from simulations (S_{prc}) with those observed (O_{prc}). This evaluation encompasses the entire distribution, from the lowest to the highest percentiles, allowing us to gauge the model's accuracy across a range of scenarios. MADp is particularly valuable for its sensitivity to systematic errors, such as persistent underestimation of high percentiles, which can significantly impact the reliability of simulation results. By penalizing these systematic errors, MADp highlights areas where improvements in the simulation model are necessary to better align with observed data. Lower MADp values indicate closer agreement between simulations and observations.

The second new metric is the corrected MAD (MADc):

$$\mathrm{MADc} = \overline{|S - O|} + \mathrm{MADp}. \tag{9}$$

In this indicator we exploit the ability of the "traditional" MAD to capture the model's skill but reduce its strong penalization of the phase error or timing error (i.e., the reproduction by the model of peaks shifted in space-time) by adding the MAD (MADp) on the percentiles as previously defined. MAD measures the average absolute difference between simulated and observed values, while MADp evaluates the average percentage deviation between them. By combining these two components, MADc provides a comprehensive evaluation of the simulation model's performance, considering both the magnitude and percentage deviations. A lower MADc value indicates better agreement between simulated and observed values, reflecting higher accuracy and reliability of the simulation model.

To quantify phase errors between observations and simulations, peaks in the hourly time series were identified using MATLAB's "find peaks" function for both observed and simulated data. The phase error was then calculated by measuring the time difference (in hours) between the occurrence of each peak in the observations and the corresponding peak in the simulations. This approach provided a direct assessment of the model's accuracy in capturing the timing of key events, such as storm surges.

The proposed metrics were also validated using an idealized time series. A sinusoidal time series was generated to represent an observed parameter. Two simulated time series were then created: one with the same amplitude as the observation but shifted in time (introducing a phase error) and the other with the same phase as the observation but with half the amplitude. Various metrics were calculated and plotted as scatter plots (Fig. S1 in the Supplement). The results indicated better performance for the simulation that underestimated the observations when assessed with Pearson correlation, RMSE, and MAD. In contrast, the time series that accurately captured the amplitude was penalized for the phase error, which negatively affected its performance on these metrics. However, the proposed MADp and MADc metrics identified it as the better model.

3 Results

The probability distribution estimates (PDEs) and empirical cumulative distribution functions (ECDFs), available in Figs. S2 to S7, show that BC-UniGe better represents the higher values of storm surge when compared with observations, particularly when considering values above the 99th percentile. However, some overestimations are noticeable for Caorle and Monfalcone with BC-UniGe. In contrast, simulations with ERA5 forcing tend to underestimate these higher values, which is more noticeable for BT-ERA5.

The performance evaluation shows that if the model performance is assessed in terms of Pearson correlation, RMSE, and MAD, the surges simulated with the ERA5 forcing fit better to the measured data (Fig. 3). The Pearson correlation coefficients obtained a range between 0.8 and 0.9 in all locations for the three simulations, with a maximum of 0.842 with BT-ERA5 in Grado (Fig. 3d). Regarding the RMSE, mean values of 0.077 m for BT-ERA5, 0.075 m for BC-ERA5, and 0.079 m for BC-UniGe were obtained, with a minimum of 0.072 m (BT-ERA5 in Grado, Fig. 3d) and a maximum of 0.094 m (BC-UniGe in Monfalcone, Fig. 3e). Similar results are obtained for MAD, which shows better performance for the simulations with ERA5 forcing at all locations. Only in Trieste does BC-UniGe achieve the same performance as BC-ERA5 for this metric. Despite the aforementioned factors, the best performance is achieved by BC-UniGe in the linear fit slope, with values above 0.8 in all locations and a maximum of 0.869 in Monfalcone (Fig. 3e). For this parameter, less favorable performance is obtained with BT-ERA5 at all locations.

For MADp, the best performance is achieved by BC-UniGe at all locations, with a mean value of 0.004 m, while less favorable results are obtained with BT-ERA5, with a mean of 0.011 m. Similar results were obtained for MADc, except in Caorle (Fig. 3c) and Monfalcone (Fig. 3e), where BC-ERA5 showed better performance, likely due to overestimation in the mentioned sites. These results underscore the importance of considering percentiles as part of the performance evaluation. BC-UniGe simulations demonstrate an improvement in representing extreme values, showing a better fit of the highest percentiles, which can be noticed in Figs. 4 and 5. Additionally, these figures indicate that BC-UniGe simulations produce a greater dispersion of data, likely due to a more frequent occurrence of phase error, which was quantified as 3.1 % higher than in BT-ERA5 and 4.5 % higher than in BC-ERA5. However, they also exhibit a better fit of the linear regression and a more accurate representation of extreme values compared to BC-ERA5, which fail to represent the most extreme events in each location.

The results of the error metrics for surge values above the 99th percentile, represented using radar charts (Fig. 6), confirm that, in general, better performance is observed with BC-UniGe, while less favorable results are obtained for BT-ERA5. Although the transition from barotropic to baroclinic configuration indicates an improvement in the representation of extremes (Weisberg and Zheng, 2008; Staneva et al., 2016; Hetzel et al., 2017; Ye et al., 2020; Muñoz et al., 2022), the utilization of UniGe forcing represents the best improvement across practically all metrics. Only in Caorle (Fig. 6c) and Monfalcone (Fig. 6e) does BC-ERA5 show better Pearson correlation, RMSE, and MAD; additionally, at the latter location MADc exhibits better performance for that simulation, likely due to overestimation of the peaks by BC-UniGe in Monfalcone. At the other locations, it is evident

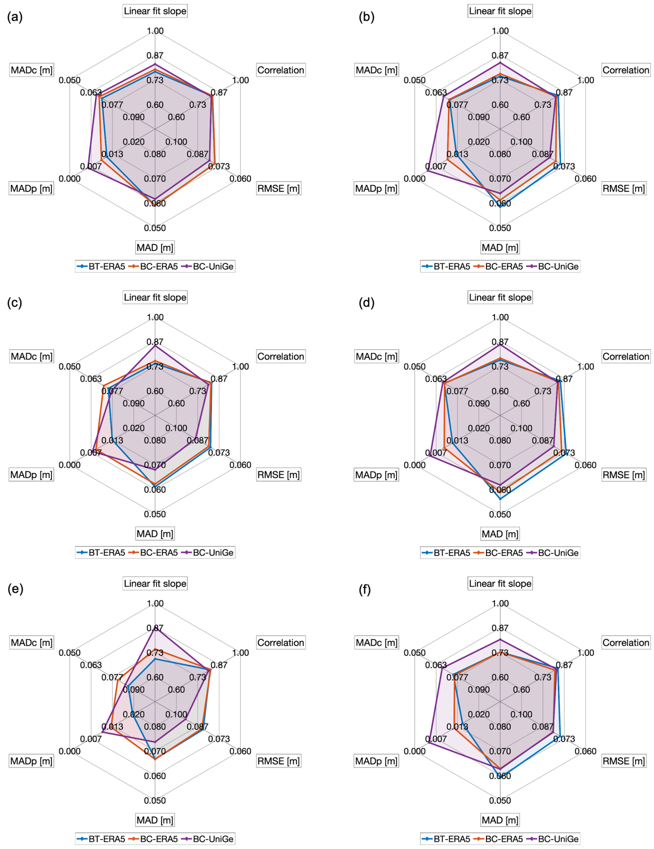

Figure 3. Radar charts of evaluation metrics for the total amount of data in all locations: **(a)** CNR platform, **(b)** Punta della Salute, **(c)** Caorle, **(d)** Grado, **(e)** Monfalcone, **(f)** Trieste. For RMSE, MADp, and MADc a reverse axis is used, this ensures that simulations covering a larger area in each metric represent a better performance (i.e., values on the fringe refer to better performance).

that BC-UniGe performs better in representing the highest storm surge values.

In order to show the capacity of the different model configurations to represent certain known storm events at each location, Fig. 7 shows time series of different storm surge events at each location. These extreme events were chosen according to the contributions of Lionello et al. (2012), Međugorac et al. (2018), Ferrarin et al. (2020), Umgiesser et al. (2021), and Giesen et al. (2021). As mentioned before, the incorporation of the UniGe forcing implies a significant improvement in the representation of extreme events, clearly evident in the peak values of the storm surge. Despite this an overestimation of some surge peaks is also observed in the events chosen at Punta della Salute (Fig. 7b), Caorle (Fig. 7c), and Monfalcone (Fig. 7e) with BC-UniGe. On the other hand, a systematic underestimation of extremes obtained in simulations with ERA5 forcing is notable in every surge peak.

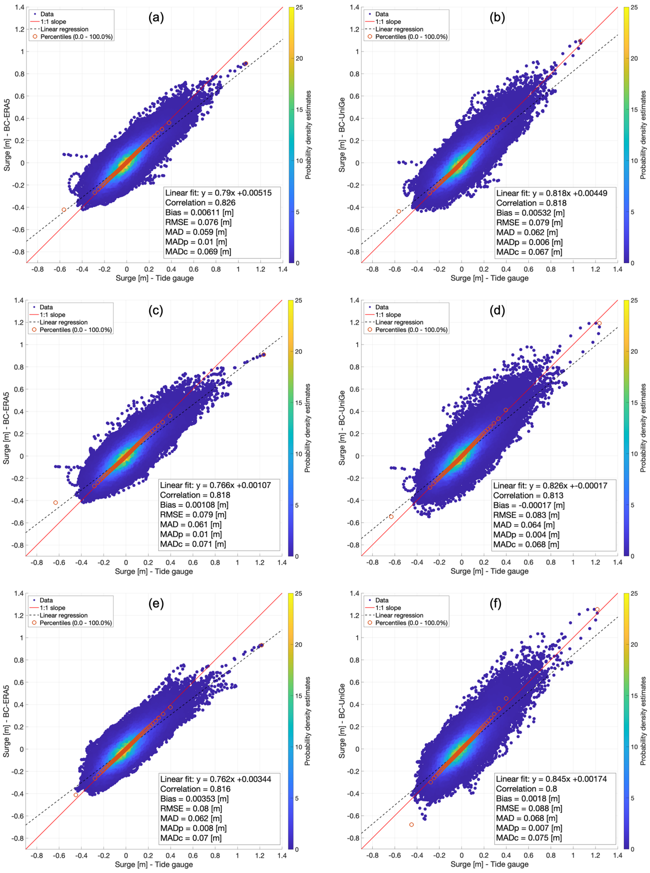

Figure 4. Scatter plots between tide gauges and baroclinic simulations for the CNR platform with BC-ERA5 **(a)** and BC-UniGe **(b)**, Punta della Salute with BC-ERA5 **(c)** and BC-UniGe **(d)**, and Caorle with BC-ERA5 **(e)** and BC-UniGe **(f)**.

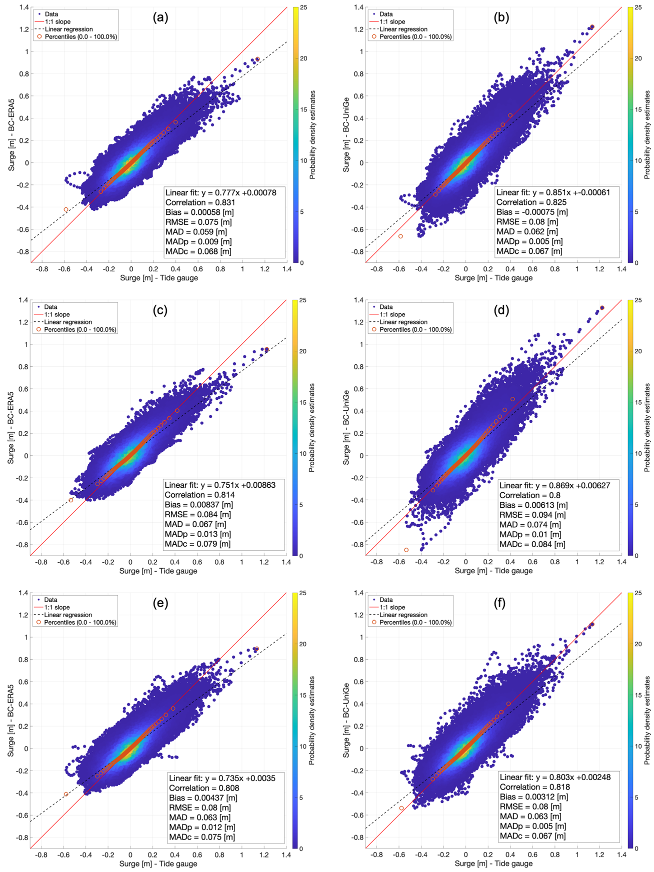

Figure 5. Scatter plots between tide gauges and baroclinic simulations for Grado with BC-ERA5 **(a)** and BC-UniGe **(b)**, Monfalcone with BC-ERA5 **(c)** and BC-UniGe **(d)**, and Trieste with BC-ERA5 **(e)** and BC-UniGe **(f)**.

4 Discussion

The utilization of different atmospheric forcing databases has revealed significant implications for the representation of storm surge in numerical simulations. Given the direct influence of wind speed and sea level pressure on this phenomenon, as represented in both forcings databases, the resulting model performances present significant differences. While simulations using ERA5 forcing generally show slightly better performance for traditional metrics such as RMSE, MAD, and the Pearson correlation coefficient, a more detailed analysis reveals that using the UniGe forcing results in better performance, especially in terms of the extreme values, when considering additional metrics.

Simulations using ERA5 forcing tend to underestimate the highest surge values, primarily due to a corresponding underestimation of extreme wind speed by this database, a variable crucially linked to surge amplitude (Campos et al., 2022). Despite this, metrics such as the Pearson correlation, RMSE, and MAD generally indicate better performance for ERA5 simulations. Conversely, the utilization of UniGe forcing shows an improvement in representing the peaks of storm surge events (with the noticeable exception of Monfalcone, where the extremes are overestimated, and where MADp present similar values for BC-ERA5 and BC-UniGe). These results demonstrate that the increase in atmospheric forcing resolution does not consistently translate into better values of all the statistical metrics.

It is important to recognize that identifying the optimal model configuration cannot rely solely on a few statistical metrics. As outlined in Sect. 3, no single simulation emerges as superior across all metrics and locations. While ERA5 simulations may demonstrate better performance on RMSE, Pearson correlation, and MAD, BC-UniGe exhibits superior performance in terms of the slope of the linear fit, MADp, and MADc.

From an epistemic point of view, BC-UniGe is a significantly more sophisticated model compared to BT-ERA5. Not only does it employ a higher-resolution forcing, it also takes into account the baroclinicity and the vertical motion within the water column, whereas the barotropic configuration of BT-ERA5 approximates the ocean as a 2D sheet that

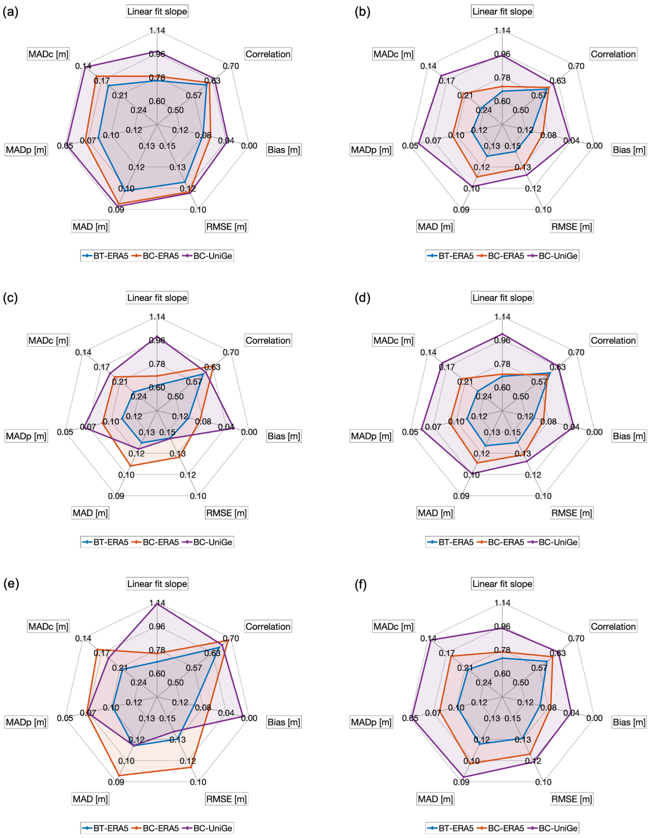

Figure 6. Radar charts of evaluation metrics for surge values above the 99th percentile of the cumulative distribution at each location: **(a)** CNR platform, **(b)** Punta della Salute, **(c)** Caorle, **(d)** Grado, **(e)** Monfalcone, and **(f)** Trieste. Bias is represented by an absolute value. In addition, for RMSE, bias, MADp, and MADc a reverse axis is used. This ensures that simulations covering a larger area for each metric represent a better performance (i.e., values on the fringe refer to better performance).

is only subject to vertically uniform motions and waves. This suggests that widespread indicators such as RMSE, Pearson correlation, and MAD, which in this case identify BT-ERA5 as the best model, should not be considered as the sole source of information in model skill assessment, since a higher-resolution forcing and a baroclinic setup are known in literature to better capture the variability of the sea levels (Weisberg and Zheng, 2008; Hetzel et al., 2017; Muñoz et al., 2022).

Similar results were found by Zampato et al. (2006) using SHYFEM with three different forcings for wind and atmospheric pressure fields: the ECMWF global model, the high-resolution LAMI (Limited Area Model Italy), and satellite QuickSCAT. In this work, the authors found well-correlated sea levels with observations near Venice using the ECMWF forcings but underestimation of the highest values. On the other hand, simulations driven by the high-resolution model (LAMI) succeeded in simulating the storm surge, giving a good reproduction of the sea level peaks. Nevertheless, the correlation with observed data was lower than in the case of ECMWF forcing.

The complexity of simulation performance evaluations is echoed in the work of Mentaschi et al. (2013), who caution against over-reliance on metrics like RMSE, NRMSE (normalized RMSE), and SI (scatter index) as indicators of model performance. These metrics may not fully capture the intricacies of natural processes such as atmospheric dynamics, ocean circulation, or wave generation and propagation. These authors mention that the RMSE and its variations tend to assume typical values of the best performance for simulations that underestimate the physical process of interest. The discrepancy between metrics and the representation of extremes highlights the need for a comprehensive understanding of model performance beyond traditional statistical measures.

These performance evaluation results are usually related to phase error in high-resolution models and RMSE "double penalty". The phase error refers to a discrepancy between the timing or phase of a simulated event and its actual occurrence on measured data. In the context of atmospheric models, phase errors can manifest as delays or advances in the timing of weather events, such as the onset of precipitation, the movement of storm systems, or the arrival of fronts. Double penalty refers to a situation where the errors in the model output are penalized twice in indicators such as RMSE and MAD, once for missing the observations and again for giving a false alarm (e.g., Gilleland et al., 2009). This is a well-known problem during performance evaluation of numerical models, and different contributions have sought to overcome it with approaches specialized in atmospheric and oceanographic fields (e.g., Ebert and Mcbride, 2000; Zingerle and Nurmi, 2008; Roberts and Lean, 2008; Mittermaier, 2014; Skok and Roberts, 2016; Crocker et al., 2020).

In RMSE, double penalty is further amplified compared to MAD, as the penalizations due to the peak mismatch are squared. This means that phase errors have a disproportionately large impact on RMSE. A more sophisticated model may be better able to capture the magnitude of the peaks, but as it is more prone to phase error compared to low-resolution ones this ability will be doubly penalized. This is the reason why a less sophisticated model employing a low-resolution forcing (BT-ERA5) appears to outperform the other two in terms of RMSE. Conversely, MAD, although it also experiences a form of double penalty, reduces the impact of this effect compared to RMSE. As a result, the performance differences between simulations, particularly above the 99th percentile, are generally more pronounced for MAD than for RMSE, better highlighting the superiority of BC-UniGe. This enhanced differentiation is likely due to MAD's linear weighting of errors, which reduces the inflated impact of large deviations that characterize RMSE.

In other words, RMSE tends to be better for "blurring" models, whereas high-resolution models, known to be more

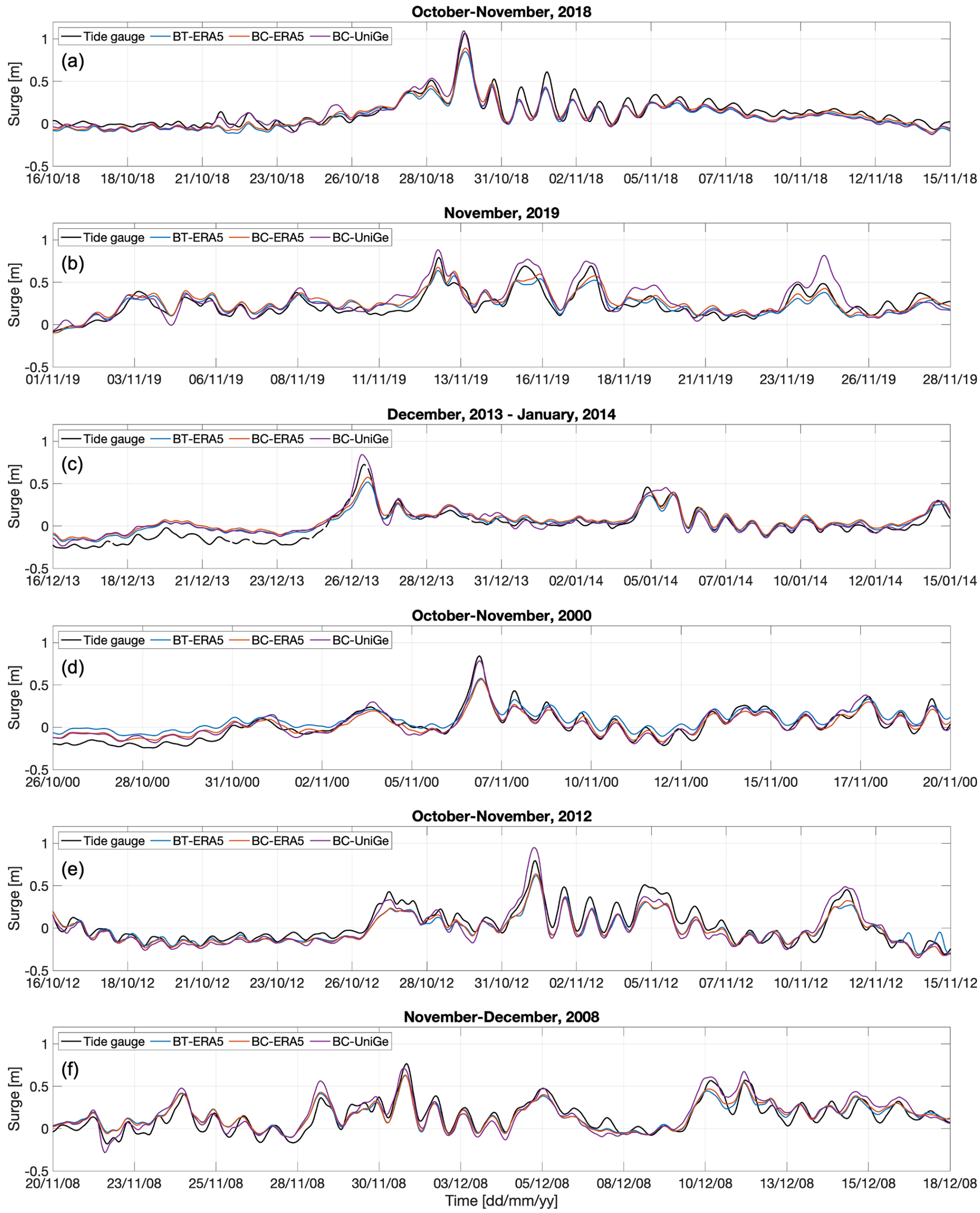

Figure 7. Time series of different storm surge events in all of the locations, showing the tidal gauge data versus the model data: **(a)** CNR platform, **(b)** Punta della Salute, **(c)** Caorle, **(d)** Grado, **(e)** Monfalcone, and **(f)** Trieste.

capable of reproducing small-scale dynamics (e.g., BC-UniGe), perform worse in terms of RMSE due to phase error (Crocker et al., 2020). Although in many aspects capturing a peak with a phase error is preferable to missing the peak entirely, this does not lead to a reduction in the RMSE.

This limitation of RMSE also impacts the Pearson correlation. Indeed, RMSE can be decomposed into a bias component and a scatter component that depends solely on the Pearson correlation (Mentaschi et al., 2013, Eq. 8). All of these considerations call for caution when claiming that one model outperforms another simply based on a better value of RMSE, MAD, or Pearson correlation.

The MADc indicator was introduced here as a possible way to correct MAD to make it less prone to the double penalty effect. The incorporation in MADc of a term that takes into account the distribution of the data (the MAD of the percentiles MADp) rewards the ability of a high-resolution and more sophisticated model to reproduce the variability in the observations without systematic errors. In other words, MADc remains more resilient to phase errors compared to other metrics, ensuring that discrepancies in the timing of events do not unduly influence the assessment of model performance. The differences between the simulation metrics are generally in the range of millimeters when considering the overall data, but these differences are significant in relative terms. For the MADc metric, BC-UniGe shows improvements ranging from 1.3 % (Grado) to 9.3 % (Trieste) compared to BT-ERA5 and from 1.6 % (Grado) to 10.3 % (Trieste) compared to BC-ERA5. The improvements are even more notable when focusing on values above the 99th percentile, where BC-UniGe outperforms BT-ERA5 by 12 % (Monfalcone) to 31.6 % (Trieste) and BC-ERA5 by 4.1 % (Caorle) to 20.2 % (Trieste).

As shown in Sect. 3, some discrepancies were observed in Caorle and Monfalcone, where BC-ERA5 achieved better performance in terms of MADc. A possible explanation for this could be related to the location of the tide gauges at these sites. The tide gauge at Caorle is situated in a protected area inside the Livenza River, a location not fully represented by the simulations due to the resolution of the coastline, even though high-resolution model data were used. A similar issue is found in Monfalcone, where the tide gauge is located in front of a breakwater not fully represented by the coastline used in the model. These factors could affect the signals obtained from observations and simulations, primarily due to local effects at the tide gauge locations.

5 Conclusions

In this study we developed high-resolution simulations of storm surge in the northern Adriatic Sea spanning from 1987 to 2020 using the model SHYFEM and employing different forcing data and physical configurations. The comparative analysis of the results highlights nuanced differences in performance metrics, particularly concerning the representation of the extreme values. Traditional metrics like Pearson correlation, RMSE, and MAD favor a simulation (BT-ERA5) forced by a coarser database and employing a less sophisticated setup (barotropic). However, a closer examination and the use of different metrics tell a different story and allow us to identify a baroclinic model forced by a high-resolution dataset (BC-UniGe) as better able to capture the variability of the water levels and, in particular, the extremes. This is because BC-UniGe is more prone to phase error than BT-ERA5 and is thus doubly penalized in indicators such as RMSE, MAD, and Pearson correlation.

The corrected MAD (MADc) introduced in this study comes as a possible way to alleviate the double penalty by adding a term that rewards the ability of a model to capture the distribution of the observations irrespective of the position of the peaks. In this study MADc is successful in identifying BC-UniGe as the best simulation in most locations. Even though this study has focused on the performance evaluation of storm surge, the analysis and proposed customized metrics (MADc and MADp) can be applied to any problem of validating a numerical model with observations by time series comparison.

These findings suggest that simply having a lower RMSE is insufficient evidence to claim that one model is superior to another. RMSE, MAD, and Pearson correlation are valuable indicators but should be used considering their limitations and complemented by other metrics, qualitative assessment, and expert judgment.

Data availability. The raw data supporting the conclusions of this article will be made available by the authors without undue reservation.

Supplement. The supplement related to this article is available online at: https://doi.org/10.5194/os-20-1513-2024-supplement.

Author contributions. RCC carried out the numerical simulations, post processing, and performance evaluation of the simulations and prepared the manuscript. LM guided the numerical simulations, post processing, and performance evaluation and contributed to the preparation of the manuscript. JA guided and supported numerical simulations. PC contributed to the performance evaluation and the preparation of the manuscript. AM, FF, IF, and MV contributed during the preparation of the manuscript. MT contributed to the performance evaluation.

Competing interests. Co-author Massimo Tondello is employed by the company HS Marine SrL. Co-author Michalis Vousdoukas is employed by the company MV Coastal and Climate Research Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Special issue statement. This article is part of the special issue "Extremes in the marine environment: analysis of multi-temporal and multi-scale dynamics using observations, models, and machine learning techniques". It is a result of the EGU General Assembly 2023, session OS4.6, Vienna, Austria, 26 April 2023.

Acknowledgements. Rodrigo Campos-Caba, Ivan Federico, and Lorenzo Mentaschi acknowledge support from the European Space Agency (ESA) under the Earth Observation Advanced science Tools for Sea level Extreme Events (EOatSEE) project.

Financial support. This research has been supported by the European Space Agency (ESA) under the Earth Observation Advanced science Tools for Sea level Extreme Events (EOatSEE) project.

Review statement. This paper was edited by Antonio Ricchi and reviewed by three anonymous referees.

References

Alessandri, J., Pinardi, N., Federico, I., and Valentini, A.: Storm Surge Ensemble Prediction System for Lagoons and Transitional Environments, Am. Meteorol. Soc., 38, 1791–1806, https://doi.org/10.1175/WAF-D-23-0040.1, 2023.

Bajo, M., Medugorac, I., Umgiesser, G., and Orlić, M.: Storm surge and seiche modelling in the Adriatic Sea and the impact of data assimilation. Q. J. Roy. Meteorol. Soc., 145, 2070–2084, https://doi.org/10.1002/qj.3544, 2019.

Bellafiore, D. and Umgiesser, G.: Hydrodynamic coastal processes in the north Adriatic investigated with a 3D finite element model, Ocean Dynam., 60, 225–273, https://doi.org/10.1007/s10236-009-0254-x, 2010.

Benetazzo, A., Davison, S., Barbariol, F., Mercogliano, P., Favaretto, C., and Sclavo, M.: Correction of ERA5 Wind for Regional Climate Projections of Sea Waves. Water (Switzerland), 14, 1590, https://doi.org/10.3390/w14101590, 2022.

Bosa, S., Petti, M., and Pascolo, S.: Improvement in the sediment management of a lagoon harbor: The case of Marano Lagunare, Italy, Water, 13, 3074, https://doi.org/10.3390/w13213074, 2021.

Burchard, H. and Petersen, O.: Models of turbulence in the marine environment-a comparative study of two-equation turbulence models, J. Mar. Syst., 21, 29–53, 1999.

Campos, R. M., Gramcianinov, C. B., de Camargo, R., and da Silva Dias, P. L.: Assessment and Calibration of ERA5 Severe Winds in the Atlantic Ocean Using Satellite Data, Remote Sens., 14, 4918, https://doi.org/10.3390/rs14194918, 2022.

Chaumillon, E., Bertin, X., Fortunato, A. B., Bajo, M., Schneider, J. L., Dezileau, L., Walsh, J. P., Michelot, A., Chauveau, E., Créach, A., Hénaff, A., Sauzeau, T., Waeles, B., Gervais, B., Jan, G., Baumann, J., Breilh, J. F., and Pedreros, R.: Storm-induced marine flooding: Lessons from a multidisciplinary approach, Earth-Sci. Rev., 165, 151–184, https://doi.org/10.1016/j.earscirev.2016.12.005, 2017.

Chen, C., Liu, H., and Beardsley, R. C.: An Unstructured Grid, Finite-Volume, Three-Dimensional, Primitive Equations Ocean Model: Application to Coastal Ocean and Estuaries, Ocean. Technol., 20, 159–186, 2003.

Chepurin, G. A., Carton, J. A., and Leuliette, E.: Sea level in ocean reanalyses and tide gauges, J. Geophys. Res.-Oceans, 119, 147–155, https://doi.org/10.102/2013JC009365, 2014.

Crocker, R., Maksymczuk, J., Mittermaier, M., Tonani, M., and Pequignet, C.: An approach to the verification of high-resolution ocean models using spatial methods, Ocean Sci., 16, 831–845, https://doi.org/10.5194/os-16-831-2020, 2020.

Danilov, S.: Ocean modeling on unstructured meshes, Ocean Modell., 69, 195–210, https://doi.org/10.1016/j.ocemod.2013.05.005, 2020.

De Vries, H., Breton, M., De Mulder, T., Krestenitis, Y., Ozer, J., Proctor, R., Ruddick, K., Salomon, J. C., and Voorrips, A.: A comparison of 2D storm surge models applied to three shallow European seas, Environ. Softw., 10, 23–42, 1995.

Deltares: Delft: Delft3D-FLOW User Manual, 1–727, https://content.oss.deltares.nl/delft3d4/ (last access: 18 November 2024), 2024.

Ebert, E. E. and Mcbride, J. L.: Verification of precipitation in weather systems: determination of systematic errors, J. Hydrol., 239, 179–202, 2000.

Escudier, R., Clementi, E., Cipollone, A., Pistoia, J., Drudi, M., Grandi, A., Lyubartsev, V., Lecci, R., Aydogdu, A., Delrosso, D., Omar, M., Masina, S., Coppini, G., and Pinardi, N.: A High Resolution Reanalysis for the Mediterranean Sea, Front. Earth Sci., 9, 702285, https://doi.org/10.3389/feart.2021.702285, 2021.

Fagherazzi, S., Palermo, C., Rulli, M. C., Carniello, L., and Defina, A.: Wind waves in shallow microtidal basins and the dynamic equilibrium of tidal flats, J. Geophys. Res.-Earth Surf., 112, F02024, https://doi.org/10.1029/2006JF000572, 2007.

Federico, I., Pinardi, N., Coppini, G., Oddo, P., Lecci, R., and Mossa, M.: Coastal ocean forecasting with an unstructured grid model in the southern Adriatic and northern Ionian seas, Nat. Hazards Earth Syst. Sci., 17, 45–59, https://doi.org/10.5194/nhess-17-45-2017, 2017.

Fernández-Montblanc, T., Vousdoukas, M. I., Mentaschi, L., and Ciavola, P.: A Pan-European high resolution storm surge hindcast. Environ. Int., 135, 105367, https://doi.org/10.1016/j.envint.2019.105367, 2020.

Ferrarin, C., Davolio, S., Bellafiore, D., Ghezzo, M., Maicu, F., Mc Kiver, W., Drofa, O., Umgiesser, G., Bajo, M., De Pascalis, F., Malguzzi, P., Zaggia, L., Lorenzetti, G., and Manfe, G.: Cross-scale operational oceanography in the Adriatic Sea, J. Operat. Oceanogr., 12, 86–103, https://doi.org/10.1080/1755876X.2019.1576275, 2019.

Ferrarin, C., Valentini, A., Vodopivec, M., Klaric, D., Massaro, G., Bajo, M., De Pascalis, F., Fadini, A., Ghezzo, M., Menegon, S., Bressan, L., Unguendoli, S., Fettich, A., Jerman, J., Ličer, M., Fustar, L., Papa, A., and Carraro, E.: Integrated sea storm management strategy: the 29 October 2018 event in the Adriatic Sea, Nat. Hazards Earth Syst. Sci., 20, 73–93, https://doi.org/10.5194/nhess-20-73-2020, 2020.

Ferrarin, C., Lionello, P., Orlić, M., Raicich, F., and Salvadori, G.: Venice as a paradigm of coastal flooding under multiple compound drivers. Sci. Rep., 12, 5754, https://doi.org/10.1038/s41598-022-09652-5, 2022.

Giesen, R., Clementi, E., Bajo, M., Federico, I., Stoffelen, A., and Santoleri, R.: The November 2019 record high water level in Venice, Italy. In: Copernicus Marine Service Ocean State Report, Issue 5, J. Operat. Oceanogr., 14:sup1, s156–s162, https://doi.org/10.1080/1755876X.2021.1946240, 2021.

Gilleland, E., Ahijevych, D., Brown, B. G., Casati, B., and Ebert, E. E.: Intercomparison of spatial forecast verification methods, Weather Forecast., 24, 1416–1430, https://doi.org/10.1175/2009WAF2222269.1, 2009.

Gumuscu, I., Islek, F., Yuksel, Y., and Sahin, C.: Spatiotemporal long-term wind and storm characteristics over the eastern Mediterranean Sea, Reg. Stud. Mar. Sci., 63, 102996, https://doi.org/10.1016/j.rsma.2023.102996, 2023.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janiskova, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J. N.: The ERA5 global reanalysis, Q. J. Roy. Meteorol. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hervouet, J.-M. and Bates, P.: The TELEMAC modelling system Special issue, Hydrol. Process., 14, 2207–2208, https://doi.org/10.1002/1099-1085(200009)14:13<2207::aid-hyp22>3.0.co;2-b, 2000.

Hetzel, Y., Janekovic, I., and Pattiaratchi, C.: Assessing the ability of storm surge models to simulate coastal trapped waves around Australia, in: Australasian Coasts & Ports 2017: Working with Nature, Engineers Australia, PIANC Australia and Institute of Professional Engineers New Zealand, https://doi.org/10.3316/informit.929951406285439, 2017.

Lionello, P., Galati, M. B., and Elvini, E.: Extreme storm surge and wind wave climate scenario simulations at the Venetian littoral, Phys. Chem. Earth, 40–41, 86–92, https://doi.org/10.1016/j.pce.2010.04.001, 2010.

Lionello, P., Cavaleri, L., Nissen, K. M., Pino, C., Raicich, F., and Ulbrich, U.: Severe marine storms in the Northern Adriatic: Characteristics and trends, Phys. Chem. Earth, 40–41, 93–105, https://doi.org/10.1016/j.pce.2010.10.002, 2012.

Lionello, P., Barriopedro, D., Ferrarin, C., Nicholls, R. J., Orlić, M., Raicich, F., Reale, M., Umgiesser, G., Vousdoukas, M., and Zanchettin, D.: Extreme floods of Venice: characteristics, dynamics, past and future evolution (review article), Nat. Hazards Earth Syst. Sci., 21, 2705–2731, https://doi.org/10.5194/nhess-21-2705-2021, 2021.

Lovato, T., Androsov, A., Romanenkov, D., and Rubino, A.: The tidal and wind induced hydrodynamics of the composite system Adriatic Sea/Lagoon of Venice, Conti. Shelf Res., 30, 692–706, https://doi.org/10.1016/j.csr.2010.01.005, 2010.

Luettich, R. A., Westerink, J. J., and Scheffner, N. W.: ADCIRC: an advanced three-dimensional circulation model for shelves, coasts, and estuaries. Report 1, Theory and methodology of ADCIRC-2DD1 and ADCIRC-3DL, 1–143, https://apps.dtic.mil/sti/tr/pdf/ADA261608.pdf (last access: 18 November 2024), 1992.

Lyard, F. H., Allain, D. J., Cancet, M., Carrère, L., and Picot, N.: FES2014 global ocean tide atlas: design and performance, Ocean Sci., 17, 615–649, https://doi.org/10.5194/os-17-615-2021, 2021.

Maicu, F., Alessandri, J., Pinardi, N., Verri, G., Umgiesser, G., Lovo, S., Turolla, S., Paccagnella, T., and Valentini, A.: Downscaling with an unstructured coastal-ocean model to the Goro Lagoon and the Po River Delta branches, Front. Mar. Sci., 8, 647781, https://doi.org/10.3389/fmars.2021.647781, 2021.

Međugorac, I., Orlić, M., Janeković, I., Pasaric, Z., and Pasaric, M.: Adriatic storm surges and related cross-basin sea-level slope, J. Mar. Syst., 181, 79–90, https://doi.org/10.1016/j.jmarsys.2018.02.005, 2018.

Mentaschi, L., Besio, G., Cassola, F., and Mazzino, A.: Problems in RMSE-based wave model validations, Ocean Modell., 72, 53–58, https://doi.org/10.1016/j.ocemod.2013.08.003, 2013.

Mentaschi, L., Besio, G., Cassola, F., and Mazzino, A.: Performance evaluation of Wavewatch III in the Mediterranean Sea, Ocean Modell., 90, 82–94, https://doi.org/10.1016/j.ocemod.2015.04.003, 2015.

Mentaschi, L., Vousdoukas, M., Montblanc, T. F., Kakoulaki, G., Voukouvalas, E., Besio, G., and Salamon, P.: Assessment of global wave models on regular and unstructured grids using the Unresolved Obstacles Source Term, Ocean Dynam., 70, 1475–1483, https://doi.org/10.1007/s10236-020-01410-3, 2020.

Mentaschi, L., Vousdoukas, M. I., García-Sánchez, G., Fernández-Montblanc, T., Roland, A., Voukouvalas, E., Federico, I., Abdolali, A., Zhang, Y. J., and Feyen, L.: A global unstructured, coupled, high-resolution hindcast of waves and storm surge, Front. Mar. Sci., 10, 1233679, https://doi.org/10.3389/fmars.2023.1233679, 2023.

Merrifield, M. A., Genz, A. S., Kontoes, C. P., and Marra, J. J.: Annual maximum water levels from tide gauges: Contributing factors and geographic patterns, J. Geophys. Res.-Oceans, 118, 2535–2546, https://doi.org/10.1002/jgrc.20173, 2013.

Micaletto, G., Barletta, I., Mocavero, S., Federico, I., Epicoco, I., Verri, G., Coppini, G., Schiano, P., Aloisio, G., and Pinardi, N.: Parallel implementation of the SHYFEM (System of HydrodYnamic Finite Element Modules) model, Geosci. Model Dev., 15, 6025–6046, https://doi.org/10.5194/gmd-15-6025-2022, 2022.

Mittermaier, M. P.: A strategy for verifying near-convection-resolving model forecasts at observing sites, Weather Forecast., 29, 185–204, https://doi.org/10.1175/WAF-D-12-00075.1, 2014.

Muis, S., Verlaan, M., Winsemius, H. C., Aerts, J. C. J. H., and Ward, P. J.: A global reanalysis of storm surges and extreme sea levels, Nat. Commun., 7, 11969, https://doi.org/10.1038/ncomms11969, 2016.

Muñoz, D. F., Yin, D., Bakhtyar, R., Moftakhari, H., Xue, Z., Mandli, K., and Ferreira, C.: Inter-Model Comparison of Delft3D-FM and 2D HEC-RAS for Total Water Level Prediction in Coastal to Inland Transition Zones, J. Am. Water Resour. Assoc., 58, 34–49, https://doi.org/10.1111/1752-1688.12952, 2022.

Park, K., Federico, I., Di Lorenzo, E., Ezer, T., Cobb, K. M., Pinardi, N., and Coppini, G.: The contribution of hurricane remote ocean forcing to storm surge along the Southeastern U.S. coast, Coast. Eng., 173, 104098, https://doi.org/10.1016/j.coastaleng.2022.104098, 2022.

Orlić, M., Kuzmić, M., and Pasarić, Z.: Response of the Adriatic Sea to the bora and sirocco forcing, Conti. Shelf Res., 14, 91–116, 1994.

Pawlowicz, R., Beardsley, B., and Lentz, S.: Classical harmonic analysis including error estimates in MATLAB using T_TIDE, Comput. Geosci., 28, 929–937, 2022.

Petti, M., Pascolo, S., Bosa, S., Bezzi, A., and Fontolan, G.: Tidal flats morphodynamics: A new conceptual model to predict their evolution over a medium-long period, Water (Switzerland), 11, 1176, https://doi.org/10.3390/w11061176, 2019.

Pineau-Guillou, L., Ardhuin, F., Bouin, M. N., Redelsperger, J. L., Chapron, B., Bidlot, J. R., and Quilfen, Y.: Strong winds in a coupled wave–atmosphere model during a North Atlantic storm event: evaluation against observations, Q. J. Roy. Meteorol. Soc., 144, 317–332, https://doi.org/10.1002/qj.3205, 2018.

Pirazzoli, P. A. and Tomasin, A.: Recent evolution of surge-related events in the Northern Adriatic Sea, J. Coastal Res., 18, 537–554, 2022.

Pringle, W. J., Wirasaet, D., Roberts, K. J., and Westerink, J. J.: Global storm tide modeling with ADCIRC v55: unstructured mesh design and performance, Geosci. Model Dev., 14, 1125–1145, https://doi.org/10.5194/gmd-14-1125-2021, 2021.

Raicich, F.: The sea level time series of Trieste, Molo Sartorio, Italy (1869–2021), Earth Syst. Sci. Data, 15, 1749–1763, https://doi.org/10.5194/essd-15-1749-2023, 2023.

Reimann, L., Vafeidis, A. T., Brown, S., Hinkel, J., and Tol, R.: Mediterranean UNESCO World Heritage at risk from coastal flooding and erosion due to sea-level rise, Nat. Commun., 9, 4161, https://doi.org/10.1038/s41467-018-06645-9, 2018.

Roberts, K. J., Pringle, W. J., and Westerink, J. J.: OceanMesh2D 1.0: MATLAB-based software for two-dimensional unstructured mesh generation in coastal ocean modeling, Geosci. Model Dev., 12, 1847–1868, https://doi.org/10.5194/gmd-12-1847-2019, 2019.

Roberts, N. M. and Lean, H. W.: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events, Mon. Weather Rev., 136, 78–97, https://doi.org/10.1175/2007MWR2123.1, 2008.

Saha, S., Moorthi, S., Pan, H. L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y. T., Chuang, H. Y., Juang, H. M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., van den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., HigginS, W., Zou, C., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R. W., Rutledge, G., and Goldberg, M.: The NCEP climate forecast system reanalysis, B. Am. Meteorol. Soc., 91, 1015–1057, https://doi.org/10.1175/2010BAMS3001.1, 2010.

Saillour, T., Cozzuto, G., Ligorio, F., Lupoi, G., and Bourban, S. E.: Modeling the world oceans with TELEMAC, in: 2020 TELEMAC-MASCARET User Conference, 86–91, https://henry.baw.de/items/a87bc7bc-fc6b-4574-b775-6179721d5d3e (last access: 19 November 2024), 2021.

Skok, G. and Roberts, N.: Analysis of Fractions Skill Score properties for random precipitation fields and ECMWF forecasts, Q. J. Roy. Meteorol. Soc., 142, 2599–2610, https://doi.org/10.1002/qj.2849, 2016.

Staneva, J., Wahle, K., Koch, W., Behrens, A., Fenoglio-Marc, L., and Stanev, E. V.: Coastal flooding: impact of waves on storm surge during extremes – a case study for the German Bight, Nat. Hazards Earth Syst. Sci., 16, 2373–2389, https://doi.org/10.5194/nhess-16-2373-2016, 2016.

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.

Toomey, T., Amores, A., Marcos, M., and Orfila, A.: Coastal sea levels and wind-waves in the Mediterranean Sea since 1950 from a high-resolution ocean reanalysis, Front. Mar. Sci., 9, 991504, https://doi.org/10.3389/fmars.2022.991504, 2022.

Trotta, F., Fenu, E., Pinardi, N., Bruciaferri, D., Giacomelli, L., Federico, I., and Coppini, G.: A structured and unstructured grid relocatable ocean platform for forecasting (SURF), Deep-Sea Res. II, 133, 54–75, https://doi.org/10.1016/j.dsr2.2016.05.004, 2016.

Umgiesser, G., Canu, D. M., Cucco, A., and Solidoro, C.: A finite element model for the Venice lagoon: Development, set up, calibration and validation, J. Marine Syst., 123–145, https://doi.org/10.1016/j.jmarsys.2004.05.009, 2004.

Umgiesser, G., Bajo, M., Ferrarin, C., Cucco, A., Lionello, P., Zanchettin, D., Papa, A., Tosoni, A., Ferla, M., Coraci, E., Morucci, S., Crosato, F., Bonometto, A., Valentini, A., Orlić, M., Haigh, I. D., Nielsen, J. W., Bertin, X., Fortunato, A. B., Pérez Gómez, B., Alvarez Fanjul, E., Paradis, D., Jourdan, D., Pasquet, A., Mourre, B., Tintoré, J., and Nicholls, R. J.: The prediction of floods in Venice: methods, models and uncertainty (review article), Nat. Hazards Earth Syst. Sci., 21, 2679–2704, https://doi.org/10.5194/nhess-21-2679-2021, 2021.

Vannucchi, V., Taddei, S., Capecchi, V., Bendoni, M., and Brandini, C.: Dynamical downscaling of era5 data on the north-western mediterranean sea: From atmosphere to high-resolution coastal wave climate, J. Mar. Sci. Eng., 9, 1–29, https://doi.org/10.3390/jmse9020208, 2021.

Vousdoukas, M. I., Clarke, J., Ranasinghe, R., Reimann, L., Khalaf, N., Duong, T. M., Ouweneel, B., Sabour, S., Iles, C. E., Trisos, C. H., Feyen, L., Mentaschi, L., and Simpson, N. P.: African heritage sites threatened as sea-level rise accelerates, Nat. Clim. Change, 12, 256–262, https://doi.org/10.1038/s41558-022-01280-1, 2022.

Vousdoukas, M. I., Mentaschi, L., Voukouvalas, E., Verlaan, M., Jevrejeva, S., Jackson, L. P., and Feyen, L.: Global probabilistic projections of extreme sea levels show intensification of coastal flood hazard, Nat. Commun., 9, 2360, https://doi.org/10.1038/s41467-018-04692-w, 2018.

Wang, X., Verlaan, M., Veenstra, J., and Lin, H. X.: 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, 2022.

Weatherall, P., Marks, K. M., Jakobsson, M., Schmitt, T., Tani, S., Arndt, J. E., Rovere, M., Chayes, D., Ferrini, V., and Wigley, R.: A new digital bathymetric model of the world's oceans, Earth Space Sci., 2, 331–345, https://doi.org/10.1002/2015EA000107, 2015.

Weisberg, R. H. and Zheng, L.: Hurricane storm surge simulations comparing three-dimensional with two-dimensional formulations based on an Ivan-like storm over the Tampa Bay, Florida region, J. Geophys. Res.-Oceans, 113, C12001, https://doi.org/10.1029/2008JC005115, 2008.

World Meteorological Organization (WMO): Guide to storm surge forecasting, WMO, 120 pp., https://community.wmo.int/en/bookstore/guide-storm-surge-forecasting (last access: 19 November 2024), 2011.

Ye, F., Zhang, Y., Yu, H., Sun, W., Moghimi, S., Myers, E., Nunez, K., Zhang, R., Wang, H., Roland, A., Martins, K., Bertin, X., Du, J., and Liu, Z.: Simulating storm surge and compound flooding events with a creek-to-ocean model: Importance of baroclinic effects, Ocean Modell., 145, 101526, https://doi.org/10.1016/j.ocemod.2019.101526, 2020.

Yu, C. S., Decouttere, C., and Berlamont, J.: Storm surge simulations in the Adriatic Sea, in: CENAS, Coastline Evolution of the Upper Adriatic Sea due to Sea Level Rise and Natural and Anthropogenic Land Subsidence, edited by: Gambolati, G., 207–232 pp., https://doi.org/10.1007/978-94-011-5147-4_10, 1998.

Zaggia, L., Lorenzetti, G., Manfé, G., Scarpa, G. M., Molinaroli, E., Parnell, K. E., Rapaglia, J. P., Gionta, M., and Soomere, T.: Fast shoreline erosion induced by ship wakes in a coastal lagoon: Field evidence and remote sensing analysis, PLoS ONE, 12, e0187210, https://doi.org/10.1371/journal.pone.0187210, 2017.

Zampato, L., Umgiesser, G., and Zecchetto, S.: Storm surge in the Adriatic Sea: observational and numerical diagnosis of an extreme event, Adv. Geosci., 7, 371–378, 2006.

Zampato, L., Umgiesser, G., and Zecchetto, S.: Sea level forecasting in Venice through high resolution meteorological fields. Estuarine, Coastal Shelf Sci., 75, 223–235, https://doi.org/10.1016/j.ecss.2007.02.024, 2007.

Zhang, Y. and Baptista, A. M.: SELFE: A semi-implicit Eulerian–Lagrangian finite-element model for cross-scale ocean circulation, Ocean Modell., 21, 71–96, https://doi.org/10.1016/j.ocemod.2007.11.005, 2008.

Zhang, Y., Ye, F., Stanev, E. V., and Grashorn, S.: Seamless cross-scale modeling with SCHISM, Ocean Model., 102, 64–81, https://doi.org/10.1016/j.ocemod.2016.05.002, 2016.

Zhang, Y. J., Fernandez-Montblanc, T., Pringle, W., Yu, H.-C., Cui, L., and Moghimi, S.: Global seamless tidal simulation using a 3D unstructured-grid model (SCHISM v5.10.0), Geosci. Model Dev., 16, 2565–2581, https://doi.org/10.5194/gmd-16-2565-2023, 2023.

Zingerle, C. and Nurmi, P.: Monitoring and verifying cloud forecasts originating from operational numerical models, Meteorol. Appl., 15, 325–330, https://doi.org/10.1002/met.73, 2008.