https://doi.org/10.5194/egusphere-2026-3411
Preprint. Discussion started: 25 June 2026

Deep-learning prediction of high-frequency sea-level oscillations in the Adriatic Sea

Iva Međugorac[1, 2], Nikola Metličić[3], Jadranka Šepić[3], Marko Rus[4], Srđan Čupić[5], Matej Kristan[4], Matjaž Ličer[2, 6, 7]

[1]University of Nova Gorica, Vipavska 13, Rožna dolina, SI-5000 Nova Gorica, Slovenia
[2]Slovenian Environment Agency, Vojkova 1b, 1000 Ljubljana, Slovenia
[3]Faculty of Science, University of Split, Ruđera Bošković 33, 21000 Split, Croatia
[4]Faculty of Computer and Information Science, University of Ljubljana, Večna pot 113, 1000 Ljubljana, Slovenia
[5]Hydrographic Institute of the Republic of Croatia, Zrinsko-Frankopanska 161, 21000 Split , Croatia
[6]National Institute of Biology, Ljubljana, Večna pot 121, 1000 Ljubljana , Slovenia
[7]Faculty of Mathematics and Physics, University of Ljubljana, Jadranska ulica 19, 1000 Ljubljana, Slovenia
Correspondence to: Iva Međugorac (iva.medugorac@gfz.hr)

Abstract. The eastern Adriatic coast is a known hotspot of strong meteorologically induced high-frequency sea-level oscillations, occurring at periods shorter than 1 hour and reaching wave heights of several metres. When highest, these oscillations are termed meteotsunamis. In this study, we test deep-learning methods for predicting maximum daily amplitudes of high-frequency ($T < 1$ hour) sea-level oscillations at two Adriatic locations, Bakar and Ploče, using convolutional neural networks driven by past sea-level observations and atmospheric predictors from the ERA5 and CERRA reanalyses. We evaluate two deep-learning architectures designed to test different approaches to representing sea-level and atmospheric forcing. The first architecture, HFNet, is based on the HIDRA family of models, a general low-frequency sea-level forecasting framework that has been extensively evaluated in the Adriatic and shown to provide a credible baseline for sea-level prediction. The second architecture, $\mathrm{HFNet_{JE}}$, extends this approach through joint encoding of atmospheric predictors and a more extensive processing of past sea-level information, with the aim of improving the representation of processes associated with high-frequency sea-level oscillations. Analysis of more than 20 years of data shows that high-frequency sea-level extremes are larger in Bakar (> 60 cm) than in Ploče (< 35 cm), occur ~6 times per year, and are most common during the warm season. Both architectures reproduce the observed variability, with higher skill for typical than for extreme events. $\mathrm{HFNet_{JE}}$ performs best overall and under typical amplitude conditions, whereas HFNet more effectively captures extreme events, although these remain systematically underestimated in both architectures. Model performance is higher at Ploče, likely because of its smaller sea-level range and simpler response to atmospheric forcing. Models forced with ERA5 consistently outperform those using the higher-resolution CERRA in predicting extremes, suggesting limited added value from increased spatial resolution. Ablation experiments indicate that several predictors are redundant for average forecasting performance, whereas extreme-event prediction generally benefits from the full predictor set. Overall, the results demonstrate the potential of deep learning for prediction of high-frequency sea-level oscillations in the Adriatic, but also highlight persistent limitations in forecasting rare high-amplitude events.

1 Introduction

Coastal floods are among the major hazards in regions worldwide, affecting an ever-growing population, infrastructure, and economies (Neumann et al., 2015). In the Mediterranean, a climate-change hot spot (Diffenbaugh and Giorgi, 2012), positive sea-level (SL) extremes occur predominantly at two temporal scales: low frequencies (LF), corresponding to storm surges (periods from several hours to several days), and high frequencies (HF), corresponding to tsunamis and meteotsunamis (periods from several minutes to several tens of minutes) (Vilibić et al., 2025). Processes that contribute to high-frequency sea-level oscillations (HFOs) include infragravity waves (e.g., Bertin et al., 2018), free and forced long ocean waves (Pugh and Woodworth, 2014), edge waves (Ursell, 1952), and seiches (Rabinovich, 2009). Intense HFOs can be generated by seismic, landslide, or volcanic tsunamis, but also by atmospheric forcing. Although the Mediterranean is highly exposed to seismic tsunamis (Euro-Mediterranean Tsunami Catalogue, 2026), atmospherically induced HFOs are far more frequent and can reach wave heights of several metres, causing strong currents, coastal flooding, and damage to harbour infrastructure (Šepić and Orlić, 2026; Vilibić et al., 2021; Monserrat et al., 2006). The most extreme of these events, known as meteorological tsunamis (meteotsunamis), occur at several locations in the Mediterranean, including sites in the Adriatic Sea (Fig. 1).

Strong atmospherically induced HFOs in the Adriatic can be broadly classified into two categories according to the dominant atmospheric structures observed during the events: small-scale and synoptic-scale systems (e.g., Šepić and Orlić, 2026; Ruić et al., 2025; Horvath et al., 2018; Šepić et al., 2016b). While the latter are often present during strong HFOs, their direct role in generating the oscillations remains uncertain, as smaller-scale atmospheric forcing may be embedded within the larger-scale flow (Ruić, 2025). Small-scale systems responsible for HFOs include spatially limited (and often sub-grid) atmospheric disturbances, such as convective systems, gravity waves, or fronts, which propagate over the open sea and induce long ocean waves through rapid pressure changes. These long ocean waves are amplified through a sequence of resonance mechanisms as they propagate toward the coast. When the speed of atmospheric disturbance matches the shallow-water long ocean wave speed, amplification occurs via Proudman resonance (Proudman, 1929), often followed by further amplification through shoaling and harbour resonance, potentially leading to destructive sea-level oscillations (Rabinovich, 2009; Monserrat et al., 2006). Since resonance depends on specific combinations of atmospheric disturbance properties and of local bathymetry, meteotsunamis are restricted to particular locations in the Adriatic Sea (Fig. 1). Despite their generation by small-scale atmospheric disturbances, this type of intense HFOs is associated with distinct synoptic-scale conditions that can be identified in the larger-scale flow (Jansà and Ramis, 2021; Šepić et al., 2016). The characteristic synoptic conditions typically involve a layered tropospheric structure with a shallow moist surface layer, a temperature inversion, and a deeper unstable layer, together with large-scale features such as low-pressure troughs and strong southwesterly flow aloft.

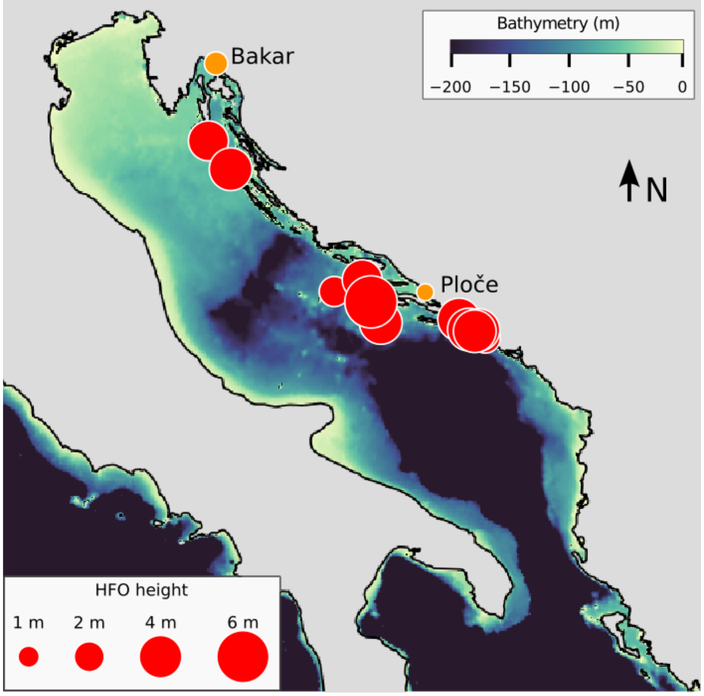

Figure 1. Bathymetry of the Adriatic Sea (depths greater than 200 m are coloured with the same colour). The displayed domain corresponds to the area over which atmospheric data used for model development were considered. Orange circles mark the tide-gauge stations analysed in this study, and red circles mark meteotsunami-prone locations. HFO heights at Bakar and Ploče were computed from available measurements (Fig. 2), while for the other locations heights are based on eyewitness reports and field studies (Šepić and Orlić, 2026). Except for Bakar and Ploče, only locations at which observed meteotsunamis had heights ≥ 2 m are shown in the plot.

In addition to small-scale patterns, pronounced HFOs in the Adriatic are often observed in the presence of synoptic-scale systems associated with Sirocco (warm and moist southeasterly flow) or Bora (cold and dry northeasterly flow) winds (Ruić et al., 2025).

Despite the relatively well-understood physical mechanisms (Vilibić et al., 2025; 2021), forecasting HFOs remains challenging. Early forecasting approaches relied on subjective recognition of synoptic atmospheric patterns (Jansà and Ramis, 2021) or empirical indices (Šepić et al., 2016a) – both of these processes can identify favourable conditions but have limited skill in predicting amplitudes. More advanced approaches use deterministic atmosphere-ocean modelling systems, such as the operational BRIFS system for the Balearic Islands (Renault et al., 2011) and the AdriSC system for the Adriatic Sea. The latter has been developed and tested but is currently not operational (Denamiel et al., 2019a). Although these systems can reproduce many observed events and provide useful warnings, they are computationally demanding and tend to underestimate extreme amplitudes. This limitation is largely due to difficulties in accurately representing the intensity and propagation of small-scale atmospheric disturbances, where even small errors can lead to large uncertainties in the SL response (Mourre et al., 2021; Šepić et al., 2016b; Orlić et al., 2010). This problem was addressed by Denamiel et al. (2019b), who developed a stochastic surrogate model for meteotsunami early warning in the eastern Adriatic. The model accounts for uncertainties in atmospheric forcing by using statistical representations of high-frequency disturbances to estimate event likelihood and magnitude. Although it successfully reproduced most observed events, with some false alarms, the approach has important limitations: it relies on synthetic atmospheric forcing, requires a large number of simulations, and remains computationally demanding, limiting its applicability in real-time forecasting.

In parallel, data-driven approaches have been increasingly applied to SL forecasting. Particularly relevant is the HIDRA family of models (Rus et al., 2026; 2025; 2023; Žust et al., 2021), a general deep-learning (DL) forecasting architecture thoroughly evaluated in the Adriatic Sea. The architecture has since been implemented along the Estonian coast (Barzandeh et al., 2025) and tested along the Spanish, Danish and German coasts (not documented). HIDRA uses a convolutional neural network (CNN) to predict hourly sea levels (including storm surges, tides, and basin-wide seiches) with lead times of up to 72 hours. Depending on the model version, the input data consist of past total or residual SL data from TG stations, together with past and forecast tides and selected meteorological and oceanographic variables, such as 10-m wind, mean sea-level pressure, sea-surface temperature, 2-m air temperature, and wave conditions. The first three HIDRA model versions provided point predictions at tide-gauge (TG) stations, while the recent HIDRA-D version (Rus et al., 2026) delivers 2D gridded SL forecasts for the entire Adriatic Sea. HIDRA3 (Rus et al., 2025) is currently operational for the northern Adriatic TG station Koper (https://vicoslab.github.io/hidra-visualization/en/) where it has been shown to outperform deterministic forecasting systems during extreme flooding events.

Despite these advances, the applicability of deep-learning approaches developed for LF sea-level variability to the HF domain remains unexplored. In contrast to LF variability (dominated by tides, storm surges, and basin-scale seiches), strong HFOs are often induced by rapidly evolving and spatially-localised atmospheric disturbances. Furthermore, strong HFOs are rare and strongly underrepresented in observational datasets. It is therefore unclear whether forecasting architectures successful for LF variability can effectively represent the processes governing HFO generation, or whether alternative strategies are required. A further limitation for developing HFO forecasting systems is the availability of long HF sea-level records. In the Mediterranean, permanent minute-resolution measurements have only recently been established at many sites (Pérez Gómez et al., 2022). However, some Adriatic locations, such as the TG stations Bakar and Ploče (Figs. 1 and 2), provide multi-decadal 1-minute records that enable the development of data-driven models. Although these sites are not prone to the most extreme HFOs (i.e., meteotsunamis), they record moderate amplitudes (up to 62 cm at Bakar and 33 cm at Ploče; Fig. 2), and their long time series provide a valuable basis for developing forecasting approaches that can later be transferred to sites with shorter records and more extreme HFOs.

The aim of this study is therefore to evaluate deep-learning approaches for predicting daily HFO amplitudes at Bakar and Ploče using past sea-level observations together with atmospheric predictors derived from the ERA5 and CERRA reanalyses. We test two CNN-based architectures that differ in their treatment of atmospheric predictors and past sea-level information.

The first architecture, HFNet, is based on the HIDRA family of models and serves as a transfer baseline for assessing whether a LF sea-level forecasting framework can be successfully adapted to HFO prediction. The second architecture, HFNet$_{\mathrm{JE}}$, extends this approach through joint encoding of atmospheric predictors and a more extensive representation of past SL variability, with the aim of improving the extraction of HFO-relevant structures. Our objective is therefore not only to assess forecasting skill, but also to examine how different architectural assumptions perform across typical and extreme HFO conditions.

The paper is structured as follows. Following the introduction (Section 1), Section 2 presents the meteorological fields and SL time series used to develop and train the proposed models. Section 3 describes the applied methods and model architectures. Section 4 presents the characteristics of HFOs at the two stations and the results of the prediction experiments. The results are discussed in Section 5, and conclusions are drawn in Section 6.

2 Meteorological and sea-level data

All datasets used in this study, including SL observations and meteorological reanalysis fields, are summarised in Table 1.

2.1 Meteorological data

Atmospheric 2D and 3D fields from the ERA5 global reanalysis (Hersbach et al., 2023a; 2023b; 2020), and the CERRA regional reanalysis (Schimanke et al., 2021a; 2021b), together with an ensemble of forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) (Leutbecher and Palmer, 2007), were used over the spatial domain shown in Fig. 1 (12–20.875°E, 39–45.875°N). Details are given in Table 1.

Table 1. Measured and simulated datasets and their properties used in this study. Time interval denotes the period for which data were available at the time the experiments were conducted.

Dataset	Temporal resolution (min)	Spatial resolution (km)	Time interval	Data reference
Bakar sea level	1		27/6/2003–19/11/2025	GFZ, https://www.ioc-sealevelmonitoring.org/station.php?code=baka
Ploče sea level	1		01/01/2003–30/12/2023	HHI, https://adriaticsea.hhi.hr/public-stations/station?id=1&name=MP+Plo%C4%8De
ERA5 reanalysis	60	~27	01/01/2003–30/11/2025	Climate Data Store (CDS), https://doi.org/10.24381/cds.adbb2d47 https://doi.org/10.24381/cds.bd0915c6

https://doi.org/10.5194/egusphere-2026-3411
Preprint. Discussion started: 25 June 2026

CERRA reanalysis	180	5.5	01/01/2003–30/09/2025	Climate Data Store, https://doi.org/10.24381/cds.622a565a https://doi.org/10.24381/cds.a39ff99f
ECMWF 3-day forecasts (50 member ensemble)	60	~27	01/08–31/08/2022, 01/07–31/08/2023, 01/10–30/11/2023	ECMWF, https://www.ecmwf.int/en/forecasts/accessing-forecasts

2.2 Sea-level time series

Sea-level records from two Adriatic tide gauges with the longest available 1-minute resolution data were used: Bakar, operated by the Department of Geophysics, Faculty of Science, University of Zagreb (GFZ), and Ploče, operated by the Hydrographic Institute of the Republic of Croatia (HHI) (Table 1, Figs. 1 and 2). Both are typical coastal stations equipped with float-type instruments installed in stilling wells. Digital instruments providing 1-minute SL measurements were installed at Bakar in 2003 (Međugorac et al., 2022) and at Ploče in 2002 (Pérez Gómez et al., 2022).

SL data were referenced to the mean over the analysed intervals, and all times are given in UTC. Raw 1-minute data were quality-controlled by identifying outliers using a Tukey 53H filter (Otnes and Enochson, 1978), followed by visual inspection to remove outliers and errors missed by the filter.

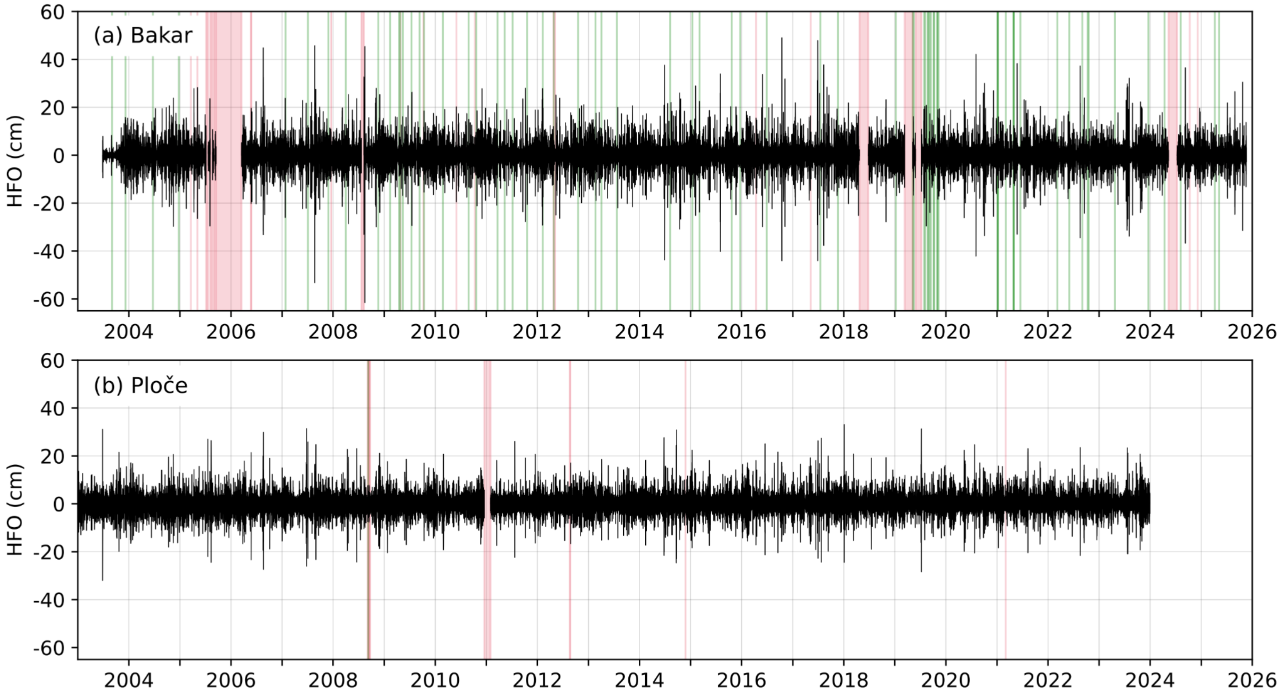

Figure 2. Available HFOs (cut off at T = 1 hour) at TG stations Bakar (**a**) and Ploče (**b**). The locations of the stations are shown in Fig. 1. Green shadings mark data gaps shorter than 5 hours, which were linearly interpolated for training purposes, while red shadings mark longer data gaps, which were not treated.

3 Methods

3.1 Sea-level and meteorological data analysis

Data gaps, more frequent at Bakar than at Ploče (Fig. 2), were treated as follows. Missing segments shorter than 5 hours were linearly interpolated (green shading in Fig. 2), while longer gaps were left unfilled (red shading in Fig. 2). This procedure was used to define 'valid days' for model training, validation, and testing, that is, time intervals without data gaps in either the input period (past 24 hours) or the target period (future 72 hours).

HFOs were extracted from the sea-level measurements by applying a digital filter with a Kaiser window and a half-power point at 1 hour.

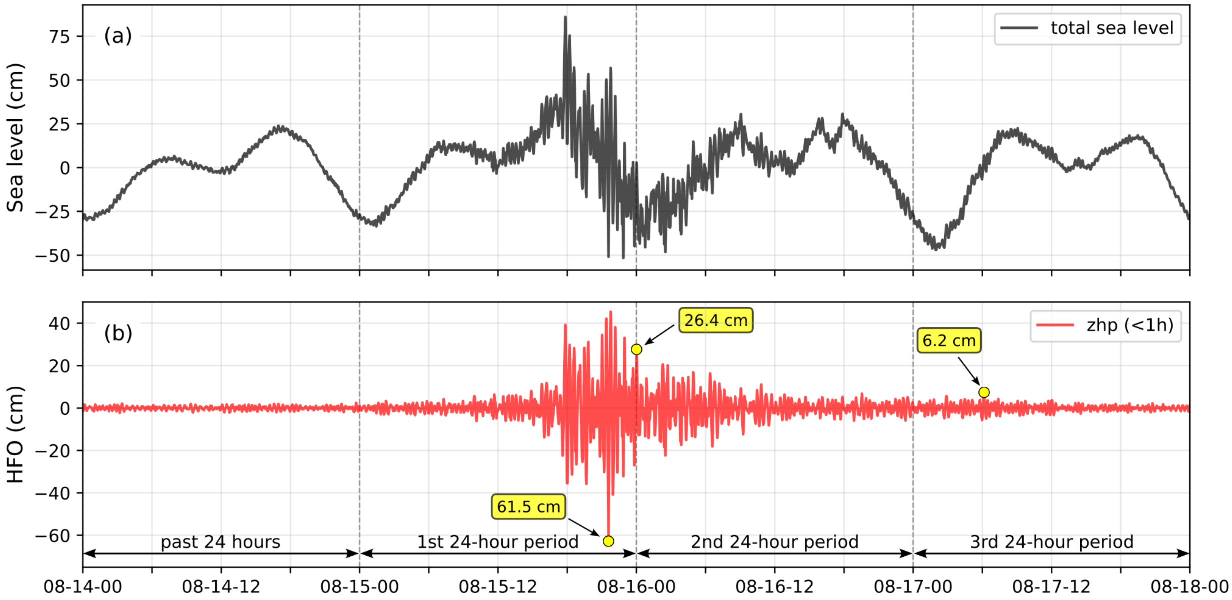

Figure 3. Four-day interval (14–18 August 2008) of (**a**) measured SL in Bakar and (**b**) the HFO series (high-pass filtered SL with a 1 hour cutoff period). This example illustrates how a single training sample is constructed: the preceding 24 hours of HFOs are used as input (together with their PSD), and the targets are defined as the maximum absolute HFO values within each of the three subsequent 24-hour periods (i.e., over the next 72 hours). In this example, the second target occurs very close to midnight.

To quantify HFO intensity within a given time interval (here, daily), we define the maximum HFO amplitude (HFO-A) as the maximum absolute value of the HFO signal within each daily interval. Our experiments, which tested several definitions of HFO strength, indicate that the selected measure provides the most robust representation of HFO intensity at the studied locations, particularly during strong events. In such cases, the sequence of extrema is not consistent: either a minimum or a maximum may occur first, and the initial extremum does not necessarily correspond to the largest amplitude (an example in Fig. 3). Consequently, measures based solely on maxima or solely on minima may underestimate the intensity of the strongest oscillation, whereas the maximum absolute HFO captures the largest amplitude regardless of its sign. This issue is further complicated by the presence of multiple oscillation modes during strong events. For example, in Bakar, in addition to the dominant mode, higher harmonics are also excited (see Section 4.1), typically with smaller amplitudes. As a result, alternative measures such as wave height defined by Šepić et al. (2016a) (i.e., the difference between the maximum and minimum HF sea level within an interval bounded by two successive falling inflection points) may also underestimate the strongest oscillation in an HFO episode.

Extreme events were defined as those when maximum daily HFO-A equaled or exceeded 20 cm in Bakar and 15 cm in Ploče, corresponding approximately to the 99.98th percentile of the available sea-level series.

To analyse the spectral characteristics of HFOs and support model development, smoothed power spectral densities (PSDs) were computed using 3-hour Hamming windows with 50% overlap.

The following meteorological fields were used: mean sea-level pressure (MSLP), air temperature (T), relative humidity (r), wind components (u, v), geopotential height (z), and the dry-atmosphere Richardson number (Ri), computed following Lindzen and Tung (1976).

The meteorological background of extreme HFO events was classified using a subjective approach based on visual inspection of ERA5 surface atmospheric fields (MSLP and (u, v) at 10 m). The classification follows three characteristic atmospheric conditions known to be associated with intense HFOs in Bakar and Ploče (Ruić et al., 2025): (i) Sirocco, (ii) Bora, and (iii) Calm conditions (i.e., summer-type meteotsunamigenic conditions). Bora conditions were characterized by the presence of a discernible MSLP gradient over the basin and northeasterly (Bora) winds exceeding 7 m/s over the entire or northern Adriatic. Sirocco conditions were characterized by the presence of a relatively strong MSLP gradient and southeasterly (Sirocco) to southwesterly winds exceeding 7 m/s over the entire Adriatic or a substantial part of it. Calm conditions were characterized by the absence of a pronounced MSLP gradient and calm or weak winds (<7 m/s).

It should be noted that Sirocco and Bora are basin- to sub-basin-scale phenomena whose properties are expected to be reasonably well represented in both ERA5 and CERRA. In contrast, Calm situations are only partially resolved. On the synoptic scale, they are typically associated with a weak pressure gradient (non-gradient MSLP field), a temperature front aloft, and enhanced southwesterly flow in the upper troposphere. However, the small-scale atmospheric features favoured by these conditions, responsible for triggering Proudman and related resonances, can have horizontal scales as small as ~25 km (Šepić et al., 2016b) and are therefore not resolved in ERA5 (∼27 km resolution in the study area) and are likely also inadequately represented in CERRA (Vilibić et al., 2025 and references therein).

3.2 Experimental strategy

We compare two CNN-based DL architectures for forecasting HFOs that differ in their treatment of atmospheric forcing and past SL information. The first architecture, HFNet (Fig. 4), is adapted from the HIDRA framework, a well-established DL architecture for LF sea-level forecasting (Rus et al., 2026; 2025; 2023; Žust et al., 2021), and serves as a transfer baseline for assessing whether a successful LF forecasting design can be extended to HFO prediction. Relative to HIDRA, HFNet extends the atmospheric input from surface 2D fields to a combination of 2D and 3D atmospheric fields, while replacing SL time-series prediction with prediction of daily HFO amplitudes. Atmospheric variables are processed largely separately before their representations are combined. The second architecture, $\mathrm{HFNet_{JE}}$ (Fig. S1 in the Supplement), modifies this approach through joint encoding of atmospheric predictors and a more extensive processing of past SL information (HFOs and PSD). Rather than presenting $\mathrm{HFNet_{JE}}$ as a stepwise extension of HFNet, we directly compare the two architectures as alternative design strategies for HFO prediction across different amplitude regimes.

Both models were applied to two TG stations (Bakar and Ploče) using two meteorological datasets (ERA5 and CERRA). Model performance was additionally evaluated using an ensemble of ECMWF forecasts consisting of 50 members generated from perturbed initial conditions (Leutbecher and Palmer, 2007). Due to memory constraints, hourly 3-day forecasts were analysed only for selected extreme events in 2022 and 2023 (Table 1). The models were trained exclusively on ERA5 and CERRA reanalysis data and not on operational ECMWF IFS forecasts.

3.3 Architecture of HFNet and $\mathrm{HFNet_{JE}}$

The HFNet architecture is shown in Fig. 4. Input variables are processed in two stages: first encoded separately and then fused and processed jointly. The model operates on 2D and 3D atmospheric fields, HFO time series, and their PSD. Although SL responds primarily to surface forcing (air pressure and wind), higher atmospheric levels may contain information relevant for HFO generation. As discussed in Section 1, HFOs with Calm background are often associated with distinct upper-atmospheric structures, whereas the small atmospheric systems responsible for their generation are inadequately resolved in the reanalysis products used here. Higher-level fields may therefore help identify larger-scale conditions favourable for HFO development. To capture these signals, the model incorporates both 2D (MSLP) and 3D (T, r, (u, v), z, and Ri) fields. The 3D fields are included at pressure levels of 500, 700, 850, and 1000 hPa. These specific variables at these (and sometimes additional) heights are typically used in analysis of synoptic conditions relevant for meteotsunami appearance (e.g., Šepić et al., 2016a). Past and future atmospheric inputs are encoded by an atmospheric encoder, while past HFOs, and their PSD (computed as described in Section 3.1) are encoded separately using dense layers. The resulting feature vectors are concatenated and passed to a regression block that predicts the maximum daily HFO-A for three forecast days. Detailed description of HFNet can be found in Appendix A.

The $\mathrm{HFNet_{JE}}$ architecture (Fig. S1 in the Supplement; detailed description in the Supplement) uses the same input variables as HFNet but processes them differently, particularly the atmospheric fields. In HFNet atmospheric variables are processed independently in the spatio-temporal stage before being combined later in the network. In contrast, $\mathrm{HFNet_{JE}}$ introduces an atmospheric encoder in which all atmospheric variables are processed jointly through both spatial and temporal modules, allowing the model to learn interactions between variables and vertical levels directly from the data. In addition, past SL information is processed through a deeper network comprising multiple convolutional, recurrent, and dense layers, rather than a single fully connected layer as in HFNet. This modification was motivated by the difference between the original HIDRA task, where past SL mainly provides information on the current LF sea-level state, and the present HFO task, where recent HF variability and spectral characteristics may indicate the presence of already excited modes that can be further amplified under favourable atmospheric forcing. Therefore, $\mathrm{HFNet}_{\mathrm{JE}}$ should be interpreted as an alternative architecture family, rather than as a single-change modification of HFNet; performance differences cannot be attributed solely to the atmospheric encoder. The regression block remains unchanged. The Richardson number was omitted as an explicit input from $\mathrm{HFNet}_{\mathrm{JE}}$ because the atmospheric encoder receives all variables simultaneously and is expected to learn the relevant dependencies internally.

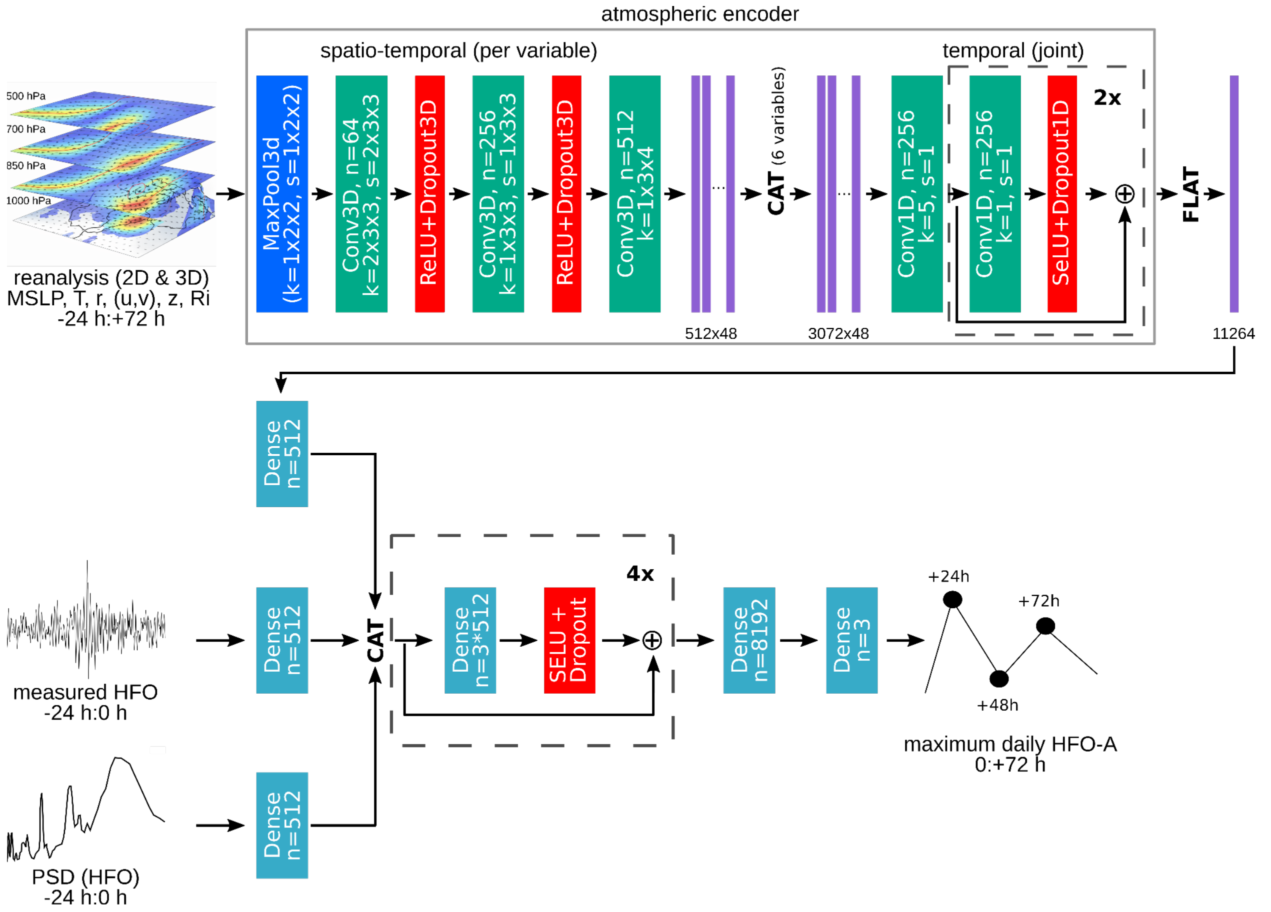

Figure 4. Architecture of the HFNet model. The atmospheric encoder consists of a spatio-temporal part, applied separately to each atmospheric variable, and a temporal part, in which the resulting features are concatenated and processed jointly. Colours denote different layer types; *n* indicates the number of output channels, *k* the kernel size, and *s* the stride.

3.4 Implementation details

Training, validation, and testing subsets were constructed to capture interannual variability in extreme events and to account for differing meteorological mechanisms classified into three categories: Bora, Sirocco, and Calm situations (see Section 3.1 and Fig. 5). Each subset therefore included large HFO events and, where possible, cases associated with three distinct meteorological backgrounds. Years 2015, 2020, and 2021 were used for validation, while 2022–2023 were reserved for testing; the remaining years were used for training, resulting in 18 training years for Bakar and 16 for Ploče. Accordingly, the training, validation, and test sets contained 5782/5766, 1097/1092, and 730/727 samples for Bakar/Ploče, respectively. Although the Bakar record spans two additional years (Fig. 2), numerous data gaps reduce the number of 'valid days', resulting in a similar number of training samples at the two stations.

Because extreme events (HFO-A ≥ 99.98th percentile) represent 0.02% of highest high-frequency sea-level oscillations (Fig. 5), training employed oversampling of extreme cases. Each batch contained 90% samples from the full training dataset and 10% from the extreme-event subset.

All input variables were normalised independently by subtracting the mean and dividing by the standard deviation. For each input variable, the mean and standard deviation were calculated over the entire training set across all spatial and temporal positions (horizontal, vertical, and time dimensions).

The dimensions of the meteorological input fields (time × level × latitude × longitude) were 96 × 1 × 56 × 72 for MSLP and 96 × 4 × 56 × 72 for all other variables in ERA5, and 32 × 1 × 153 × 148 for MSLP and 32 × 4 × 153 × 148 for all other variables in CERRA. The target variable was the maximum daily HFO-A (an example in Fig. 3).

All networks were trained using AdamW optimizer (Loshchilov and Hutter, 2017) with learning rate of 1e-5; no learning rate scheduler was used. Models were trained for 80 epochs. Additional training (e.g., up to 300 epochs) yielded no further reduction in loss or improvement in performance.

All results are based on models selected using the best validation loss criterion. Specifically, the selected epoch corresponds to the smallest maximum validation loss among all validation samples, emphasising the model's ability to capture extreme events rather than average conditions.

Batch sizes were 128 for HFNet and 32 for $\mathrm{HFNet_{JE}}$. While 128 was optimal for HFNet in terms of stability and performance, the $\mathrm{HFNet_{JE}}$ batch size was constrained by hardware limitations, despite potential benefits from larger batches. Each experiment was repeated three times to account for variability due to random initialisations of model parameters.

Training times were 13 hours (ERA5) and 19 hours (CERRA) for HFNet, and 23 hours (ERA5) and 37 hours (CERRA) for $\mathrm{HFNet_{JE}}$. Experiments were performed on NVIDIA A100-SXM4-40GB GPUs using PyTorch. Because of the large $\mathrm{HFNet_{JE}}$ input size, PyTorch's Distributed Data Parallel (DDP) framework was used with a per-GPU batch size of 8 samples to increase the number of batches and stabilize loss estimates.

4 Results

4.1 HFO characteristics and meteorological backgrounds

The SL record spans 8181 days in Bakar and 7668 days in Ploče. However, Bakar contains substantially more data gaps than Ploče (green and red shading in Fig. 2), resulting in a lower fraction of 'valid days' available for training (93.11%, corresponding to 7618 'valid days') compared with Ploče (98.9%, corresponding to 7584 'valid days'). Consequently, the two stations yielded a very similar number of samples in the training, validation, and test datasets (Section 3.4).

PSDs (not shown) indicate that, at periods shorter than 1 hour, three dominant spectral peaks appear in Bakar (20.0 min, 7.8 min, 4.3 min), whereas two periods dominate in Ploče (30.0 minute and 8.6 minute). Note that the spectra were smoothed (3-hour Hamming windows, then averaged), therefore, smaller peaks already reported in the literature (for Bakar cf. Goldberg and Kempni, 1938; their Table 9) are not present in spectra.

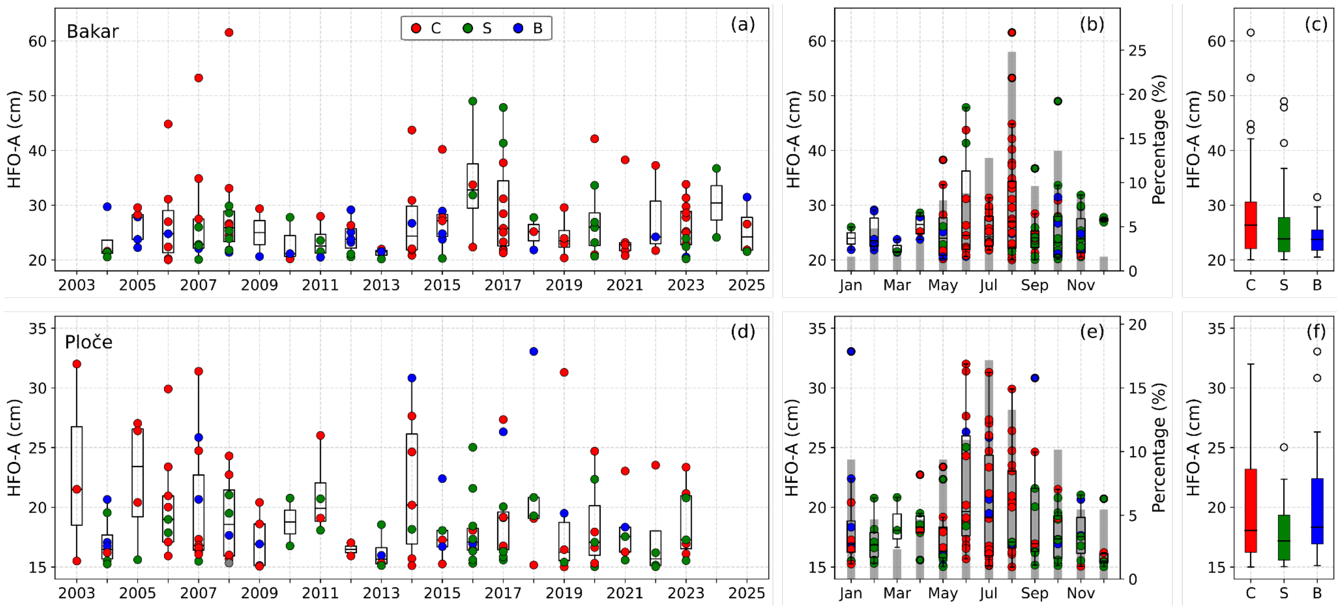

Figure 5. Distribution of extreme daily HFO-As in Bakar (≥ 20 cm): (**a**) over the study period, (**b**) by season, and (**c**) by weather type. Distribution of extreme daily HFO-As in Ploče (≥ 15 cm): (**d**) over the study period, (**e**) by season, and (**f**) by weather type. Boxplots show the median, the 25th and 75th percentiles, whiskers extending to the farthest data within 1.5x the interquartile range, and outliers beyond the whiskers. Coloured dots indicate different types of meteorological background: C - Calm situation, S - Sirocco wind, B - Bora wind.

Extreme daily HFO-As are larger in Bakar (> 60 cm) than in Ploče (< 35 cm) (Fig. 5) and occur, on average, about six times per year. The annual distribution of extremes is highly uneven at both stations: some years contain no or only minor events, while others exhibit events with larger amplitudes. Seasonally, intense HFOs occur throughout the year but are most frequent during the warm season – peaking in August at Bakar and during June–September at Ploče. Amplitudes tend to be larger in the warm season as well, although large HFOs also occur outside this period (e.g., notice recorded extremes in May and October at Bakar and in January at Ploče).

Focusing on event types, identified based on a subjective assessment as described in Section 3.1, the largest events in Bakar are most often associated with Calm conditions or Sirocco, whereas Bora typically generates lower HFO-A. The three event types also produce distinct SL responses in terms of frequency content (not shown). During Calm and Sirocco events, the HFO signal at Bakar is dominated by a ~20-minute period, indicating that this mode is predominantly excited, with no significant contribution from shorter periods. In contrast, Bora additionally excites shorter periods (7.8 min, 4.3 min). Notably, both Sirocco and Bora were relatively weak during the extracted extreme events compared to typical episodes of these wind regimes, as maximum Adriatic wind speeds in the corresponding composite (event-averaged) wind fields did not exceed ~12 m/s and ~15 m/s, respectively. In Ploče, the largest events are primarily associated with Calm conditions or Bora. In contrast to Bakar, all three event types produce very similar SL responses in terms of excited frequencies, with a dominant period of ~30 min. As in Bakar, the extracted Sirocco and Bora events were weak relative to typical episodes of these winds, with maximum Adriatic wind speeds in the composite fields not exceeding ~10 m/s. During Bora events in Ploče, the wind field is characterized by northeasterly flow over the northern Adriatic and northwesterly flow over the middle and southern Adriatic. This differs from Bora events in Bakar, where Bora is either confined to the northern Adriatic or extends across the entire basin, while the middle and southern Adriatic remain dominated by southeasterly winds.

Around 20% of all extreme events (Fig. 5) occur simultaneously at both stations, i.e. on the same day (25 events). These synchronized extremes occur mostly in summer and autumn and are primarily associated with Calm conditions (16 events) and Bora (8 events), while only one event is related to Sirocco.

As shown in Fig. 5, although more than 20 years of data are available at both stations, extreme events are rare and relatively weak (the largest amplitude in Bakar is ~60 cm). Consequently, the models have only a limited number of examples from which to learn. In fact the experiments without oversampling (setup HFNet-ERA5, Tables 2 and 4) showed lower model errors for regular events at both stations: by approximately 12.5% in both RMSE and MAE at Bakar, and by 16% in RMSE and 17.7% in MAE at Ploče. However, errors for extreme events increased. At Bakar, RMSE increased by 15% and MAE by 16.8%, while at Ploče the increase was substantially larger, reaching 39% in RMSE and 51% in MAE. This justifies the use of extreme-event oversampling applied in this study. The small number of extremes also motivates the use of an amplitude-dependent weighted loss function, in which larger HFOs are assigned greater weights to enhance sensitivity to extreme events during training, potentially improving model performance (see Section 5).

4.2 Prediction of HFOs

Results of predicted HFOs are presented for configurations (targets and lead times) that yielded the most skillful forecasts, although a broader range of configurations was tested and is discussed in Section 5. Results for the Bakar station are summarized in Tables 2 and 3 and Fig. 6, while results for the Ploče station are shown in Tables 4 and 5 and Fig. 7.

Table 2. Mean performance metrics (RMSE, MAE, relative error, bias, and accuracy) for the 24-hour lead-time forecasts using ERA5 input at the Bakar station, averaged over three runs. Metrics are reported for all HFO-As (All), regular events (HFO-A < 20 cm; Reg), and extreme events (HFO-A ≥ 20 cm; Ext). Accuracy is defined as the proportion of predictions differing from observations by less than one standard deviation. Underlined values denote optimal values within each category (All, Reg, and Ext).

Model	Amplitude type	Events	RMSE (cm)	MAE (cm)	Relative error (%)	Bias (cm)	Accuracy (%)
HFNet	All	730	3.15	2.22	41.87	0.90	89.95
HFNet$_{JE}$			2.57	1.54	24.14	-0.35	93.97
HFNet	Reg	715	2.88	2.11	42.19	1.04	89.93
HFNet$_{JE}$			2.07	1.39	23.96	-0.22	95.37
HFNet	Ext	15	9.12	7.46	26.53	-5.72	42.22
HFNet$_{JE}$			10.87	8.95	32.64	-6.37	28.89

At the Bakar station, both CNN architectures reproduce the observed variability of daily HFO-A (Fig. 6). Most predictions cluster close to the one-to-one line, indicating that the models capture the overall relationship between atmospheric forcing and HFO-A. Prediction skill for both models using both reanalysis datasets (Tables 2 and 3) is highest for the more frequent regular events (Reg), whereas larger events (Ext) show greater scatter (Fig. 6) and are mostly underestimated.

Table 3. Mean performance metrics (RMSE, MAE, relative error, bias, and accuracy) for the 24-hour lead-time forecasts using CERRA input at the Bakar station, averaged over three runs. Metrics are reported for all HFO-As (All), regular events (HFO-A < 20 cm; Reg), and extreme events (HFO-A ≥ 20 cm; Ext). Accuracy is defined as the proportion of predictions differing from observations by less than one standard deviation. Underlined values denote optimal values within each category (All, Reg, and Ext).

Model	Amplitude type	Events	RMSE (cm)	MAE (cm)	Relative error (%)	Bias (cm)	Accuracy (%)
HFNet	All	730	2.97	2.02	36.70	0.24	91.14
HFNet$_{JE}$			2.75	1.60	24.10	-0.56	93.29
HFNet	Reg	715	2.50	1.85	36.72	0.43	92.73
HFNet$_{JE}$			2.18	1.43	23.84	-0.43	94.78
HFNet	Ext	15	11.47	10.02	36.04	-8.82	15.56
HFNet$_{JE}$			11.91	9.78	36.60	-6.68	22.22

Comparison of results obtained using ERA5 and CERRA input data, independent of model architecture, shows relatively small differences for all and regular events, with slightly better scores obtained when CERRA input is used. In contrast, differences are larger for extreme events, for which better scores are obtained with ERA5 input across all metrics. This suggests that CERRA, with its higher spatial resolution (5.5 km versus 27 km), performs somewhat better for regular events, whereas ERA5, with its higher temporal but coarser spatial resolution (1 hour versus 3 hours), performs better for forecasting extremes.

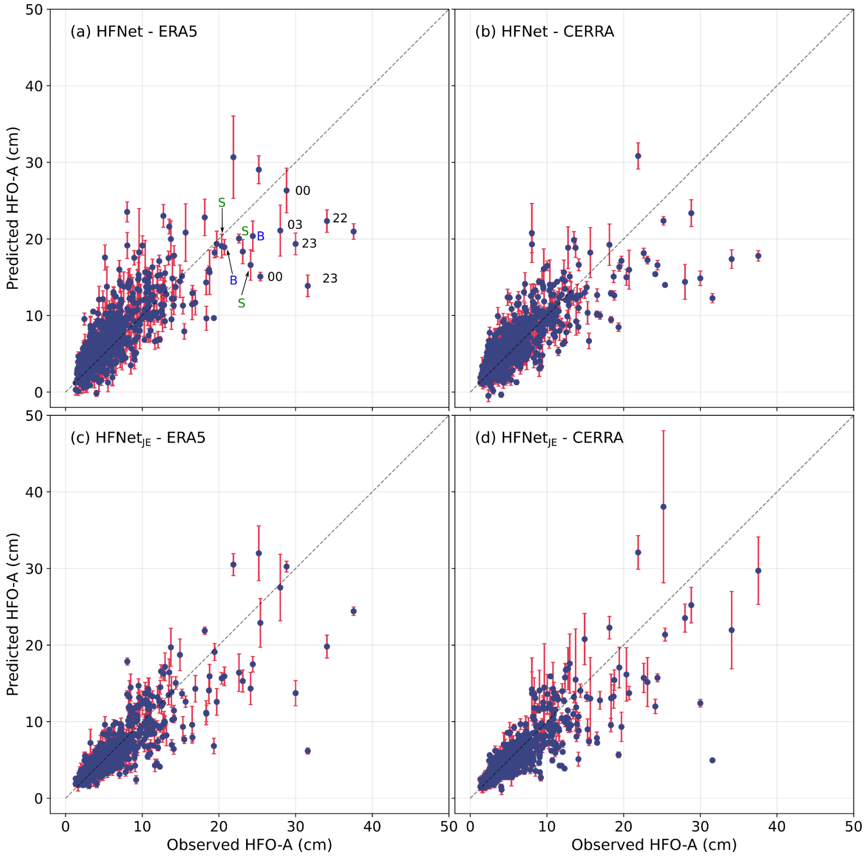

Figure 6. Scatter plots of observed and predicted HFO-A at the Bakar station for 2022–2023 with a 24-hour lead time. Blue dots show the mean predictions across three runs, and red bars indicate the corresponding standard deviations: (**a**) HFNet with ERA5, (**b**) HFNet with CERRA, (**c**) HFNet$_{JE}$ with ERA5, and (**d**) HFNet$_{JE}$ with CERRA. Numbers and letters shown next to selected extreme events in panel (a) indicate the hour of occurrence (for events near midnight) and the event type (S – Sirocco, B – Bora). All other extreme events are associated with Calm conditions (C) and are not individually labelled for clarity.

Comparison of the two architectures, independent of the meteorological input data, shows that according to most metrics HFNet$_{JE}$ performs better for regular events and consequently for all events, which are dominated by the more frequent regular cases. Relative to HFNet, HFNet$_{JE}$ reduces RMSE and MAE by about 13% and 26%, respectively, for all events, and by about 21% and 29% for regular events. In contrast, extreme events (HFO-A ≥ 20 cm) are generally better predicted by the adapted HIDRA model, HFNet, for which RMSE and MAE are lower by approximately 10% and 7%, in comparison to the ones for HFNet$_{JE}$. It is worth noting that adding the Richardson number as an additional input variable to the HFNet$_{JE}$-ERA5 configuration (Table 2) resulted in errors for regular events that were comparable to those reported here, while errors for extreme events increased by approximately 13% in RMSE and 10% in MAE (not shown). This may suggest that HFNet$_{JE}$

internally learnt relationships among atmospheric variables and captured indicators of atmospheric instability more effectively than the explicitly prescribed Richardson number under extreme conditions.

Among all tested configurations, HFNet$_{JE}$-ERA5 (Table 2) provides the best overall performance for the full dataset (All) and for regular events (Reg), whereas HFNet-ERA5 (Table 2) performs best for extreme events (Ext). In addition, HFNet$_{JE}$ always predicted positive HFO-A values, whereas HFNet occasionally produced negative amplitudes (Fig. 6a,b), which are physically unplausible, although such cases were rare. Notably, in HFNet experiments using both ERA5 and CERRA input, relative error decreases with increasing event amplitude, resulting in lower relative errors for extreme than for regular events (Tables 2 and 3).

Model performance at the Ploče station (Tables 4 and 5) is better than at Bakar and both architectures reproduce the general relationship between observed and predicted HFO-A (Fig. 7). Compared to Bakar, RMSE and MAE at Ploče are lower for nearly all configurations and event groups, except HFNet$_{JE}$-ERA5-All and HFNet$_{JE}$-CERRA-Reg. Depending on the configuration, RMSE is reduced up to 52%, while MAE is reduced up to 58%. This likely reflects a simpler prediction problem at Ploče, where the range of HFO-A is smaller (Fig. 5) and the frequency response is less complex, with a single dominant mode excited under all three meteorological backgrounds (Section 4.1). Extreme events are also less biased and more accurately predicted at Ploče (Tables 4 and 5) than at Bakar (Tables 2 and 3).

Table 4. Mean performance metrics (RMSE, MAE, relative error, bias, and accuracy) for the 24-hour lead-time forecasts using ERA5 input at the Ploče station, averaged over three runs. Metrics are reported for all HFO-As (All), regular events (HFO-A < 20 cm; Reg), and extreme events (HFO-A ≥ 20 cm; Ext). Accuracy is defined as the proportion of predictions differing from observations by less than one standard deviation. Underlined values denote optimal values within each category (All, Reg, and Ext).

Model	Amplitude type	Events	RMSE (cm)	MAE (cm)	Relative error (%)	Bias (cm)	Accuracy (%)
HFNet	All	727	2.67	2.12	51.23	1.25	78.56
HFNet$_{JE}$			2.22	1.45	29.31	0.02	89.45
HFNet	Reg	716	2.64	2.09	51.72	1.31	78.96
HFNet$_{JE}$			2.05	1.38	29.28	0.07	90.36
HFNet	Ext	11	4.87	3.75	18.95	-2.99	54.54
HFNet$_{JE}$			7.27	6.02	31.64	-3.16	30.30

Comparison of results obtained using ERA5 and CERRA input data, independent of model architecture, reveals a less consistent signal than at Bakar. Differences between the two datasets are relatively small for all and regular events, with slightly better scores generally obtained using CERRA input. For extreme events, the comparison is less consistent, with some metrics favouring CERRA and others ERA5, although CERRA performs better according to most metrics. Overall, results at the Ploče station suggest that the higher spatial resolution of CERRA modestly improves the representation of regular events and may also benefit the prediction of extremes.

Comparison of the two architectures, independent of the meteorological input data, shows the same general behaviour as at Bakar, with a clear signal that HFNet$_{\mathrm{JE}}$ performs better for regular and consequently all events. Relative to HFNet, HFNet$_{\mathrm{JE}}$ reduces RMSE and MAE by about 12% and 27%, respectively, for all events, and by about 16% and 28% for regular events, values comparable to those obtained at Bakar. In contrast, extreme events (HFO-A ≥ 15 cm) are generally better predicted by HFNet, for which RMSE and MAE are lower by substantially larger amounts than at Bakar, by about 25% and 29%, respectively.

Among all tested configurations, HFNet$_{\mathrm{JE}}$-ERA5 (Table 4) provides the best overall performance for the full dataset (All) and for regular events (Reg), whereas HFNet-ERA5 (Table 4) performs best for extreme events (Ext). In Ploče, both models predict only physically realistic positive HFO-A values.

Both stations exhibit the same overall behaviour: HFNet$_{\mathrm{JE}}$ performs better for regular and consequently all events, whereas HFNet performs better for extreme events. Differences between ERA5 and CERRA are generally small for regular and all events, while ERA5 tends to be more reliable for forecasting extremes. At both stations, extreme events are predicted more accurately and consistently by HFNet (Tables 2 and 4), which shows smaller variability across runs (Figs. 6 and 7), lower RMSE, MAE, and relative error, and higher accuracy than HFNet$_{\mathrm{JE}}$. Nevertheless, all configurations systematically underestimate extreme events, and prediction uncertainty increases with HFO amplitude, particularly in HFNet$_{\mathrm{JE}}$ simulations, indicating greater sensitivity to initial conditions. This suggests that the main limitation arises from data imbalance rather than model architecture, with both models being well constrained for frequent moderate events but much less constrained for rare extremes because of the limited number of training examples.

Table 5. Mean performance metrics (RMSE, MAE, relative error, bias, and accuracy) for the 24-hour lead-time forecasts using CERRA input at the Ploče station, averaged over three runs. Metrics are reported for all HFO-As (All), regular events (HFO-A < 20 cm; Reg), and extreme events (HFO-A ≥ 20 cm; Ext). Accuracy is defined as the proportion of predictions differing from observations by less than one standard deviation. Underlined values denote optimal values within each category (All, Reg, and Ext).

Model	Amplitude type	Events	RMSE (cm)	MAE (cm)	Relative error (%)	Bias (cm)	Accuracy (%)
HFNet	All	727	2.55	1.95	46.05	0.93	81.20
HFNet$_{\mathrm{JE}}$			2.35	1.54	33.42	0.35	89.41
HFNet	Reg	716	2.47	1.91	46.43	1.00	81.66
HFNet$_{\mathrm{JE}}$			2.22	1.49	33.53	0.41	90.22
HFNet	Ext	11	5.56	4.18	21.11	-3.63	51.51
HFNet$_{\mathrm{JE}}$			6.67	5.12	26.23	-3.52	36.36

https://doi.org/10.5194/egusphere-2026-3411
Preprint. Discussion started: 25 June 2026

When tested on training data (not shown), both models reproduce the training data well. However, HFNet predictions appear more tightly clustered around the one-to-one line, particularly for larger amplitudes, while HFNet$_{JE}$ shows greater dispersion and stronger underestimation for extremes. This behaviour is consistent with the results obtained on the independent test data, where HFNet performs better for extreme events.

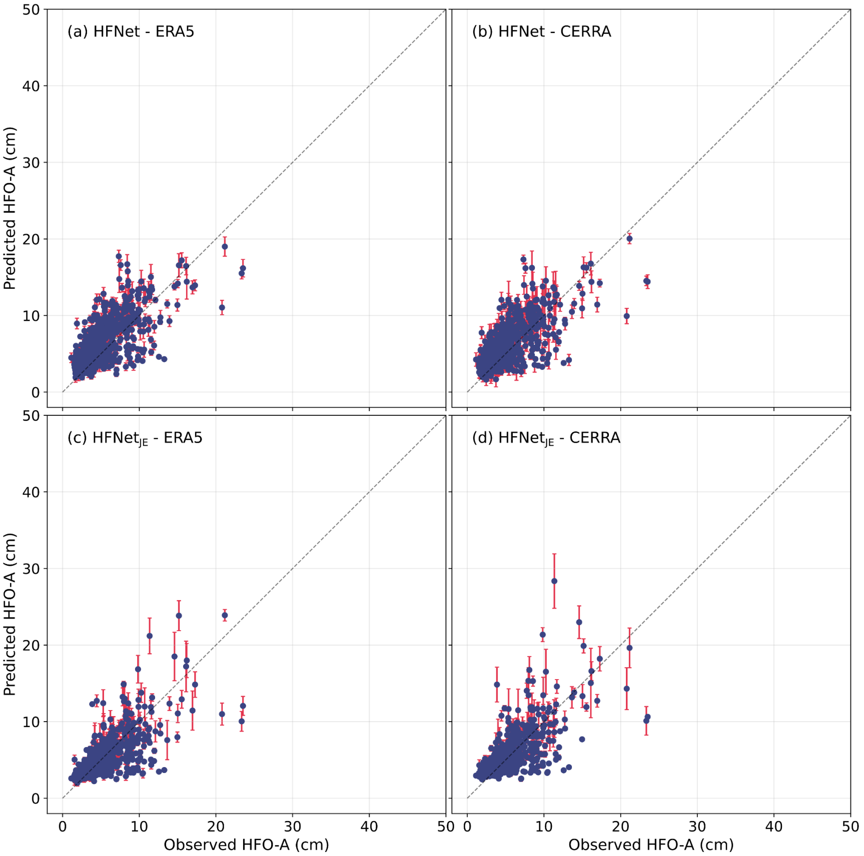

Figure 7. As in Fig. 6 but for Ploče station.

4.3 Ablation study

The ablation analysis averaged across both stations and all amplitude categories is presented in Table 6, while station-specific and amplitude-specific results are provided in the Supplement (Tables S1 and S2, Figs. S2 and S3 in the Supplement). The analysis focuses on the HFNet-ERA5 configuration because it achieved the best performance for extreme (Ext) amplitudes at both stations, which is the primary focus of this study. Table 6 highlights the sensitivity of model performance to the choice of input predictors. Each reduced configuration yields modest improvements across all metrics. Overall, the largest reduction in error is obtained for HFNet$_{-Ri}$, with RMSE and MAE decreasing by around 10%. When considering each station and each subset separately (All, Reg, and Ext), multiple reduced configurations improve performance for the overall and regular-event cases. In particular, the exclusion of the Richardson number (HFNet$_{\text{-Ri}}$, Tables S1 and S2 in the Supplement) results in consistently lower RMSE and MAE, reduced relative error, near-neutral bias, and slightly higher accuracy, suggesting that this predictor may not contribute beneficial information for more common conditions, or that its influence is already implicitly captured by other variables. During regular events, the Richardson number often attains comparably large values associated with the absence of an unstable layer (i.e., strong wind shear) above a stable layer capable of supporting gravity-wave propagation. While the Richardson number exhibits little variability under these conditions, the corresponding SL responses remain variable. Consequently, inclusion of the Richardson number may degrade forecast performance for regular events, as it provides little information for distinguishing among them, although it remains a valuable predictor during extreme events. Similar, albeit weaker, improvements for the overall and regular-event cases are observed when excluding MSLP, relative humidity, wind components, or geopotential. However, these gains are generally accompanied by a deterioration in performance for extreme events, for which the full model tends to perform better. This behaviour indicates that these predictors likely contain information that becomes particularly relevant under high-amplitude conditions, even if their contribution to average performance is limited.

Table 6. Mean performance metrics, averaged over three runs, both stations (Bakar and Ploče), and all amplitudes for the HFNet ablation experiments. Variants exclude individual predictors: HFNet$_{\text{-PSD}}$ (spectrum); HFNet$_{\text{-Ri}}$ (Richardson number); HFNet$_{\text{-MSLP}}$ (mean sea-level pressure); HFNet$_{\text{-r}}$ (relative humidity); HFNet$_{\text{-HFO}}$ (previous 24 hours HFOs); HFNet$_{\text{-T}}$ (air temperature); HFNet$_{\text{-(u,v)}}$ (wind components); HFNet$_{\text{-z}}$ (geopotential). The reference HFNet model includes all predictors. Optimal values are underlined.

Model variations	RMSE (cm)	MAE (cm)	Relative error (%)	Bias (cm)	Accuracy (%)
HFNet	2.92	2.17	46.55	1.07	83.77
HFNet$_{\text{-PSD}}$	2.85	2.15	47.47	1.00	83.81
HFNet$_{\text{-Ri}}$	<u>2.65</u>	<u>1.9</u>	<u>39.04</u>	<u>0.40</u>	<u>86.75</u>
HFNet$_{\text{-MSLP}}$	2.77	2.04	43.43	0.69	85.44
HFNet$_{\text{-r}}$	2.79	2.02	42.39	0.56	85.99
HFNet$_{\text{-HFO}}$	2.77	2.04	43.53	0.94	85.30
HFNet$_{\text{-T}}$	2.88	2.06	43.01	0.82	85.16
HFNet$_{\text{-(u,v)}}$	2.81	2.08	44.91	0.77	85.12
HFNet$_{\text{-z}}$	2.86	2.05	42.78	0.70	84.93

A notable exception is the exclusion of the previous 24 hours of HFO time series (HFNet$_{\text{-HFO}}$; Tables S1 and S2 in the

Supplement), which leads to improvements across all subsets, including extremes. This may indicate redundancy with other predictors or limitations in how this variable is represented in the current model configuration. As noted, HFNet is based on the HIDRA architecture adapted for the HF sea-level forecasting problem. In the original HIDRA framework, past SL information was processed using only a single dense layer because, in the LF setting, this input mainly provided information about the current SL state. For the HF sea-level problem, however, recent variability may contain additional predictive information that is not adequately captured by such a simple representation.

Across all experiments, extreme events exhibit a pronounced negative bias, indicating a systematic underestimation of large amplitudes. This bias becomes stronger when atmospheric predictors are removed, highlighting their role in representing high-amplitude conditions. Overall, the results suggest that models optimized for average conditions do not necessarily perform best for extremes, and that retaining a broader set of physically relevant predictors is important for capturing such events. At the same time, the improved performance in some reduced configurations points to potential redundancy among certain inputs.

Figures S2 and S3 (in the Supplement) show the dependence of model error on event amplitude and highlight substantial differences between configurations at higher HFO-A values. For low-amplitude events (below ~15 cm), all experiments show similar performance, with only minor differences in MAE, indicating limited sensitivity to predictor selection in this range. In contrast, as amplitude increases, the spread between configurations becomes more pronounced, with errors increasing rapidly across all models. This suggests that the representation of higher-amplitude events is more sensitive to the choice of input variables. In particular, the exclusion of large-scale atmospheric predictors such as geopotential, and MSLP generally leads to higher errors at larger amplitudes, indicating their importance for capturing extreme conditions. Conversely, the configuration without HFO input consistently yields among the lowest MAE values across most amplitude bins, including the highest ones, suggesting that this predictor may be redundant or not optimally represented in the current setup. As the number of events decreases with increasing amplitude, the estimates for the highest bins are subject to greater uncertainty.

4.4 Prediction of HFOs using ECMWF forecasts

Since reanalysis incorporates assimilated measurements and therefore does not represent a true forecast, we additionally evaluated model performance using ECMWF 3-day forecasts. Based on the results presented so far, HFNet achieved the best performance for extreme amplitudes at both stations. Furthermore, the ablation analysis showed that the HFNet configuration without HFO input (HFNet$_{\text{-HFO}}$; Tables S1 and S2 in the Supplement) performed best for extreme events while also improving predictions for overall and regular amplitudes. Therefore, this configuration was selected for the forecast evaluation. In these experiments, the meteorological input consisted of the previous 24 hours of ERA5 reanalysis combined with ECMWF 3-day forecast data. For SL input, only the PSD was retained. We present results for selected time intervals between 2022 and 2023, during which numerous extremes were observed. Results for the Bakar station are shown in Figs. 8

https://doi.org/10.5194/egusphere-2026-3411
Preprint. Discussion started: 25 June 2026

and 9, while corresponding results for the Ploče station are provided in the Supplement (Figs. S4 and S5).

Figure 8 shows that the spread of predicted values increases with lead time. At 24 hours, points are closely clustered around the one-to-one line, whereas at 48 hours and 72 hours they become progressively more dispersed. Deviations from the diagonal are more pronounced at higher HFO-A values. This increase in scatter is reflected in larger differences between predicted and observed values, corresponding to higher RMSE and a more noticeable negative bias at higher amplitudes.

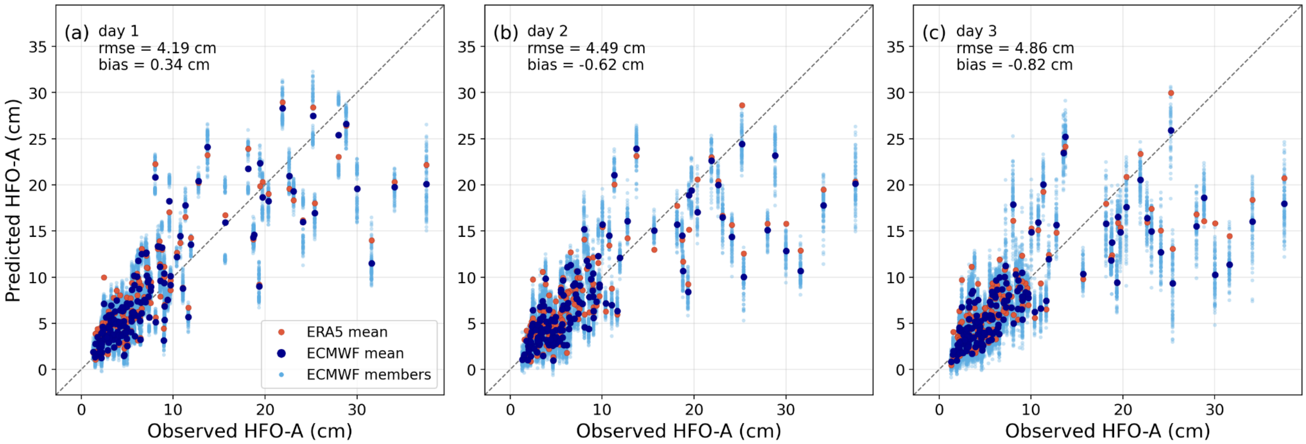

Figure 8. Scatter plots of observed and predicted HFO-A (model: HFNet$_{\text{HFO}}$) at the Bakar station for August 2022 and July, August, October, and November 2023. Panels show lead times of (**a**) 24 hours, (**b**) 48 hours, and (**c**) 72 hours. Dark blue dots represent mean predictions from the 50-member ECMWF ensemble (averaged across three runs), red dots indicate corresponding means based on ERA5 (also averaged across three runs), and light blue dots show individual ECMWF ensemble members.

Figure 9 shows that the predicted HFO-A time series reproduces the main temporal features of the observed signal across all selected periods, including the timing of individual peaks and intervening low-amplitude intervals. The ensemble-mean predictions follow the observed variability, although peak amplitudes are usually lower than the corresponding observed values. These differences are most evident during short-duration, high-amplitude events, where the predicted peaks appear smoother and less pronounced. Outside of peak events, predicted and observed values are more closely aligned, with smaller deviations throughout periods of moderate variability. Temporal offsets between predicted and observed peaks (Fig. 9a, b) are occasionally visible. The largest HFO-A events, associated with Calm-type atmospheric conditions, exhibit the greatest underestimation, whereas events linked to Sirocco forcing show closer agreement between predicted and observed values.

As noted, analogue predictions for the Ploče station are provided in the Supplement. Compared to Bakar, RMSE values in Ploče are lower for all three forecast days (Fig. S4 in the Supplement), consistent with experiments using reanalysis data. In addition, extreme events predicted using ECMWF forecasts more closely match those obtained from ERA5 fields, and the ensemble spread is smaller, particularly for the 24-hour lead time. All three forecasts, however, show a tendency to overestimate HFO-A (i.e., a positive bias) in particular for events with amplitudes lower than 15 cm. Events with higher amplitudes are typically underestimated. The HFO-A time series (Fig. S5 in the Supplement) lead to conclusions similar to those for the Bakar station, although fewer extreme events are present.

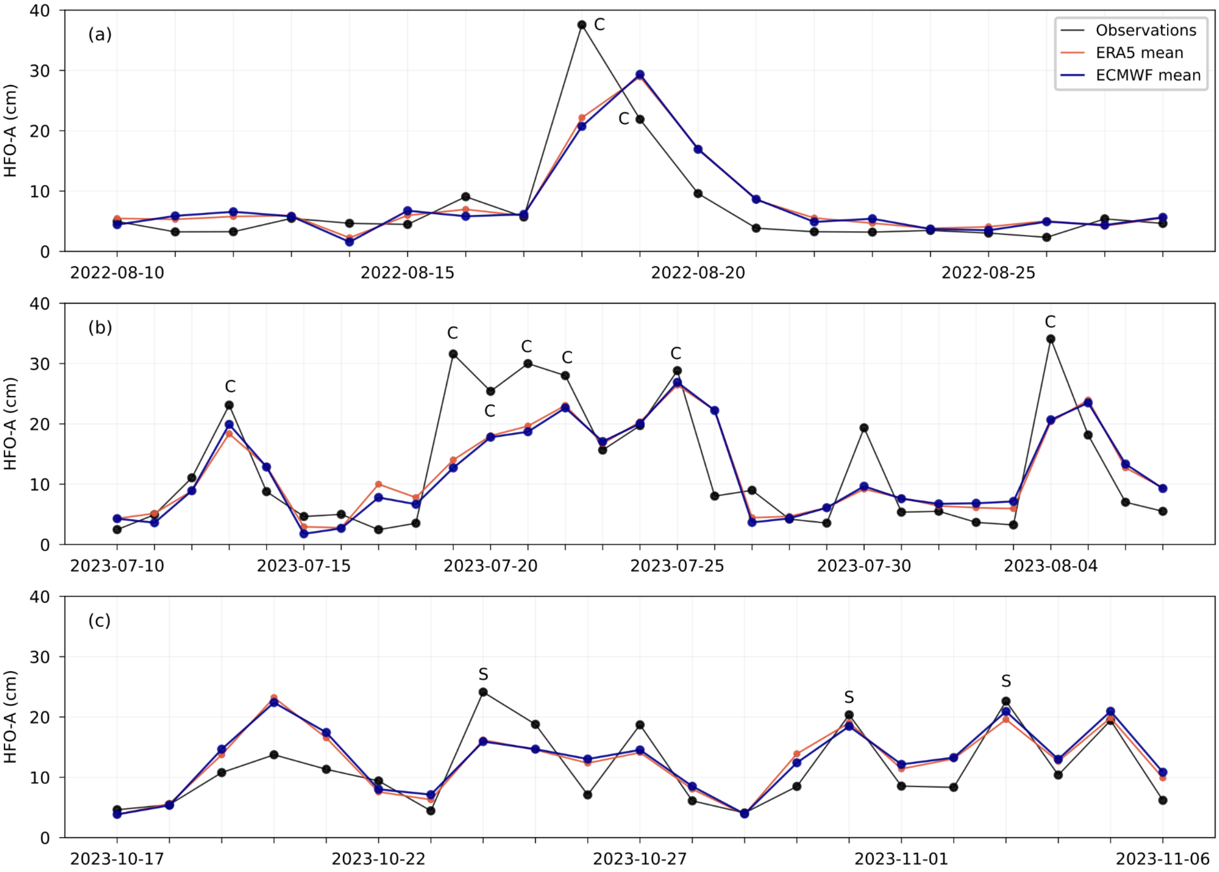

Figure 9. Time series of observed and predicted HFO-A (model: HFNet$_{\text{-HFO}}$) at the Bakar station with a 24-hour lead-time for selected periods: (**a**) August 2022, (**b**) July–August 2023, and (**c**) October–November 2023. Dark blue dots represent ensemble-mean predictions from the 50-member ECMWF forecasts (averaged across three runs), while red dots indicate the corresponding ERA5-based means. Letters marking observed values equaling or exceeding 20 cm denote the event class (see Section 4.1).

5 Discussion

In this study, we explored the predictability of daily amplitudes of HFOs at two Adriatic locations using more than 20 years of 1-minute observations combined with atmospheric reanalysis and forecast data. Within a broader experimental framework, multiple model configurations were tested, including variations in architecture, input variables, targets, and loss functions, in order to identify an optimal setup.

The experiments suggest that model performance depends strongly on event amplitude. A clear difference emerges between regular and extreme events, with large-amplitude oscillations remaining substantially more difficult to predict. This

https://doi.org/10.5194/egusphere-2026-3411
Preprint. Discussion started: 25 June 2026

behaviour is consistent across architectures and input datasets, indicating that the main limitation is not model design alone, but the intrinsic difficulty of predicting rare, large-amplitude events.

In this context, we further examined the role of lead times and target definitions. The results suggest that shorter lead times do not necessarily improve the predictability of HFO-As. In an experiment where HFO-A was predicted every 6 hours up to 24 hours ahead (setup as HFNet-ERA5; Table 2), forecast skill improved for overall and typical events, but decreased significantly for extremes. A similar trade-off was observed when directly forecasting HFO time series in the same setup. In this case, the loss function was defined in the frequency domain (FFT amplitude) rather than on the HFO signal itself, to avoid penalizing phase errors and to focus on the amplitude of oscillations.

We additionally tested an amplitude-weighted loss function, defined as $L=\alpha\, L(Reg)+\beta\, L(Ext)$, with $\alpha<\beta$, where Reg and Ext amplitudes were defined in Table 2. Several weight combinations were evaluated to increase sensitivity to extreme events. However, this approach had a limited impact on performance. For example, in the HFNet-ERA5 setup for the Bakar station, it produced similar metrics for All and Reg amplitudes, while performance for Ext events was slightly degraded.

A similar aspect was considered for input variables. We evaluated whether including total sea level provides additional predictive information. This was motivated by the fact that intense HFOs are associated with total SL extremes and contribute to compound events (Ruić et al., 2023). Including total sea level (HFNet-ERA5 setup) improved performance for overall and typical amplitudes, but degraded performance for extremes.

In addition, sources of uncertainty arise both from the definition of daily targets and from differences in HFO-generating mechanisms. Events occurring near the boundaries of the 24-hour windows (e.g., the second target in Fig. 3 and the extremes marked in Fig. 6) may be split between consecutive prediction intervals. This can lead to inconsistencies between predicted and observed amplitudes (Fig. 6a). The effect is particularly relevant for longer-lasting or multi-peak events and may contribute to apparent prediction errors that are not directly related to model performance. At the same time, Fig. 6a shows no clear dependence of prediction skill on the type of HFO-generating mechanism (Sirocco, Bora, and Calm conditions). Cases associated with Sirocco and Bora do not exhibit better predictive performance than Calm events, proving that within all three synoptic scale settings the processes which are responsible for the most extreme events are not properly reproduced by atmospheric models nor/or recognised by our methods.

From an evaluation perspective, a direct comparison with a hydrodynamic baseline forecast for Bakar and Ploče is not possible, as there is currently no operational HFO forecasting system for the Adriatic Sea. Although advanced deterministic modelling frameworks such as AdriSC (Denamiel et al., 2019a) have been developed and tested for the region, they are not operated in real time any more. As a result, no established hydrodynamic reference forecast is available for direct benchmarking against the approaches evaluated in this study.

A key question that follows is whether the limitations in HFO predictability arise from the modelling approach itself (e.g., the atmospheric encoders) or from the limited information content of the atmospheric input fields. In this study, the latter was recognized from the outset: reanalysis products such as ERA5 (~30 km, 1 hour) and CERRA (5.5 km, 3 hour) cannot resolve the small-scale, rapidly evolving atmospheric disturbances responsible for generating strong HFOs. This limitation is not specific to reanalysis products but reflects a broader challenge in atmospheric modelling. Even high-resolution numerical models struggle to adequately reproduce meteotsunamigenic disturbances, often not reproducing correctly their intensity, misrepresenting their timing, and introducing positional errors of tens of kilometres (e.g., Vilibić et al., 2025; Rahimian et al., 2022; Horvath et al., 2018; Belušić et al., 2007). These deficiencies arise from both insufficient resolution and model parameterizations, including the treatment of air-sea coupling, and persist even at kilometre-scale resolutions. As a result, atmospheric forcing for meteotsunami studies is often simplified using idealized pressure disturbances, highlighting the difficulty of fully representing the relevant processes in operational or research-grade atmospheric models.

In this light, our hypothesis was therefore not that the input fields fully describe the relevant processes, but that deep learning models might still extract sufficient information from large-scale atmospheric patterns to provide useful HFO predictions. This idea is supported by previous work, such as Šepić et al. (2016a), who constructed an index based on large-scale atmospheric variables to predict meteotsunami occurrence. Their index successfully captured conditions favourable for meteotsunami generation and demonstrated predictive skill despite relying only on large-scale atmospheric features. However, the approach was primarily effective in identifying the likelihood of events rather than accurately quantifying their intensity. In this context, our approach can be seen as an improved data-driven extension of such methods.

Nevertheless, our results suggest that this strategy has inherent limitations: while the models perform reasonably well for typical conditions, their skill decreases for extreme events, which are most strongly controlled by unresolved small-scale processes. This indicates that the predictability of HFOs is fundamentally constrained by the information content of the atmospheric fields, and that further improvements may require either higher-resolution inputs or additional sources of information.

Accordingly, improving HFO predictability likely requires augmenting atmospheric inputs with additional sources of information that better capture the relevant forcing mechanisms. In particular, incorporating observations of air pressure disturbances (e.g., from microbarograph stations in the Adriatic; Šepić and Vilibić, 2011) could help detect small, rapidly evolving signals that are not represented in reanalysis fields. Radar reflectivity atmospheric data may further provide insight into convective systems associated with strong HFOs. In addition, when creating 2D forecasts covering multiple coastal stretches, including static parameters such as bathymetry and land-sea masks could help account for local resonance properties and site-specific responses. Finally, extending the spatial domain of reanalysis inputs may allow the model to better capture synoptic conditions associated with intense HFOs.

6 Conclusions

Using more than 20 years of minute-resolution sea-level measurements at Bakar and Ploče, combined with atmospheric reanalysis (ERA5 and CERRA) and forecast data (ECMWF), this study systematically evaluated the potential of deep-learning models to predict daily amplitudes of high-frequency sea-level oscillations in the Adriatic Sea. The main conclusions are:

1. HFO extremes at both stations occur preferentially during the warm season and are associated mainly with Calm and Sirocco conditions in Bakar, and Calm and Bora conditions in Ploče.
2. Forecast skill is substantially higher for regular than for extreme HFO-As, with extreme events being systematically underestimated.
3. Joint encoding of atmospheric input features (i.e., early input fusion) with more extensive SL input processing (HFNet_{JE}) yielded superior performance for overall and regular events, whereas individual variable encoding (i.e., late input fusion) with simpler SL processing (HFNet) proved more robust for extreme events, highlighting a trade-off between average and extreme-event performance.
4. HFO-As are more predictable at Ploče than at Bakar, likely due to a simpler and less variable local sea-level response.
5. The higher-spatial-resolution CERRA reanalysis does not provide a consistent improvement over ERA5, suggesting that the additional information contained in CERRA, including its finer spatial resolution, does not enhance predictive skill in the present setup.
6. Predictors that improve average model performance do not necessarily improve the representation of extremes.
7. The main limitation of the proposed approach most likely arises from the limited information content of atmospheric inputs for rare, small-scale meteotsunamigenic forcing.

Overall, deep learning provides a computationally efficient framework for predicting HFO amplitudes and shows clear skill for regular conditions. However, reliable prediction of rare high-amplitude events remains challenging and will likely require improved representation of atmospheric forcing through additional observations or higher-fidelity inputs.

Appendix A: Detailed description of HFNet model

The atmospheric encoder (Fig. 4) consists of two sequential blocks. In the spatio-temporal block, atmospheric variables are encoded separately to extract spatial structures. The resulting features are then concatenated and processed jointly in the temporal block, which captures temporal dependencies. Each atmospheric variable is represented as a tensor of size $b \times l \times 96 \times 56 \times 72$ (batch size, level, time, latitude, longitude). The input first undergoes initial downsampling using a 3D max-pooling operation with a kernel size and stride of $1 \times 2 \times 2$, where the first dimension corresponds to time and the remaining two to space. This operation reduces the spatial resolution while preserving the temporal dimension, resulting in tensors of size $b \times l \times 96 \times 28 \times 36$. The downsampled tensors are then processed by a sequence of 3D convolutional operations forming the spatio-temporal block. First, a 3D convolutional layer with 64 kernels of size $2 \times 3 \times 3$ and stride $2 \times 3 \times 3$ is applied, followed by a rectified linear unit (ReLU) and 3D dropout. Next, a second 3D convolutional layer with 256 kernels of size $1 \times 3 \times 3$ and stride $1 \times 3 \times 3$ is applied, again followed by ReLU and 3D dropout. Finally, a third 3D convolutional layer with 512 kernels of size $1 \times 3 \times 4$ completes the spatio-temporal processing. Encodings of six variables from the first step (each having a size of 512×48) are concatenated into a feature vector 3072×48 and then processed by a second stage (temporal part) consisting of 1D convolution with 256 kernels of temporal size 5 to increase the receptive field of the temporal layer. Next, two 1D convolutions with 256 kernels of temporal size 1 were applied and are followed by a scaled exponential linear unit (SeLU) activation, one-dimensional dropout, and a residual connection. These layers enable the extraction of more complex features from the atmospheric fields. After flattening, the output is a single vector of 11264 atmospheric features.

In the next step, the 1D features extracted by the atmospheric encoder and the two SL-related inputs (past HFOs and PSD) are each processed separately using a single dense layer, yielding three feature vectors of size 512. These vectors are then concatenated and passed through four residual blocks. Each residual block consists of a fully connected layer that preserves the feature dimension (3·512), followed by a SeLU activation and dropout, with a residual connection applied across the block. The resulting combined features are subsequently mapped to a feature vector of size 8192 using a dense layer. Finally, a fully connected output layer maps these features to the target values, representing the maximum daily HFO-A for the next three days.

Data and code availability

The 1-minute sea level measured in Bakar was obtained from GFZ (Department of Geophysics, Faculty of Science, University of Zagreb) which from recently has been publicly available in the IOC database (https://www.ioc-sealevelmonitoring.org/station.php?code=baka). The 1-minute sea level measured in Ploče was obtained from Port authority of Ploče. The spatial fields of the ERA5 reanalysis are publicly available in the Copernicus Climate Change Service Climate Data Store (single levels: https://doi.org/10.24381/cds.adbb2d47; pressure levels: https://doi.org/10.24381/cds.bd0915c6). The spatial fields of the CERRA reanalysis are publicly available in the Copernicus Climate Change Service Climate Data Store (single levels: https://doi.org/10.24381/cds.622a565a; pressure levels: https://doi.org/10.24381/cds.a39ff99f). ECMWF forecasts are accessible to ECMWF member states (including the authors' countries) and their authorized users via ECMWF facilities (https://www.ecmwf.int/en/forecasts/accessing-forecasts). The HFNet model code is available on GitHub (https://github.com/imedugorac/HFNet-Adriatic), and a subset of the training samples from TG Bakar, provided for demonstration purposes, is available on Zenodo (https://doi.org/10.5281/zenodo.20583597).

Author contributions

All authors contributed to the conceptualization of the study. NM performed the initial analysis at the Bakar station, while IM carried out the remaining analyses. SC prepared the data for the Ploče station, and IM prepared the data for the Bakar station. JŠ contributed to the interpretation of the results. IM and ML wrote the manuscript, and all authors contributed to its editing. ML, with support from MK, supervised the analysis.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Disclaimer

Acknowledgements

We thank the Slovenian National Supercomputing Network (SLING consortium) for access to the VEGA computing cluster. We also thank the personnel of HHI and GFZ for maintaining the tide-gauge stations and for collecting and quality-controlling the data. IM research within Marie Skłodowska-Curie Postdoctoral Fellowship Programme SMASH was supported by the SMASH Consortium host institution Slovenian Environment Agency (ARSO).

Financial support

This research has been supported by: the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie Postdoctoral Fellowship Programme, SMASH co-funded under the grant agreement No. 101081355 and co-funded by the Republic of Slovenia and the European Union from the European Regional Development Fund; the Croatian Science Foundation projects CroClimExtremes (IP-2022-10-4144), and “Young Researchers' Career Development Project – Training New Doctoral Students” (DOK-NPOO-2023-10-7076); ERC projects SHEXtreme (853045), and MeD-Track (101213756); the STIM–REI 2.0 project (PK.1.1.10.0005), the ZMAJ project funded by the European Union – NextGenerationEU, the VAL project (PK.3.4.17.0021) funded by the European Regional Development Fund (ERDF); the Slovenian Research and Innovation Agency (ARIS) projects (P1-0237 and P2-0214).

References

https://doi.org/10.5194/egusphere-2026-3411
Preprint. Discussion started: 25 June 2026

Barzandeh, A., Ličer, M., Rus, M., Kristan, M., Maljutenko, I., Elken, J., Lagemaa, P., and Uiboupin, R.: Application of the HIDRA2 deep-learning model for sea level forecasting along the Estonian coast of the Baltic Sea, Ocean Science, 21(4), 1315–1327, https://doi.org/10.5194/os-21-1315-2025, 2025.

Belušić, D., Grisogono, B., and Bencetić Klaić, K.: Atmospheric origin of the devastating coupled air-sea event in the east Adriatic, Journal of Geophysical Research, 112, D17111, https://doi.org/10.1029/2006JD008204, 2007.

Bertin, X., Bakker, A., van Dongeren, A., Coco, G., Andre, G., Ardhuin, F., Bonneton, P., Bouchette, F., Castelle, B., Crawford, W., Davidson, M., Deen, M., Dodet, G., Guérin, T., Inch, K., Leckler, F., McCall, R., Muller, H., Olabarrieta, M., and Tissier, M.: Infragravity waves: From driving mechanisms to impacts, Earth-Science Reviews, 177, 774–799, https://doi.org/10.1016/j.earscirev.2018.01.002, 2018.

Denamiel, C., Šepić, J., Ivanković, D., and Vilibić, I.: The Adriatic Sea and coast modelling suite: evaluation of the meteotsunami forecast component, Ocean Modelling, 135, 71–93, https://doi.org/10.1016/j.ocemod.2019.02.003, 2019a.

Denamiel, C., Šepić, J., Huan, X., Bernardet, C., and Vilibić, I.: Stochastic surrogate model for meteotsunami early warning system in the eastern Adriatic Sea, Journal of Geophysical Research: Oceans, 124(11), 8485–8499, https://doi.org/10.1029/2019JC015574, 2019b.

Diffenbaugh, N.S. and Giorgi, F.: Climate change hotspots in the CMIP5 global climate model ensemble, Climatic Change, 114, 813–822, https://doi.org/10.1007/s10584-012-0570-x, 2012.

Euro-Mediterranean Tsunami Catalogue: https://ingv.maps.arcgis.com/apps/webappviewer/index.html?id=effd15b2bf1843e5a7ba47c03b5972d5, last access: 23 January 2026.

Goldberg, J. and Kempni, K.: On the sea level oscillations in Bakar Bay and on the problem of bay seiches in general, Prirodoslovna Istraživanja Kraljevine Jugoslavije, 21, 9–235, 1938.

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., Janisková, 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, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J-N.: ERA5 hourly data on pressure levels from 1940 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS), https://doi.org/10.24381/cds.bd0915c6, last access: 20 September 2024, 2023a.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J. N.: ERA5 hourly data on single levels from 1940 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS), https://doi.org/10.24381/cds.adbb2d47, last access: 20 September 2025, 2023b.

Horvath, K., Šepić, J., and Telišman Prtenjak, M.: Atmospheric forcing conducive for the Adriatic 25 June 2014 meteotsunami event, Pure and Applied Geophysics, 175(11), 3817–3837, https://doi.org/10.1007/s00024-018-1902-1, 2018.

Jansà, A. and Ramis, C.: The Balearic rissaga: from pioneering research to present-day knowledge, Natural Hazards, 106, 1269–1297, https://doi.org/10.1007/s11069-020-04221-3, 2021.

Leutbecher, M. and Palmer, T.: Ensemble forecasting, Technical Memorandum, European Centre for Medium-Range Weather Forecasts (ECMWF), https://doi.org/10.21957/c0hq4yg78, 2007.

Lindzen, R.S. and Tung, K.K.: Banded convective activity and ducted gravity waves, Monthly Weather Review, 104(12), 1602–1617, https://doi.org/10.1175/1520-0493(1976)104<1602:BCAADG>2.0.CO;2, 1976.

Loshchilov, I. and Hutter, F.: Decoupled weight decay regularization, arXiv [preprint], https://doi.org/10.48550/arXiv.1711.05101, 2017.

Međugorac, I., Pasarić, M., and Orlić, M.: Long-term measurements at Bakar tide-gauge station (east Adriatic), Geofizika, 39, 149–162, https://doi.org/10.15233/gfz.2022.39.8, 2022.

Monserrat, S., Vilibić, I., and Rabinovich, A. B.: Meteotsunamis: atmospherically induced destructive ocean waves in the tsunami frequency band, Natural Hazards and Earth System Sciences, 6, 1035–1051, https://doi.org/10.5194/nhess-6-1035-2006, 2006.

Mourre, B., Santana, A., Buils, A., Gautreau, L., Ličer, M., Jansà, A., Casas, B., Amengual, B., and Tintoré, J.: On the potential of ensemble forecasting for the prediction of meteotsunamis in the Balearic Islands: sensitivity to atmospheric model parameterizations, Natural Hazards, 106, 1315–1336, https://doi.org/10.1007/s11069-020-03908-x, 2021.

Neumann, B., Vafeidis, A.T., Zimmermann, J., and Nicholls, R.J.: Future coastal population growth and exposure to sea-level rise and coastal flooding - a global assessment, PloS ONE, 10(3), p.e0118571, https://doi.org/10.1371/journal.pone.0118571, 2015.

Orlić, M., Belušić, D., Janeković, I., and Pasarić, M.: Fresh evidence relating the great Adriatic surge of 21 June 1978 to mesoscale atmospheric forcing, Journal of Geophysical Research, 115, C0601, https://doi.org/10.1029/2009JC005777, 2010.

Otnes, R.K. and Enochson, L.: Applied time series analysis, vol 1, 2nd edn. Wiley, New York, 1978.

Pérez Gómez, B., Vilibić, I., Šepić, J., Međugorac, I., Ličer, M., Testut, L., Fraboul, C., Marcos, M., Abdellaoui, H., Álvarez Fanjul, E., Barbalić, D., Casas, B., Castaño-Tierno, A., Čupić, S., Drago, A., Fraile, M. A., Galliano, D. A., Gauci, A., Gloginja, B., Martín Guijarro, V., Jeromel, M., Larrad Revuelto, M., Lazar, A., Keskin, I. H., Medvedev, I., Menassri, A., Meslem, M. A., Mihanović, H., Morucci, S., Niculescu, D., Quijano De Benito, J. M., Pascual, J., Palazov, A., Picone, M., Raicich, F., Said, M., Salat, J., Sezen, E., Simav, M., Sylaios, G., Tel, E., Tintoré, J., Zaimi, K., and Zodiatis, G.: Coastal sea level monitoring in the Mediterranean and Black seas, Ocean Science, 18, 997–1053, https://doi.org/10.5194/os-18-997-2022, 2022.

Proudman, J.: The effects on the sea of changes in atmospheric pressure, Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society, 2, 197–209, https://doi.org/10.1111/j.1365-246X.1929.tb05408.x, 1929.

Pugh, D. and Woodworth, P.: Sea-level science: understanding tides, surges, tsunamis and mean sea-level changes, Cambridge University Press, 395 pp., ISBN 9781107028197, 2014.

Rabinovich, A. B.: Seiches and Harbor Oscillations, in: Handbook of Coastal and Ocean Engineering, World Scientific, 193–236, https://doi.org/10.1142/9789812819307_0009, 2009.

Rahimian, M., Beyramzadeh, M., and Siadatmousavi, S. M.: The skill assessment of Weather and Research Forecasting and WAVEWATCH-III models during recent meteotsunami event in the Persian Gulf, Frontiers in Marine Science, 9, 834151, https://doi.org/10.3389/fmars.2022.834151, 2022.

Renault, L., Vizoso, G., Jansà, A., Wilkin, J., and Tintoré, J.: Toward the predictability of meteotsunamis in the Balearic Sea using regional nested atmosphere and ocean models, Geophysical Research Letters, 38(10), L10601, https://doi.org/10.1029/2011GL047361, 2011.

Ruić, K., Šepić, J., Mlinar, M., and Međugorac I.: Contribution of high-frequency (T < 2 h) sea level oscillations to the Adriatic sea level maxima, Natural Hazards, 116, 3747–3777, https://doi.org/10.1007/s11069-023-05834-0, 2023.

Ruić, K., Šepić, J., and Vojković, M.: Synoptic patterns associated with high-frequency sea level extremes in the Adriatic Sea, Ocean Science, 21(3), 1183–1203, https://doi.org/10.5194/os-21-1183-2025, 2025.

Ruić, K.: Extreme sea-level oscillations in the Adriatic - contributions of various processes, associated synoptic conditions and numerical modelling. Ph.D. thesis, Faculty of Science, Department of Geophysics, University of Zagreb, Croatia, 137, 2025.

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, 2023.

Rus, M., Mihanović, H., Ličer, M., and Kristan, M.: HIDRA3: a deep-learning model for multipoint ensemble sea level forecasting in the presence of tide gauge sensor failures, Geoscientific Model Development, 18, 605–620, https://doi.org/10.5194/gmd-18-605-2025, 2025.

Rus, M., Ličer, M., and Kristan, M.: HIDRA-D: deep-learning model for dense sea level forecasting using sparse altimetry and tide gauge data, Geoscientific Model Development, 19(5), 2177–2195, https://doi.org/10.5194/gmd-19-2177-2026, 2026.

Schimanke S., Ridal M., Le Moigne P., Berggren L., Undén P., Randriamampianina R., Andrea U., Bazile E., Bertelsen A., Brousseau P., Dahlgren P., Edvinsson L., El Said A., Glinton M., Hopsch S., Isaksson L., Mladek R., Olsson E., Verrelle A., and Wang Z.Q.: CERRA sub-daily regional reanalysis data for Europe on single levels from 1984 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS), https://doi.org/10.24381/cds.622a565a, last access: 14 January 2026, 2021a.

Schimanke S., Ridal M., Le Moigne P., Berggren L., Undén P., Randriamampianina R., Andrea U., Bazile E., Bertelsen A., Brousseau P., Dahlgren P., Edvinsson L., El Said A., Glinton M., Hopsch S., Isaksson L., Mladek R., Olsson E., Verrelle A., and Wang Z.Q.: CERRA sub-daily regional reanalysis data for Europe on pressure levels from 1984 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS), https://doi.org/10.24381/cds.a39ff99f, last access: 14 January

2026, 2021b.

Šepić, J. and Vilibić, I.: The development and implementation of a real-time meteotsunami warning network for the Adriatic Sea, Natural Hazards and Earth System Sciences, 11, 83–91, https://doi.org/10.5194/nhess-11-83-2011, 2011.

Šepić, J., Vilibić, I., and Monserrat, S.: Quantifying the probability of meteotsunami occurrence from synoptic atmospheric patterns, Geophysical Research Letters, 43(19), 10377–10384, https://doi.org/10.1002/2016GL070754, 2016a.

Šepić, J. and Orlić, M.: Adriatic meteotsunami catalogue, https://projekti.pmfst.unist.hr/floods/meteotsunamis/, last access: 14 January 2026.

Šepić, J., Međugorac, I., Janeković, I., Dunić, N., and Vilibić, I.: Multi-meteotsunami event in the Adriatic Sea generated by atmospheric disturbances of 25-26 June 2014, Pure and Applied Geophysics, 173(12), 4117, https://doi.org/10.1007/s00024-016-1249-4, 2016b.

Ursell, F.: Edge waves on a sloping beach. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, 214, 79–97, https://doi.org/10.1098/rspa.1952.0152, 1952.

Vilibić, I., Denamiel, C., Zemunik, P., and Monserrat, S.: The Mediterranean and Black Sea meteotsunamis: an overview, Natural Hazards, 106, 1223–1267, https://doi.org/10.1007/s11069-020-04306-z, 2021.

Vilibić, I., Zemunik Selak, P., and Šepić, J.: Meteorological tsunamis: from local hazard to global relevance, Reviews of Geophysics, 63(4), e2024RG000867, https://doi.org/10.1029/2024RG000867, 2025.

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