



# Application of HIDRA2 Deep Learning Model for Sea Level Forecasting Along the Estonian Coast of the Baltic Sea

Amirhossein Barzandeh<sup>1</sup>, Marko Rus<sup>2,3</sup>, Matjaž Ličer<sup>2,4</sup>, Ilja Maljutenko<sup>1</sup>, Jüri Elken<sup>1</sup>, Priidik Lagemaa<sup>1</sup>, and Rivo Uiboupin<sup>1</sup>

<sup>1</sup>Department of Marine Systems, Tallinn University of Technology, Tallinn, Estonia

<sup>2</sup>Slovenian Environment Agency, Ljubljana, Slovenia

<sup>3</sup>Faculty of Computer and Information Science, Visual Cognitive Systems Lab, University of Ljubljana, Ljubljana, Slovenia

<sup>4</sup>National Institute of Biology, Marine Biology Station, Piran, Slovenia

**Correspondence:** Amirhossein Barzandeh (amirhossein.barzandeh@taltech.ee)

**Abstract.** Sea level predictions, typically derived from 3D hydrodynamic models, are computationally intensive and subject to uncertainties stemming from physical representation and inaccuracies in initial or boundary conditions. As a complementary alternative, data-driven machine learning models provide a computationally efficient solution with comparable accuracy. This study employs the deep learning model HIDRA2 to forecast hourly sea levels at five coastal stations along the Estonian coast-  
5 line of the Baltic Sea, evaluating its performance across various forecast lead times. Compared to the regional NEMO<sub>BAL</sub> and subregional NEMO<sub>EST</sub> hydrodynamic models, HIDRA2 consistently delivers superior results, particularly across all sea level ranges and stations. While HIDRA2 struggles to capture high-frequency variability above  $(6 \text{ h})^{-1}$ , it effectively reproduces energy in lower-frequency bands below  $(18 \text{ h})^{-1}$ . Errors tend to average out over longer time windows encompassing multiple seiche periods, enabling HIDRA2 to surpass the overall performance of the NEMO models. These findings underscore  
10 HIDRA2's potential as a robust, efficient, and reliable tool for operational sea level forecasting and coastal management in the Eastern Baltic Sea region.

## 1 Introduction

The importance of sea level variability has grown significantly in the context of climate change as higher global temperatures accelerate its complexity (Bindoff et al., 2007; Cazenave et al., 2014). The increasing frequency and severity of extreme  
15 weather events, compounded by rising sea levels due to climate change, present pressing challenges for coastal communities worldwide. Accurate and timely forecasting of sea level is essential for effective disaster preparedness, risk mitigation, and coastal management. However, forecasting sea level variability is inherently complex and requires sophisticated models that account for various drivers and uncertainties. Traditionally, sea level forecasting has been accomplished using hydrodynamic models, but these models are subject to a range of uncertainties induced by errors in atmospheric and open boundary forcing  
20 fields, as well as physical flux parametrization (Church et al., 2011; Miles et al., 2014; Khojasteh et al., 2021). Errors in regional sea level prediction can arise from inaccurate initial and boundary conditions, models' approximated descriptions of physical

environments (e.g., coupling with waves, atmosphere, ice, and runoff), and the nonlinear nature of planetary dynamics (Carson et al., 2019; Ponte et al., 2019; Hamlington et al., 2020).

Despite these challenges, advancements such as assimilation of observation data in hydrodynamic models have shown  
25 promise in improving the accuracy of sea level prediction (Bajo et al., 2019; Barron et al., 2004; Tanajura et al., 2015; Liu and Fu, 2018). However, the computational cost of both ocean modeling and data assimilation adds to the complexity and resource requirements of accurately predicting sea levels (Byrne et al., 2023; Bajo et al., 2023). In this regard, the use of data-driven models as a surrogate system can be a promising addition to classic hydrodynamic models for addressing these challenges (Sonnewald et al., 2021). In recent years, machine learning techniques have demonstrated significant potential in  
30 enhancing the accuracy and reliability of predictive models in oceanography (Ahmad, 2019; Lou et al., 2023; Jahanmard et al., 2023). Recent studies have highlighted the effectiveness of deep learning methods in predicting sea level heights across diverse scenarios and basins (Rus et al., 2023; Balogun and Adebisi, 2021; Rajabi-Kiasari et al., 2023; Bahari et al., 2023). However, these studies imply that, unlike hydrodynamic models, deep-learning models require particular structural design and adjustments depending on various objectives, and the specific characteristics of the study area. Additionally, ensemble-  
35 based approaches in deep learning, widely employed in practical applications, distinguish themselves from traditional machine learning methods by generating and integrating multiple hypotheses from training data to improve prediction accuracy. This allows for the incorporation of more advanced and diverse algorithms, making it better suited to tackle specific challenges in operational applications (Zhang et al., 2022; Ganaie et al., 2022).

The Baltic Sea, a large, high-latitude, semi-enclosed, and nearly tideless sea located in northern Europe is a vital marine  
40 region characterized by its unique geographical and ecological features. It comprises several sub-basins and is connected to the Atlantic Ocean through the narrow Danish Straits in the southwest. This study specifically focuses on the Eastern Baltic Sea, which spans from approximately 21.5°E to 30.5°E longitude and 56.5°N to 61°N latitude. This region encompasses the northeastern Baltic Proper, the Gulf of Finland, and the Gulf of Riga—areas of strategic importance due to the heavy vessel traffic between their populated shores and islands. Understanding sea level changes in the Eastern Baltic Sea is therefore vital.  
45 Sea level fluctuations affect navigation routes, harbor operations, and maritime safety, making accurate knowledge of these variations essential for stakeholders to manage risks, optimize vessel operations, and ensure efficient trade flows.

Research on sea level variations in the Baltic Sea has revealed a wide range of patterns and trends influenced by numerous environmental factors that vary both spatially and temporally (Andersson, 2002; Ekman, 2009; Hünicke and Zorita, 2016). Historical tide gauge records, used by Jevrejeva et al. (2003) to reconstruct past sea level changes, demonstrate significant  
50 interannual and decadal variability attributed to atmospheric pressure fluctuations and shifting wind patterns. Furthermore, the intricate structure of the Estonian coastline, combined with its numerous small islands, poses additional challenges for projecting the impacts of climate change and sea-level rise in this region (Johansson et al., 2004; Kont et al., 2008). Although predictions of sea level height on longer temporal scales, such as monthly or weekly, tend to be more accurate due to the smoothing of short-term fluctuations caused by local wind tilt and storm surges (Samuelsson and Stigebrandt, 1996; Scotto  
55 et al., 2009; Elken et al., 2024), short-term sea level changes in the Baltic Sea are heavily influenced by dynamic factors such as wind, precipitation, and the redistribution of water through the Danish Straits (Hünicke et al., 2015). Extreme sea levels



on the Baltic coasts, resulting from increasingly frequent storm surges, are influenced by three main factors: the volume of water in the Baltic Sea's basins, which sets the initial sea level; the direction, speed and duration of tangential wind stresses, whether shoreward or seaward; and the rapid passage of deep low-pressure systems that deform the sea surface, producing seiche-like oscillations that further alter sea levels (Kulikov and Medvedev, 2013; Wolski et al., 2014; Weisse et al., 2021; Wolski and Wiśniewski, 2021; Rutgeresson et al., 2022). Seiche oscillations are especially significant in the Baltic Sea due to their prolonged decay over several days. Both the Gulf of Riga and the Gulf of Finland can experience their own seiches with different characteristics (e.g., the 5-hour seiche period of the Gulf of Riga, which differs from the recognized seiche period in the Gulf of Finland). These seiches may interact with those from neighboring basins (e.g., the approximately 27-hour seiche period of the entire Baltic Sea), leading to complex phase relationships and further complicating sea level predictions (Metzner et al., 2000; Suursaar et al., 2002; Jönsson et al., 2008). This intricate interplay of atmospheric, oceanic, and coastal dynamics makes the prediction of sea level in the Baltic Sea a challenging task.

In response to the increasing demand for accurate and efficient forecasting methods in coastal regions, this study explores the potential of a deep-learning ensemble model. By utilizing deep neural networks, we aim to improve sea level prediction accuracy in the Eastern Baltic Sea by applying our approach to a selection of sample stations. The primary focus of this study is to evaluate the effectiveness of a deep-learning ensemble model HIDRA2 (Rus et al., 2023) for sea level forecasting in the study area in multiple lead time windows. A secondary objective is to compare its performance with that of regional and subregional hydrodynamic models currently used for operational purposes in the study area.

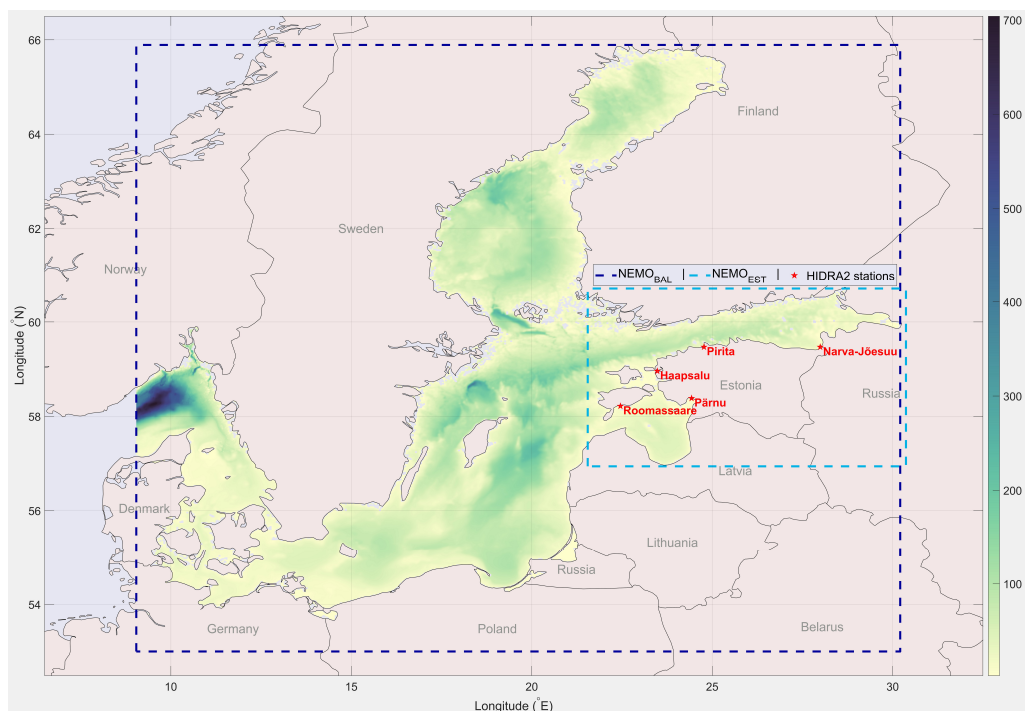
## 2 Data and Methods

### 2.1 Tide gauge data

The hourly sea surface height (SSH) time series were extracted and evaluated as measures of sea level using the available data from five observation stations: Narva-Jõesuu, Pritia, Haapsalu, Pärnu, and Roomassaare (Figure 1). These stations are situated along the Estonian coast, in the Gulf of Finland and the Gulf of Riga. The datasets were provided by the Estonian Environmental Agency (<https://www.ilmateenistus.ee/>).

### 2.2 HIDRA2: A deep-learning ensemble sea level forecasting model

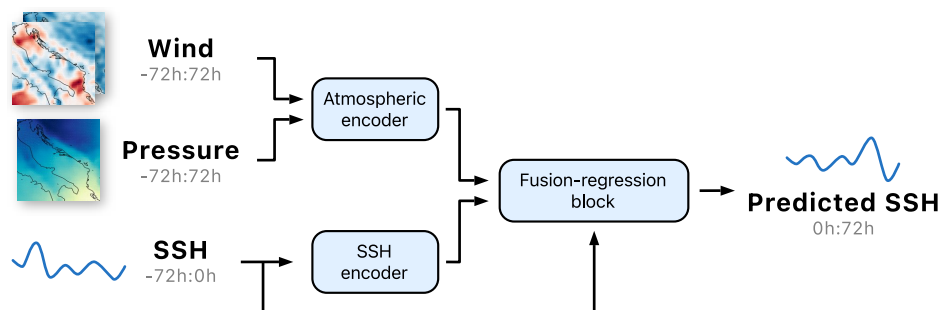
HIDRA2 is a state-of-the-art deep-learning model for predicting SSH at specific geographic locations, presented in depth in Rus et al. (2023). The model utilizes historical and predicted atmospheric data, tidal information, and observed SSH measurements to generate precise hourly SSH predictions for a 72-hour window. Unlike models that forecast the residual SSH, i.e., the difference between SSH and tide, HIDRA2 predicts the full SSH. The HIDRA2 architecture, illustrated in Figure 2, incorporates specialized encoders for processing both atmospheric data and sea level inputs. The Atmospheric encoder processes wind and pressure sequences, subsequently merging them into a unified feature representation. The SSH encoder processes historical SSH data. In the Fusion-regression block, all extracted features with the preceding 72-hour raw SSH data are fused



**Figure 1.** Study area: The domain of each model is illustrated, with red stars indicating the locations of the stations where HIDRA2 has been applied.

and regressed into final SSH forecasts. These encoders consist of a series of convolution layers, activation functions, dropout layers, and skip connections, forming a robust and deep neural network. Details of the encoding architecture are presented in (Rus et al., 2023). The specific difference in this study for the Baltic Sea application is that, unlike in Rus et al. (2023), HIDRA2 was trained on the full SSH data rather than on separate tidal and SSH signals. This decision was made because tides in the Baltic are very weak, and the explicit inclusion of tides as a separate input signal did not improve forecasting performance. Consequently, we removed the tidal encoder block and fused only the atmospheric and SSH features, augmented by the raw SSH measurements, to produce the final set of 72-hour SSH predictions.

For the training of HIDRA2, we utilized atmospheric data fields from the European Centre for Medium-Range Weather Forecasts (ECMWF) Ensemble Prediction System (Leutbecher and Palmer, 2007). For all locations, we used 10-meter winds and mean sea level air pressure from a single atmospheric ensemble member, covering a longitudinal range from 16.25°E to 28.5°E and a latitudinal range from 54.25°N to 64°N. The original meteorological data, with a domain size of 40 × 50, were subsampled to a 9 × 12 grid using linear interpolation to match HIDRA2's required input size. The training data for SSH at the coastal stations were obtained from the Estonian Environmental Agency. It included hourly automated SSH observations from the period 2010 to 2019. Following the methodology outlined by Rus et al. (2023), we trained a separate HIDRA2 model for each of the five geographic locations depicted in Figure 1.



**Figure 2.** This diagram illustrates the HIDRA2 architecture.

### 2.3 NEMO<sub>BAL</sub>: A forecast product based on a regional hydrodynamic modelling

We extracted the Baltic Sea Physics Analysis and Forecast datasets (BALTICSEA\_ANALYSIS\_FORECAST\_PHY\_003\_006) from the Copernicus Marine Service for each station in the study area over the study period. This forecast product is based on the NEMO 4.0 ocean engine, specifically configured for the North Sea and Baltic Sea as Nemo-Nordic 2.0 (Madec et al., 2017; Kärnä et al., 2021), and has been adapted and validated by the Baltic Monitoring and Forecasting Centre (BALMFC) to address the needs of the Baltic Sea region. For brevity, we refer to it as NEMO<sub>BAL</sub> in this study. The product offers a spatial resolution of 1 nautical mile and is updated twice daily, providing new six-day forecasts. Consequently, the dataset records available from Copernicus include only 12-hour forecasts. The detailed validation report is given in Jandt-Scheelke et al. (2023).

### 2.4 NEMO<sub>EST</sub>: A subregional hydrodynamic forecasting model

NEMO-EST05, is the Estonian adaptation of the Baltic Sea regional configuration Nemo-Nordic 2.0 with enhanced horizontal and vertical resolution, tailored for national operational purposes (hereafter referred to as NEMO<sub>EST</sub>). Detailed specifications of this model are provided in Maljutenko et al. (2022), and an example of its utilization is presented in Pärt et al. (2023). NEMO<sub>EST</sub> uses the same NEMO 4.0 ocean engine with a horizontal resolution of 0.5 nautical mile spanning from 21.5°E to 30.5°E and from 56.5°N to 60.5°N. The meteorological forcing for the model system is obtained from the ECMWF 24-hour integrated forecasting system (Owens and Hewson, 2018). Boundary conditions for salinity, temperature, and sea levels are derived from the same NEMO<sub>BAL</sub>, following the methodology outlined by (Elken et al., 2021). Model initialization in the October 2022 is from the NEMO<sub>BAL</sub>. It can be inferred that NEMO<sub>EST</sub> serves as a down-scaled version of NEMO<sub>BAL</sub>. Therefore, using both models in the present study also provides implicit insights into how the spatial resolution of hydrodynamic models as well as their domain extent affects SSH forecasting.

### 2.5 Forecasting and evaluation framework

HIDRA2 and NEMO<sub>EST</sub> provide SSH forecasts for the next 72 hours, segmented into three 24-hour intervals: 0 to 24 hours (24), 24 to 48 hours (48), and 48 to 72 hours (72), with daily updates. This segmentation allows for the evaluation of model



125 performance at different lead times up to 72 hours. Regardless of the training time window used for the deep learning model,  
the testing period of all models in the present study spans one year, from April 1, 2023, to March 31, 2024 (in total, 8784  
hourly time instances). This period aligns with the initiation of archiving 72-hour SSH fields in NEMO<sub>EST</sub> from its daily  
forecast cycles. For comparison, NEMO<sub>BAL</sub> was also considered as the publicly available SSH forecast for the study area.  
To ensure a fair comparison and address potential bias from the reference sea level in the hydrodynamic models (NEMO<sub>BAL</sub>  
130 and NEMO<sub>EST</sub>), the outputs were adjusted by removing their mean value over the study period and then adding the mean  
observed SSH at each station. This adjustment to the datums of the local observation system ensures that the outputs of the  
hydrodynamic models are comparable to the observed values, whereas HIDRA2, as a data-driven model, is not subject to  
this type of reference sea level bias. Additionally, HIDRA2, as an ensemble forecasting system, produces 50 distinct forecast  
ensembles at each time point. In the analysis, unless specifically examining individual ensembles, the HIDRA2 forecasts are  
135 represented by the average of all 50 ensembles for each time instance and station. Using the mean of all ensembles as the final  
forecast is appropriate because the ensembles represent probabilities without any inherent priority in real-world occurrences.

## 2.6 Validation metrics

The validations and comparisons in the present study include the Pearson correlation coefficient (Correlation) and the Root  
Mean Square Deviation (RMSD). Each is defined as follows.

140 RMSD is given by Equation (1):

$$\text{RMSD} = \left( \frac{1}{n} \sum_{i=1}^n (o_i - m_i)^2 \right)^{1/2} \quad (1)$$

where  $n$  represents the total number of observations, and  $o_i$  and  $m_i$  are the observed and modeled sea level values, respectively.

Correlation is defined by Equation (2):

$$\text{Correlation} = \frac{1}{\sigma_m \sigma_o} \sum_{i=1}^n (m_i - \bar{m})(o_i - \bar{o}) \quad (2)$$

145 where  $\bar{m}$  and  $\bar{o}$  are the mean values of the modeled and observed datasets, respectively.

The standard deviations of the modeled and observed datasets,  $\sigma_m$  and  $\sigma_o$ , are calculated as shown in Equation (3):

$$\sigma_m = \left( \sum_{i=1}^n (m_i - \bar{m})^2 \right)^{1/2}, \quad \sigma_o = \left( \sum_{j=1}^n (o_j - \bar{o})^2 \right)^{1/2} \quad (3)$$

## 3 Results

### 3.1 Comparison of models' overall performance

150 Table 1 presents the RMSD values for the 12-hour, 24-hour, and 72-hour lead-time forecasts available during this study across  
all tide gauge stations. At the 24-hour lead time, the HIDRA2 ensemble mean consistently achieves the lowest RMSD at each



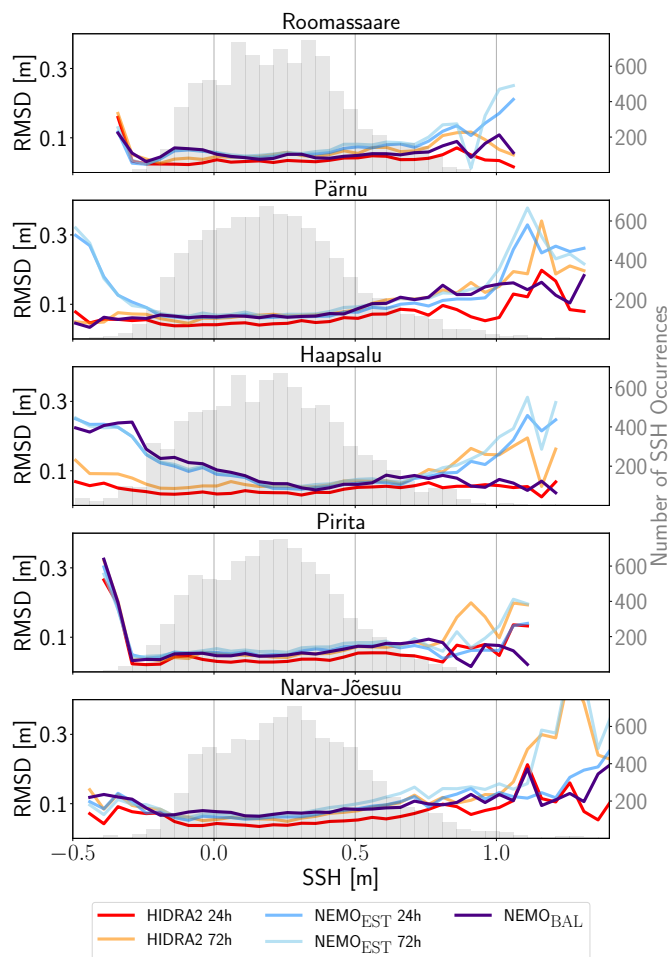
station, outperforming even NEMO<sub>BAL</sub>, which inherently includes 12-hour lead-time forecasting. HIDRA2 with a 72-hour lead-time also demonstrates performance comparable to NEMO<sub>BAL</sub> at Roomassaare and Pärnu. Remarkably, HIDRA2 exhibits substantially lower RMSD at other stations, even with extended forecast lead times.

155 Figure 3 illustrates the forecast RMSD by sea level bin (0.05 m) for all models at different lead times. The RMSDs for each station are calculated using sea level data from April 2023 to April 2024. HIDRA2 with a 24-hour lead time generally outperformed the other models, achieving the lowest RMSD across most of the sea level distribution. However, in the tails of the distribution (SSH > 1.00 m or SSH < -0.40 m), RMSDs increased for all models. At extreme sea levels, HIDRA2 and NEMO<sub>BAL</sub> demonstrated similarly robust performance, albeit with differences depending on lead time. HIDRA2 showed  
 160 solid performance at the 24-hour lead time but struggled in the distribution tails at the 72-hour lead time. Figure 3 highlights that HIDRA2's advantage over NEMO<sub>BAL</sub> primarily stems from its superior performance in the bulk of the distribution and at extreme low SSH values. This distinction is particularly evident at the Haapsalu station.

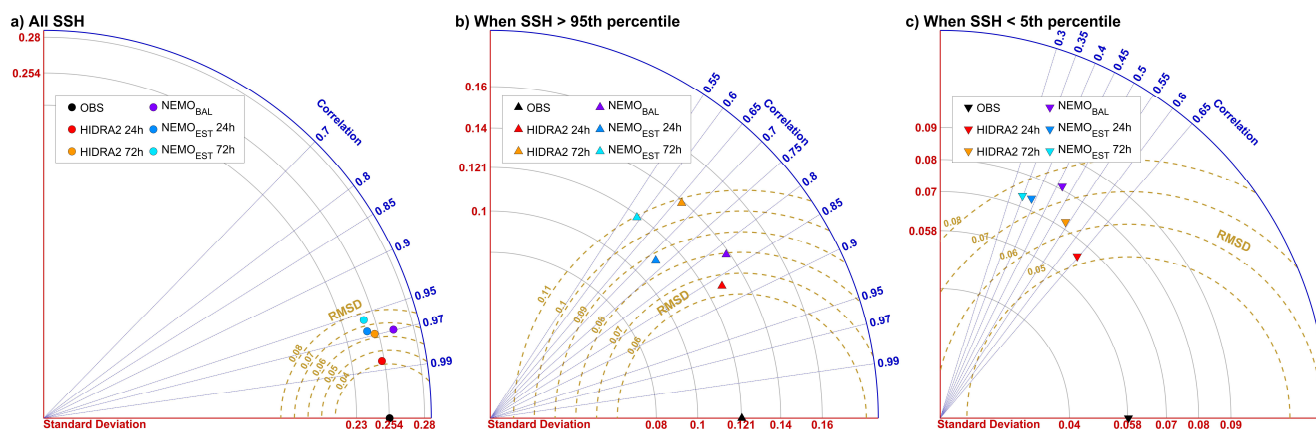
**Table 1.** RMSD metrics of the three forecast models for the April 2023 to April 2024 period. Results are separated by the forecast lead times (12 h, 24 h, and 72 h). Lowest RMSDs are typed bold.

Station	12 h RMSD (m)		24 h RMSD (m)		72 h RMSD (m)	
	NEMO <sub>BAL</sub>	NEMO <sub>EST</sub>	NEMO <sub>EST</sub>	HIDRA2	NEMO <sub>EST</sub>	HIDRA2
Roomassaare	0.051	0.058	<b>0.033</b>	0.063	<b>0.051</b>	
Pärnu	0.074	0.074	<b>0.051</b>	0.083	<b>0.074</b>	
Haapsalu	0.099	0.090	<b>0.040</b>	0.095	<b>0.063</b>	
Pritia	0.056	0.059	<b>0.039</b>	0.067	<b>0.054</b>	
Narva-Jõesuu	0.079	0.074	<b>0.039</b>	0.090	<b>0.075</b>	

Additionally, the accuracy of each model, integrated across all study stations, is compared using Taylor diagrams (Taylor, 2001) in Figure 4a. The performances are also separately assessed during extreme negative and positive observed SSH events  
 165 (Figure 4 b & c). Extreme negative SSH values are defined as those falling below the 5th percentile of observed SSH during the study period at each station, while extreme positive values exceed the 95th percentile (Cannaby et al., 2016; Mentaschi et al., 2023). The 5th and 95th percentile thresholds for each station are provided in Table 2. HIDRA2 consistently outperformed both hydrodynamic models, showing the lowest RMSD and highest correlation coefficients across most observed SSH ranges. Although HIDRA2's accuracy declines for extreme SSH values, particularly at extended lead times, it still surpasses both  
 170 regional and subregional hydrodynamic models, except at the 72-hour lead time. Among the hydrodynamic models, the regional model NEMO<sub>BAL</sub> performs best during extreme SSH events, while the subregional model NEMO<sub>EST</sub> performs better under non-extreme high SSH conditions but struggles to accurately predict extreme SSH values.



**Figure 3.** RMSD of all forecast models (colored lines, see legend at the bottom) at different sea level values. Grey histogram in the back (right axis) shows the distribution of occurrences of a given sea level bin.



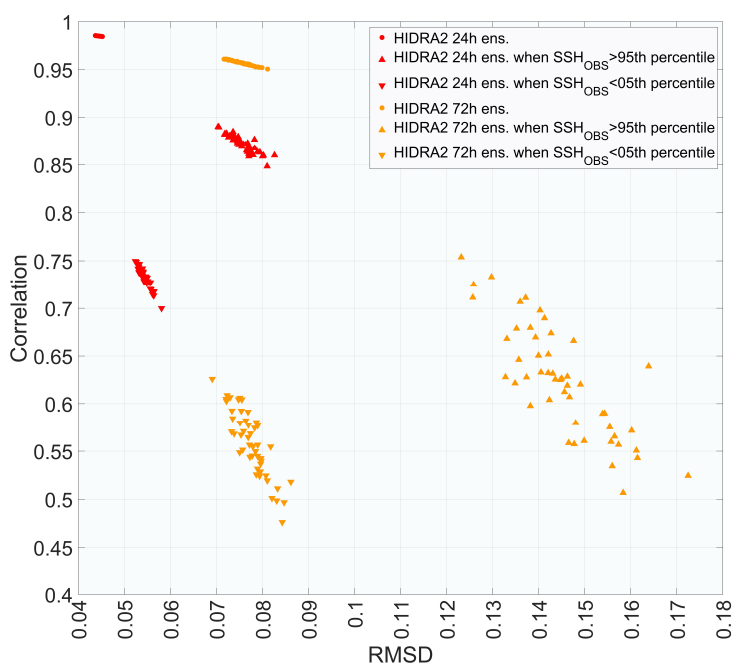
**Figure 4.** Taylor diagrams comparing forecasts: (a) all hourly time instances, (b) positive extreme SSH, and (c) negative extreme SSH.





**Table 2.** Thresholds for SSH used to identify extreme positive and negative storm surges at each station.

Stations	SSH <sub>05th-percentile</sub>	SSH <sub>95th-percentile</sub>
Roomasaare	-0.11	0.67
Pärnu	-0.16	0.71
Haapsalu	-0.21	0.66
Pritia	-0.15	0.60
Narva-Jõesuu	-0.10	0.76

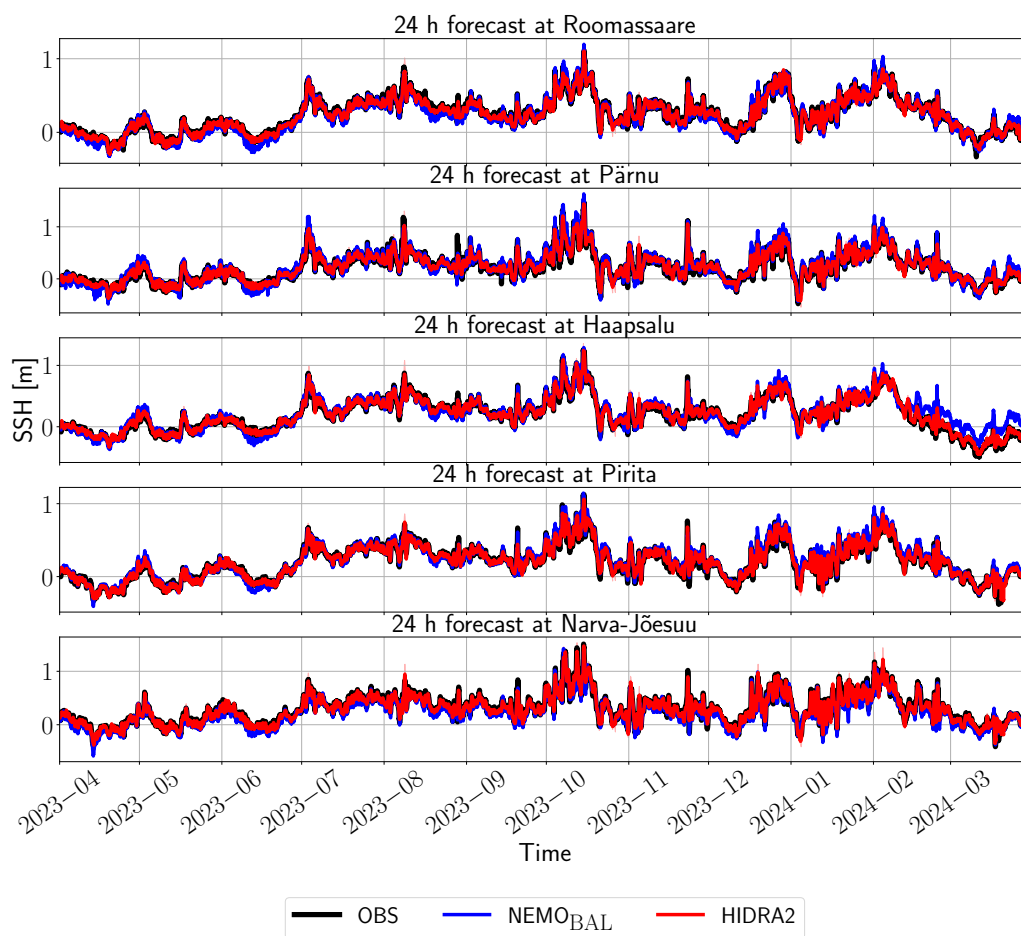


**Figure 5.** Performance of individual HIDRA2 ensemble members at 24-hour and 72-hour lead times and a further assessment under the conditions of the positive and negative extreme observed SSH.

We perform an independent analysis of the predictive accuracy across the 50 members of the HIDRA2 ensemble, evaluating each member's forecasts of SSH values. As expected, increasing the lead time of forecasts results in reduced accuracy of each predicted ensemble member. Both the correlation coefficient and RMSD vary across different lead times, and the spread of ensemble predicted values increases with longer lead times. This spread is especially pronounced during extreme observed SSH values. For negative extreme SSH events, correlation is lower with a lower RMSD, while for positive extreme SSH events, correlation is higher but with a higher RMSD.

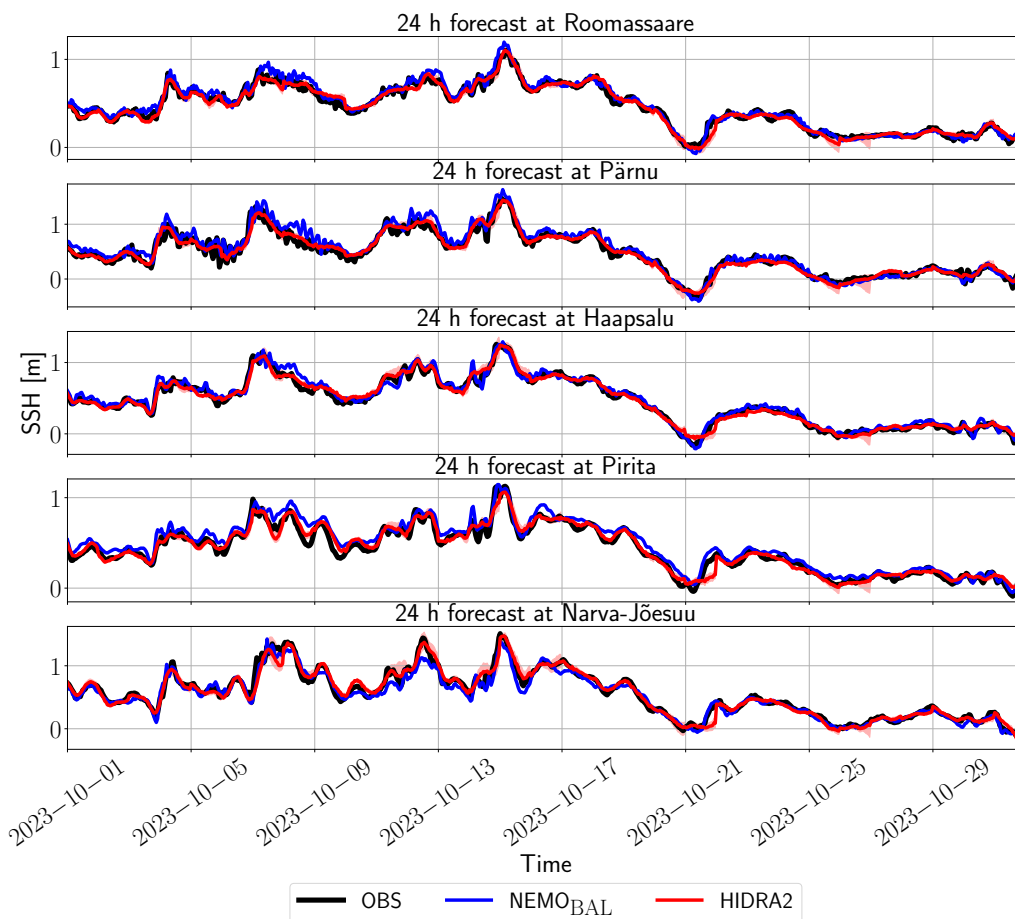
### 3.2 Specific challenges in model performance

180 Figure 6 presents the predictions of HIDRA2 with a 24-hour lead time, alongside those of NEMO<sub>BAL</sub>, for the entire study period at each station. For clarity, the performance of NEMO<sub>EST</sub>, which is the least accurate (as detailed in Section 3.1), is not included in the figure. Both models exhibit stable behavior over the timescales considered.



**Figure 6.** Observed SSH (black line) versus NEMO<sub>BAL</sub> (blue line) and HIDRA2 ensemble mean (red line) forecasts with a 24 h lead time at various stations between April 2023 and April 2024.

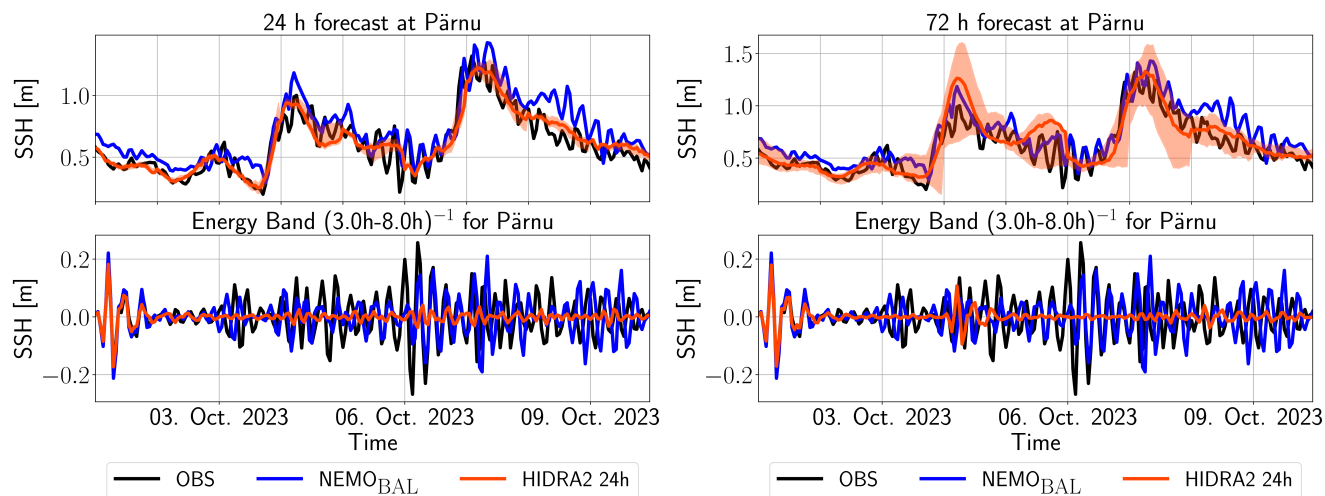
185 However, a closer examination, as shown in Figure 7, reveals that HIDRA2 predictions are overly smoothed, lacking the ability to fully capture high-frequency variability (with periods below 6 hours). In contrast, NEMO<sub>BAL</sub> demonstrates better performance in this energy band. Figure 7 highlights model forecasts with a 24-hour lead time during October 2023, a period marked by a series of extreme sea-level events.



**Figure 7.** Same as Figure 6, but only for October 2023.

A visual comparison of model predictions on Figure 7 between 5 and 9 October 2023 is quite telling. Sea levels in Pärnu indicate the presence of the fundamental seiche with a period around 5 h. NEMO<sub>BAL</sub> is reproducing this excitation very clearly while it is practically completely absent from the HIDRA2 forecast. A similar scenario occurs on 13 October - we have high frequency variability in NEMO<sub>BAL</sub> and a completely smooth HIDRA2 ensemble mean. One might argue that this smoothing stems from the fact that we are working with the ensemble mean in which otherwise present oscillations in different ensemble members cancel out. This is however not the case. None of the ensemble members exhibit enough excitations in the Pärnu seiche energy band (not shown).

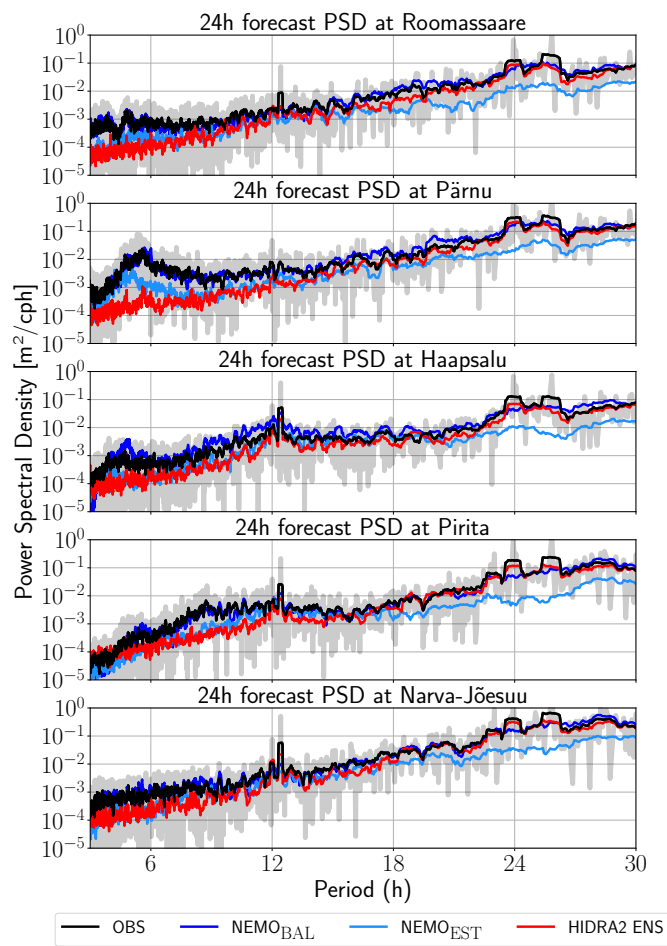
The limitations of HIDRA2 during seiche excitations can be evaluated by applying a band-pass filter to the sea-level signal in the Pärnu seiche energy band for the period of excited seiches between 5 and 9 October 2023. We use a 6th order Butterworth filter with a sample rate of 1 h and low and high cutoff frequencies of  $(3\text{h})^{-1}$  and  $(8\text{h})^{-1}$ , respectively (Figure 8). A clear excitation of the seiche is visible in filtered observations in Figure 8 from 3 October on NEMO<sub>BAL</sub> captures this excitation



**Figure 8. Top:** Sea levels in Pärnu during the seiche between 5 - 9 October 2023 for 24 h (left column) and 72 h (right column) lead times. Orange area denotes the limits of the HIDRA2 ensemble envelope. **Bottom:** Bandpass filtered sea levels in the frequency band  $(3h)^{-1} - (8h)^{-1}$  during the same period and for the same lead times. Note different vertical scales on both plots.

to some extent and successfully matches the seiche phase during periods of highest amplitudes. In contrast, the HIDRA2 ensemble mean is overly smooth, rendering it less capable of reproducing sea-level variability within the seiche frequency band. However, it adheres more closely to the overall observations than NEMO<sub>BAL</sub>. HIDRA2's limited accuracy during seiche excursions leads to deviations from instantaneous sea-level values, increasing the short-term errors. However, over periods of several days, HIDRA2 exhibits minimal bias. This indicates that while individual hourly predictions may show higher error, HIDRA2's forecasts do not consistently over- or underestimate SSH over longer time spans, resulting in minimal systematic bias.

This experiment also demonstrates the value of the ensemble prediction. While separate members of the HIDRA2 ensemble (and consequently their ensemble mean as well) are not capable of capturing the high-frequency variability of the seiche (not shown), this limitation is partly compensated by the ensemble spread, which to some extent envelops the actual amplitude of the seiches. Therefore the sea levels during the seiche will often still be within the ensemble envelope although the seiche itself is poorly reproduced in each individual ensemble member. This ensemble approach in HIDRA2 is especially useful for addressing the challenges of triggering events and managing uncertainty in sea level forecasts. Triggering events often involve abrupt and transient atmospheric or oceanic changes, like storms or pressure systems, which lead to rapid sea level shifts. While HIDRA2's ensemble mean may not directly capture these fluctuations, the ensemble spread creates a "buffer" around the observed values. This means that during trigger events, even if individual ensemble members miss exact timings or magnitudes, the ensemble as a whole may still envelop the observed extremes, providing a probabilistic signal of potential risks.



**Figure 9.** Power spectral densities [ $\text{m}^2/\text{cph}$ ] of sea levels from different sources (colored lines, see legend at the bottom) at each station (rows). Denotation cph indicates cycles per hour.

To quantify  $\text{NEMO}_{\text{BAL}}$ , and  $\text{HIDRA2}$  behavior in different spectral bands, we computed power spectral densities (PSD) for observations and model forecasts for each of the stations, as shown on Figure 9. PSD was calculated by normalizing the squared magnitude of the Fast Fourier Transform coefficients at each frequency, followed by an inversion of frequencies to periods. This allowed us to identify dominant seiche periods and the (ir)relevance of tidal forcing (peaks around 12 and 24 h).  
220 In Pärnu, for example, the fundamental seiche peak close to 6 h period is quite prominent, while the semi-diurnal 12 h peak does not stand out in its energy band, indicating that semi-diurnal tides are not driving the dynamics at this station. At other stations, for example, semi-diurnal tides are a bit more pronounced.

The spectra reveal that  $\text{HIDRA2}$  faces challenges in fully capturing seiche dynamics and tends to underestimate energy density in the subinertial energy band (at the latitude of Pärnu, this corresponds to periods below 13 h or rather frequencies  
225 above  $(13 \text{ h})^{-1}$ ) - in this band  $\text{NEMO}_{\text{BAL}}$  accurately reproduces the observed amount of energy.  $\text{HIDRA2}$  spectra are however



the closest to observations spectra in the band with periods above 18 h. Especially at diurnal tide band HIDRA2 shows slightly better performance than NEMO<sub>BAL</sub>, while both HIDRA2 and NEMO<sub>BAL</sub> outperform NEMO<sub>EST</sub>. This corresponds to the solid overall performance of HIDRA2 and its poor reproduction of high frequency variability. NEMO<sub>EST</sub> on the other hand lacks almost an order of magnitude of energy in the energy band with periods above 18 h, but indicates a decent reproduction of the  
230 Pärnu seiche.

#### 4 Discussion

While the hydrodynamic model NEMO<sub>BAL</sub> may be particularly more effective at detecting specific oscillations, such as the excited seiches, HIDRA2's adaptability makes it better suited for a wider range of sea level conditions. HIDRA2 consistently achieves lower RMSD values and frequently outperforms hydrodynamic models in predictive accuracy. This suggests that  
235 the accuracy of hydrodynamic models in capturing short-term oscillations largely stems from their foundation in physical equations, including the use of temporal derivatives with high-resolution time-stepping, and their broad modeling domains, which enable responsive adaptation to changing environmental conditions across various scales. This approach suggests a deeper, more intrinsic understanding of the system's natural behavior. As a result, although data-driven models are adept at detecting data value fluctuations, they lack the inherent physical insight that hydrodynamic models possess (Zhao et al., 2024).  
240 Nonetheless, future advancements in areas such as physics-informed deep learning can provide strong potential to enhance data-driven models for ocean parameter prediction, equipping them with a more nuanced, system-aware perspective (Raissi et al., 2019; Donnelly et al., 2024). However, developing an effective physics-informed deep learning model tailored to sea level prediction remains significantly more challenging than for other oceanic parameters, as sea level dynamics involve continuity and momentum equations that govern the conservation of mass and momentum under complex external influences.

245 An additional advantage of deep learning models, such as HIDRA2, lies in their computational efficiency. In the present study, HIDRA2 demonstrates the ability to generate 50 ensemble predictions for each 72-hour forecast at each station in approximately 30 seconds using a current typical personal computer or laptop (with performance rates typically ranging within the teraflops scale), using only atmospheric input data (wind and mean sea level pressure), background sea level, and the trained network file. In contrast, running a high-resolution hydrodynamic model for even a single ensemble simulation requires sig-  
250 nificant computational resources, a large amount of input data, and considerable processing time, often necessitating dedicated high-performance computing facilities. Consequently, HIDRA2's efficiency not only enhances forecast accessibility but also enables a more adaptable and resource-effective approach to sea level prediction in operational settings, providing valuable support for optimizing processes in coastal management.

Furthermore, one inherent limitation of hydrodynamic models, particularly in accounting for long-term variability, lies in the  
255 incorporation of a realistic vertical datum, which requires adjustments for dynamic geoid height variations and land movements (Jahanmard et al., 2021; Mostafavi et al., 2023). These adjustments pose additional challenges to the accuracy of traditional models, which rely on stable reference levels and may struggle to account for local geophysical changes (Jahanmard et al., 2022). This constraint further underscores the advantage of deep learning models like HIDRA2, which do not rely on prede-



260 fined or geophysically biased dynamic topography but instead adapt more flexibly to complex patterns within large datasets, capturing broad behaviors and fine details with high fidelity.

Besides, validation metrics, such as RMSD and correlation coefficient, presented in this study underscore HIDRA2's superior performance compared to other sea level forecasting methods, as demonstrated in studies e.g. Ishida et al. (2020); Shahabi and Tahvildari (2024); Dong et al. (2024). Despite the influence of various regional factors on short-term SSH forecasts, the primary distinction between HIDRA2 and methods used in previous studies is the utilization of a deep-learning "ensemble" approach. Thus, the effectiveness of HIDRA2 in this case study underscores the importance of incorporating diverse ensembles to account for varying short-term atmospheric conditions and climate scenarios (Gröger et al., 2019; Hieronymus and Hieronymus, 2023). For operational applications, particularly in decision-making and initiating specific response phases, HIDRA2 can be more effective when the triggering limits are set to the maximum and minimum ensemble values, as these represent the most probable scenarios. Although the probability of reaching the maximum or minimum ensemble values in typical cases is around 2%—given the 50 ensembles used in this study—the ensemble spread, which almost always encompasses the absolute SSH values, suggests that HIDRA2 is a more reliable tool for operational forecasting and triggering. This reliability is especially notable when compared to other forecasting systems that do not provide ensemble predictions. Future studies could further explore optimizing this ensemble range. In the present study, it has been shown that the spread of ensemble-predicted values increases with longer lead times, with this spread becoming particularly pronounced during extreme observed SSH values. When considering a separate evaluation of negative and positive extreme SSH events, the results show that for negative extreme SSH events, the correlation is lower with a narrower RMSD, while for positive extremes, the correlation is higher but accompanied by a wider RMSD. Here, the significance of comparing RMSDs may be less relevant, as higher RMSD values can be acceptable when target values are higher. But, in general, the lower correlation during negative SSH extremes suggests that while the model captures some general trends or patterns, it does not align well with the exact fluctuations or magnitudes of observed values. From a data-driven modeling perspective, this may be explained by data imbalance: negative extreme events are rarer than positive extremes in the study area (Conte and Lionello, 2013; Jensen et al., 2022), providing HIDRA2 with fewer examples of negative events during training. Consequently, HIDRA2 may perform better for positive extreme events, having learned from their relatively higher frequency. This observation demonstrates the model's enhanced capability to capture more common extreme event types.

## 285 5 Conclusions

This study evaluates an implementation of HIDRA2, a deep-learning sea level model, alongside two numerical models, NEMO<sub>EST</sub>, a subregional hydrodynamic model with a spatial resolution of 0.5 nautical miles, and NEMO<sub>BAL</sub>, a regional hydrodynamic model with a spatial resolution of 1 nautical mile and a 12-hour lead time. The study area focuses on five tide gauge stations along the Estonian coast of the Baltic Sea: Narva-Jõesuu, Pritia, Haapsalu, Pärnu, and Roomasaare, which provided hourly SSH observations for the study period from April 1, 2023, to March 31, 2024. In summary, based on the



averaged correlation coefficient and RMSD across the study period between modeled and observed SSH, the results indicate that HIDRA2 generally outperforms the hydrodynamic models.

While accuracy decreases in predicting extreme sea level events, the observed values generally fall within the predicted ensemble spread of HIDRA2, highlighting its overall reliability. However, some open issues remain. HIDRA2 did not perfectly reproduce seiches in Pärnu bay and in general lacks energy in the subinertial band with periods below 13 h. Reproduction of sea level dynamics in this band is something that will have to be addressed in further work. This can be partly offset by an ensemble spread but the spread with 24 h lead time is underestimated while the spread for 72 h seems overestimated, as can be seen from Figure 8. This remains a challenge for future investigations.

Nevertheless, by utilizing HIDRA2's ensemble spread, stakeholders can gain greater confidence in the forecast range rather than relying solely on a deterministic forecast. For example, if the ensemble spread widens around an event, it indicates a higher level of uncertainty and suggests preparing for a range of scenarios, rather than a single outcome. Ensemble models like HIDRA2 can be set to generate alerts when the spread surpasses a specific threshold, signaling a high probability of an unusual event. This probabilistic signaling can be essential for coastal management, providing early warnings for potential risks in the face of triggering events, even if the ensemble mean remains stable.

Deep-learning ensemble models, such as HIDRA2, are pertinent for advancing the development of Digital Twins and associated impact models (Li et al., 2023). These models, utilizing ensemble-based techniques, are particularly effective in capturing the complex, non-linear relationships in SSH data across diverse scales. By integrating multiple predictive models, ensemble approaches enhance the accuracy and robustness of forecasts, making them valuable for the creation of Digital Twins of the Earth systems. This, in turn, supports more precise impact assessments and decision-making processes in coastal management and risk mitigation.

*Code availability.* The persistent version of the GitHub repository containing HIDRA2 code is available at <https://doi.org/10.5281/zenodo.7307365>.

*Data availability.* SSH observation data for each studied station are available from the Estonian Environmental Agency (<https://www.ilmateenistus.ee/>). Atmospheric inputs used for training and running the models in this study can be obtained from ECMWF (<https://www.ecmwf.int/en/forecasts/>). Baltic Sea Physics Analysis and Forecast datasets (referred to as NEMO<sub>BAL</sub> in this study), are accessible through the EU Copernicus Marine Service under the dataset identifier BALTICSEA\_ANALYSISFORECAST\_PHY\_003\_006 (<http://marine.copernicus.eu/>).

*Author contributions.* All authors contributed to discussions and editing of the paper. **A.B.** conducted model runs and data extractions for the study area, pre- and post-processed the outputs, analyzed and prepared the final results, and drafted the original manuscript. **M.R.** prepared and trained the HIDRA2 model for the case study. **M.L.** performed spectral analyses and wrote parts of the manuscript. **A.B.** and **M.L.**





plotted the Figures. **M.L.**, **I.M.**, and **R.U.** contributed to the design and conceptualization of the study and provided advisory input on the methodology. **I.M.**, **J.E.**, and **P.L.** conducted the NEMO<sub>EST</sub> runs and prepared and preprocessed the related data.

*Competing interests.* The authors declare that they have no conflict of interest.

*Financial support.* This research was funded by the EU through agreement DE\_330\_MF between ECMWF and Météo-France. Additional support was provided by the EU and the Estonian Research Council under project TEM-TA38 (Digital Twin of Marine Renewable Energy). Furthermore, this work received funding through the AdapEST project ("Implementation of national climate change adaptation activities in Estonia, VEU23019"), supported by the European Climate, Infrastructure and Environment Executive Agency (CINEA) via the LIFE programme. **M.L.** acknowledges the financial support from the Slovenian Research and Innovation Agency ARIS (contract no. P1-0237).



## References

- 330 Ahmad, H.: Machine learning applications in oceanography, *Aquatic Research*, 2, 161–169, <https://doi.org/10.3153/AR19014>, 2019.
- Andersson, H.: Influence of long-term regional and large-scale atmospheric circulation on the Baltic sea level, *Tellus A Dyn. Meteorol. Oceanogr.*, 54, 76–88, <https://doi.org/10.3402/tellusa.v54i1.12125>, 2002.
- Bahari, N. A. A. B. S., Ahmed, A. N., Chong, K. L., Lai, V., Huang, Y. F., Koo, C. H., Ng, J. L., and El-Shafie, A.: Predicting Sea Level Rise Using Artificial Intelligence: A Review, *Archives of Computational Methods in Engineering*, 30, 4045–4062, <https://doi.org/10.1007/s11831-023-09934-9>, 2023.
- 335 Bajo, M., Međugorac, I., Umgiesser, G., and Orlić, M.: Storm surge and seiche modelling in the Adriatic Sea and the impact of data assimilation, *Quarterly Journal of the Royal Meteorological Society*, 145, 2070–2084, <https://doi.org/https://doi.org/10.1002/qj.3544>, 2019.
- Bajo, M., Ferrarin, C., Umgiesser, G., Bonometto, A., and Coraci, E.: Modelling the barotropic sea level in the Mediterranean Sea using data assimilation, *Ocean Sci.*, 19, 559–579, <https://doi.org/10.5194/os-19-559-2023>, 2023.
- 340 Balogun, A.-L. and Adebisi, N.: Sea level prediction using ARIMA, SVR and LSTM neural network: assessing the impact of ensemble Ocean-Atmospheric processes on models' accuracy, *Geomatics, Natural Hazards and Risk*, 12, 653–674, <https://doi.org/10.1080/19475705.2021.1887372>, 2021.
- BALTICSEA\_ANALYSIS\_FORECAST\_PHY\_003\_006: E.U. Copernicus Marine Service Information (CMEMS). Marine Data Store (MDS), <https://doi.org/10.48670/moi-00010>, accessed on 02-Apr-2024.
- 345 Barron, C., Birol Kara, A., Hurlburt, H., Rowley, C., and Smedstad, L.: Sea surface height predictions from the global Navy Coastal Ocean Model during 1998–2001, *J. Atmos. Ocean. Technol.*, 21, 1876–1893, <https://doi.org/10.1175/JTECH-1680.1>, 2004.
- Bindoff, N., Willebrand, J., Artale, V., Cazenave, A., Gregory, J., Gulev, S., Hanawa, K., Le Quere, C., Levitus, S., and Nojiri, Y.: Observations: oceanic climate change and sea level, <https://nora.nerc.ac.uk/id/eprint/15400>, 2007.
- Byrne, D., Horsburgh, K., and Williams, J.: Variational data assimilation of sea surface height into a regional storm surge model: Benefits and limitations, *J. Oper. Oceanogr.*, 16, 1–14, <https://doi.org/10.1080/1755876X.2021.1884405>, 2023.
- 350 Cannaby, H., Palmer, M. D., Howard, T., Bricheno, L., Calvert, D., Krijnen, J., Wood, R., Tinker, J., Bunney, C., Harle, J., Saulter, A., O'Neill, C., Bellingham, C., and Lowe, J.: Projected sea level rise and changes in extreme storm surge and wave events during the 21st century in the region of Singapore, *Ocean Science*, 12, 613–632, <https://doi.org/10.5194/os-12-613-2016>, 2016.
- Carson, M., Lyu, K., Richter, K., Becker, M., Domingues, C. M., Han, W., and Zanna, L.: Climate model uncertainty and trend detection in regional sea level projections: A review, *Surveys in Geophysics*, 40, 1631–1653, <https://doi.org/10.1007/s10712-019-09559-3>, 2019.
- 355 Cazenave, A., Dieng, H.-B., Meyssignac, B., Von Schuckmann, K., Decharme, B., and Berthier, E.: The rate of sea-level rise, *Nat. Clim. Chang.*, 4, 358–361, <https://doi.org/10.1038/nclimate2159>, 2014.
- Church, J., Gregory, J., White, N., Platten, S., and Mitrovica, J.: Understanding and projecting sea level change, *Oceanography*, 24, 130–143, <https://doi.org/10.5670/oceanog.2011.33>, 2011.
- 360 Conte, D. and Lionello, P.: Characteristics of large positive and negative surges in the Mediterranean Sea and their attenuation in future climate scenarios, *Global and Planetary Change*, 111, 159–173, <https://doi.org/j.gloplacha.2013.09.006>, 2013.
- Dong, Z., Hu, H., Liu, H., Baiyin, B., Mu, X., Wen, J., Liu, D., Chen, L., Ming, G., Chen, X., and Li, X.: Superior performance of hybrid model in ungauged basins for real-time hourly water level forecasting – A case study on the Lancang-Mekong mainstream, *Journal of Hydrology*, 633, 130941, <https://doi.org/10.1016/j.jhydrol.2024.130941>, 2024.



- 365 Donnelly, J., Daneshkhah, A., and Abolfathi, S.: Physics-informed neural networks as surrogate models of hydrodynamic simulators, *Science of The Total Environment*, 912, 168 814, <https://doi.org/https://doi.org/10.1016/j.scitotenv.2023.168814>, 2024.
- Ekman, M.: The changing level of the Baltic Sea during 300 years: a clue to understanding the Earth, <https://www.historicalgeophysics.ax/The%20Changing%20Level%20of%20the%20Baltic%20Sea.pdf>, 2009.
- Elken, J., Maljutenko, I., Lagemaa, P., and Verjovkina, S.: Mere operatiivmodelisüsteemi NEMO kasutuselevõtt ja töölerakendamine  
370 mereala operatiivprognooside parandamiseks, Tech. Rep. 82, TTÜ Meresüsteemide Inst., [https://keskkonnaportaal.ee/sites/default/files/andmekataloog/lisakirjeldus/NEMO\\_II\\_aruanne2021\\_v2.pdf](https://keskkonnaportaal.ee/sites/default/files/andmekataloog/lisakirjeldus/NEMO_II_aruanne2021_v2.pdf), 2021.
- Elken, J., Barzandeh, A., Maljutenko, I., and Rikka, S.: Reconstruction of Baltic Gridded Sea Levels from Tide Gauge and Altimetry Observations Using Spatiotemporal Statistics from Reanalysis, *Remote Sensing*, 16, <https://doi.org/10.3390/rs16152702>, 2024.
- Ganaie, M., Hu, M., Malik, A., Tanveer, M., and Suganthan, P.: Ensemble deep learning: A review, *Engineering Applications of Artificial Intelligence*, 115, 105 151, <https://doi.org/10.1016/j.engappai.2022.105151>, 2022.  
375
- Gröger, M., Arneborg, L., Dieterich, C., Höglund, A., and Meier, H. E. M.: Summer hydrographic changes in the Baltic Sea, Kattegat and Skagerrak projected in an ensemble of climate scenarios downscaled with a coupled regional ocean–sea ice–atmosphere model, *Climate Dynamics*, 53, 5945–5966, <https://doi.org/10.1007/s00382-019-04908-9>, 2019.
- Hamlington, B. D., Gardner, A. S., Ivins, E., Lenaerts, J. T., Reager, J., Trossman, D. S., Zaron, E. D., Adhikari, S., Arendt, A., Aschwanden,  
380 A., et al.: Understanding of contemporary regional sea-level change and the implications for the future, *Reviews of Geophysics*, 58, e2019RG000 672, <https://doi.org/10.1029/2019RG000672>, 2020.
- Hieronimus, M. and Hieronymus, F.: A Novel Machine Learning Based Bias Correction Method and Its Application to Sea Level in an Ensemble of Downscaled Climate Projections, *Tellus A: Dynamic Meteorology and Oceanography*, <https://doi.org/10.16993/tellusa.3216>, 2023.
- 385 Hünicke, B. and Zorita, E.: Statistical Analysis of the Acceleration of Baltic Mean Sea-Level Rise, 1900–2012, *Frontiers in Marine Science*, 3, <https://doi.org/10.3389/fmars.2016.00125>, 2016.
- Hünicke, B., Zorita, E., Soomere, T., Madsen, K., Johansson, M., and Suursaar, U.: Recent change—Sea level and wind waves, in: *Second Assess. Clim. Chang. Balt. Sea basin*, pp. 155–185, Springer, [https://doi.org/10.1007/978-3-319-16006-1\\_9](https://doi.org/10.1007/978-3-319-16006-1_9), 2015.
- Ishida, K., Tsujimoto, G., Ercan, A., Tu, T., Kiyama, M., and Amagasaki, M.: Hourly-scale coastal sea level modeling  
390 in a changing climate using long short-term memory neural network, *Science of The Total Environment*, 720, 137 613, <https://doi.org/10.1016/j.scitotenv.2020.137613>, 2020.
- Jahanmard, V., Delpeche-Ellmann, N., and Ellmann, A.: Realistic dynamic topography through coupling geoid and hydrodynamic models of the Baltic Sea, *Cont. Shelf Res.*, 222, 104 421, <https://doi.org/10.1016/j.csr.2021.104421>, 2021.
- Jahanmard, V., Delpeche-Ellmann, N., and Ellmann, A.: Towards realistic dynamic topography from coast to offshore by incorporating  
395 hydrodynamic and geoid models, *Ocean Model.*, 180, 102 124, <https://doi.org/10.1016/j.ocemod.2022.102124>, 2022.
- Jahanmard, V., Hordoir, R., Delpeche-Ellmann, N., and Ellmann, A.: Quantification of hydrodynamic model sea level bias utilizing deep learning and synergistic integration of data sources, *Ocean Modelling*, 186, 102 286, <https://doi.org/https://doi.org/10.1016/j.ocemod.2023.102286>, 2023.
- Jandt-Scheelke, S., Panteleit, T., Verjovkina, S., Lagemaa, P., Spruch, L., and Morrison, H.: Quality Information Document: Baltic Sea Physical Analysis and Forecasting Product (BALTICSEA\_ANALYSIS\_FORECAST\_PHY\_003\_006), <https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-BAL-QUID-003-006.pdf>, cMEMS product quality coordination team, 2023.  
400



- Jensen, C., Mahavadi, T., Schade, N. H., Hache, I., and Kruschke, T.: Negative Storm Surges in the Elbe Estuary—Large-Scale Meteorological Conditions and Future Climate Change, *Atmosphere*, 13, <https://doi.org/10.3390/atmos13101634>, 2022.
- Jevrejeva, S., Moore, J., and Grinsted, A.: Influence of the Arctic Oscillation and El Niño-Southern Oscillation (ENSO) on ice conditions in the Baltic Sea: The wavelet approach, *J. Geophys. Res. Atmos.*, 108, <https://doi.org/10.1029/2003JD003417>, 2003.
- 405 Johansson, M., Kahma, K., Boman, H., and Launiainen, J.: Scenarios for sea level on the Finnish coast, *Boreal Environ. Res.*, 9, 153, <https://www.borenav.net/BER/archive/pdfs/ber9/ber9-153.pdf>, 2004.
- Jönsson, B., Döös, K., Nycander, J., and Lundberg, P.: Standing waves in the Gulf of Finland and their relationship to the basin-wide Baltic seiches, *Journal of Geophysical Research: Oceans*, 113, <https://doi.org/https://doi.org/10.1029/2006JC003862>, 2008.
- 410 Khojasteh, D., Glamore, W., Heimhuber, V., and Felder, S.: Sea level rise impacts on estuarine dynamics: A review, *Sci. Total Environ.*, 780, 146 470, <https://doi.org/10.1016/j.scitotenv.2021.146470>, 2021.
- Kont, A., Jaagus, J., Aunap, R., Ratas, U., and Ravis, R.: Implications of sea-level rise for Estonia, *J. Coast. Res.*, 24, 423–431, <https://doi.org/10.2112/07A-0015.1>, 2008.
- Kulikov, E. and Medvedev, I.: Variability of the Baltic Sea level and floods in the Gulf of Finland, *Oceanology*, 53, 145–151, <https://doi.org/10.1134/S0001437013020094>, 2013.
- 415 Kärnä, T., Ljungemyr, P., Falahat, S., Ringgaard, I., Axell, L., Korabel, V., Murawski, J., Maljutenko, I., Lindenthal, A., and Jandt-Scheelke, S.: Nemo-Nordic 2.0: Operational marine forecast model for the Baltic Sea, *Geosci. Model Dev.*, 14, 5731–5749, <https://doi.org/10.5194/gmd-14-5731-2021>, 2021.
- Leutbecher, M. and Palmer, T.: Ensemble forecasting, <https://doi.org/10.21957/c0hq4yg78>, 2007.
- 420 Li, X., Feng, M., Ran, Y., Su, Y., Liu, F., Huang, C., Shen, H., Xiao, Q., Su, J., Yuan, S., et al.: Big Data in Earth system science and progress towards a digital twin, *Nature Reviews Earth & Environment*, 4, 319–332, <https://doi.org/10.1038/s43017-023-00409-w>, 2023.
- Liu, Y. and Fu, W.: Assimilating high-resolution sea surface temperature data improves the ocean forecast potential in the Baltic Sea, *Ocean Science*, 14, 525–541, <https://doi.org/10.5194/os-14-525-2018>, 2018.
- Lou, R., Lv, Z., Dang, S., Su, T., and Li, X.: Application of machine learning in ocean data, *Multimedia Systems*, 29, 1815–1824, <https://doi.org/10.1007/s00530-020-00733-x>, 2023.
- 425 Madec, G., Bourdallé-Badie, R., Bouttier, P.-A., Bricaud, C., Bruciaferri, D., Calvert, D., Chanut, J., Clementi, E., Coward, A., and Delrosso, D.: NEMO ocean engine, [https://epic.awi.de/id/eprint/39698/1/NEMO\\_book\\_v6039.pdf](https://epic.awi.de/id/eprint/39698/1/NEMO_book_v6039.pdf), 2017.
- Maljutenko, I., Elken, J., Lagemaa, P., and Verjovkina, S.: Mere operatiivmodelisüsteemi NEMO kasutuselevõtt ja töölerakendamise mereala operatiivprognooside parandamiseks, Tech. rep., TTÜ Meresüsteemide Inst., <https://kliimaministerium.ee/sites/default/files/documents/2022-10/Mere%20operatiivmodelis%C3%BCsteemi%20NEMO%20kasutuselev%C3%B5tt%20ja%20t%C3%B6%C3%B6lerakendamise%20mereala%20operatiivprognooside%20parandamiseks.pdf>, 2022.
- 430 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, *Frontiers in Marine Science*, 10, <https://doi.org/10.3389/fmars.2023.1233679>, 2023.
- 435 Metzner, M., Gade, M., Hennings, I., and Rabinovich, A. B.: The observation of seiches in the Baltic Sea using a multi data set of water levels, *Journal of Marine Systems*, 24, 67–84, [https://doi.org/https://doi.org/10.1016/S0924-7963\(99\)00079-2](https://doi.org/https://doi.org/10.1016/S0924-7963(99)00079-2), 2000.
- Miles, E., Spillman, C., Church, J., and McIntosh, P.: Seasonal prediction of global sea level anomalies using an ocean–atmosphere dynamical model, *Clim. Dyn.*, 43, 2131–2145, <https://doi.org/10.1007/s00382-013-2039-7>, 2014.



- Mostafavi, M., Delpeche-Ellmann, N., Ellmann, A., and Jahanmard, V.: Determination of Accurate Dynamic Topography for the Baltic Sea  
440 Using Satellite Altimetry and a Marine Geoid Model, *Remote Sensing*, 15, <https://doi.org/10.3390/rs15082189>, 2023.
- Owens, R. and Hewson, T.: ECMWF Forecast User Guide, Tech. rep., Reading, <https://doi.org/10.21957/m1cs7h>, 2018.
- Ponte, R. M., Carson, M., Cirano, M., Domingues, C. M., Jevrejeva, S., Marcos, M., Mitchum, G., Van De Wal, R., Woodworth, P. L., Ablain,  
M., et al.: Towards comprehensive observing and modeling systems for monitoring and predicting regional to coastal sea level, *Frontiers  
in Marine Science*, 6, 437, <https://doi.org/10.3389/fmars.2019.00437>, 2019.
- 445 Pärt, S., Björkqvist, J.-V., Alari, V., Maljutenko, I., and Uiboupin, R.: An ocean–wave–trajectory forecasting system for the eastern  
Baltic Sea: Validation against drifting buoys and implementation for oil spill modeling, *Marine Pollution Bulletin*, 195, 115497,  
<https://doi.org/https://doi.org/10.1016/j.marpolbul.2023.115497>, 2023.
- Raissi, M., Perdikaris, P., and Karniadakis, G.: Physics-informed neural networks: A deep learning framework for solving for-  
ward and inverse problems involving nonlinear partial differential equations, *Journal of Computational Physics*, 378, 686–707,  
450 <https://doi.org/10.1016/j.jcp.2018.10.045>, 2019.
- Rajabi-Kiasari, S., Delpeche-Ellmann, N., and Ellmann, A.: Forecasting of absolute dynamic topography using deep learning algorithm with  
application to the Baltic Sea, *Computers & Geosciences*, 178, 105406, <https://doi.org/10.1016/j.cageo.2023.105406>, 2023.
- Rus, M., Fettich, A., Kristan, M., and Ličer, M.: HIDRA2: deep-learning ensemble sea level and storm tide forecasting in the presence of  
seiches – the case of the northern Adriatic, *Geoscientific Model Development*, 16, 271–288, <https://doi.org/10.5194/gmd-16-271-2023>,  
455 2023.
- Rutgersson, A., Kjellström, E., Haapala, J., Stendel, M., Danilovich, I., Drews, M., Jylhä, K., Kujala, P., Larsén, X. G., Halsnæs, K., Lehtonen,  
I., Luomaranta, A., Nilsson, E., Olsson, T., Särkkä, J., Tuomi, L., and Wasmund, N.: Natural hazards and extreme events in the Baltic Sea  
region, *Earth System Dynamics*, 13, 251–301, <https://doi.org/10.5194/esd-13-251-2022>, 2022.
- Samuelsson, M. and Stigebrandt, A.: Main characteristics of the long-term sea level variability in the Baltic sea, *Tellus A: Dynamic Meteo-  
460 rology and Oceanography*, 48, 672–683, <https://doi.org/10.1034/j.1600-0870.1996.t01-4-00006.x>, 1996.
- Scotto, M., Barbosa, S. M., and Alonso, A. M.: Model-based clustering of Baltic sea-level, *Applied Ocean Research*, 31, 4–11,  
<https://doi.org/10.1016/j.apor.2009.03.001>, 2009.
- Shahabi, A. and Tahvildari, N.: A deep-learning model for rapid spatiotemporal prediction of coastal water levels, *Coastal Engineering*, 190,  
104504, <https://doi.org/10.1016/j.coastaleng.2024.104504>, 2024.
- 465 Sonnewald, M., Lguensat, R., Jones, D. C., Dueben, P. D., Brajard, J., and Balaji, V.: Bridging observations, theory and numerical simulation  
of the ocean using machine learning, *Environmental Research Letters*, 16, 073008, <https://doi.org/10.1088/1748-9326/ac0eb0>, 2021.
- Suursaar, Ü., Kullas, T., and Otsmann, M.: A model study of the sea level variations in the Gulf of Riga and the Väinameri Sea, *Continental  
Shelf Research*, 22, 2001–2019, [https://doi.org/10.1016/S0278-4343\(02\)00046-8](https://doi.org/10.1016/S0278-4343(02)00046-8), 2002.
- Tanajura, C. A. S., Lima, L. N., and Belyaev, K. P.: Assimilation of satellite surface-height anomalies data into a Hybrid Coordinate Ocean  
470 Model (HYCOM) over the Atlantic Ocean, *Oceanology*, 55, 667–678, <https://doi.org/10.1134/S0001437015050161>, 2015.
- Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res. Atmos.*, 106, 7183–7192,  
<https://doi.org/10.1029/2000JD900719>, 2001.
- Weisse, R., Dailidienė, I., Hünicke, B., Kahma, K., Madsen, K., Omstedt, A., Parnell, K., Schöne, T., Soomere, T., Zhang, W., and Zorita,  
E.: Sea level dynamics and coastal erosion in the Baltic Sea region, *Earth System Dynamics*, 12, 871–898, [https://doi.org/10.5194/esd-12-  
871-2021](https://doi.org/10.5194/esd-12-<br/>475 871-2021), 2021.



- Wolski, T. and Wiśniewski, B.: Characteristics and long-term variability of occurrences of Storm surges in the Baltic Sea, *Atmosphere*, 12, 1679, <https://doi.org/10.3390/atmos12121679>, 2021.
- Wolski, T., Wiśniewski, B., Giza, A., Kowalewska-Kalkowska, H., Boman, H., Grabbi-Kaiv, S., Hammarklint, T., Holfort, J., and Žydrune Lydeikaitė: Extreme sea levels at selected stations on the Baltic Sea coast, *Oceanologia*, 56, 259–290, <https://doi.org/10.5697/oc.56-2.259>, 480 2014.
- Zhang, Y., Liu, J., and Shen, W.: A review of ensemble learning algorithms used in remote sensing applications, *Appl. Sci.*, 12, 8654, <https://doi.org/10.3390/app12178654>, 2022.
- Zhao, Q., Peng, S., Wang, J., Li, S., Hou, Z., and Zhong, G.: Applications of deep learning in physical oceanography: a comprehensive review, *Frontiers in Marine Science*, 11, <https://doi.org/10.3389/fmars.2024.1396322>, 2024.