Authors' reply to Anonymous Reviewer's comments on the manuscript https://doi.org/10.5194/egusphere-2026-3411

General evaluation

The MS considers an important and novel topic: how to predict the occurrence of high-frequency (HF) sea-level oscillations of high amplitudes, based on the available forcing data of operational observations and routine meteorological and oceanographic predictions. While sea level data have been observed in recent decades with HF 1-minute intervals, the forcing factors have been acquired on a low-frequency (LF) hourly temporal resolution. This means that possible forcing mechanisms for HF sea level oscillations (such as atmospheric convective systems, gravity waves, fronts, or Proudman resonance, etc.), as nicely presented in the introduction and discussion, are not covered by available data. However, previous studies have established some statistical link between synoptic forcing and HF oscillations.

To overcome the driver-response resolution gap, the MS tests deep-learning methods for predicting maximum daily amplitudes of high-frequency (T < 1 hour) sea-level oscillations at selected geographical sites in the Adriatic Sea. The convolutional neural networks were driven by past HF sea-level observations and LF atmospheric predictors from the two available reanalyses. One of the models tested in the HF range, HFNet, is based on the known HIDRA family of models that work well for the LF range. The second, more HF-adjusted model, HFNetJE, considers more HF details of past sea level observations and uses joint encoding with atmospheric predictions. Both models reproduce the observed daily HF maximum amplitudes, whereas HFNetJE performs better under typical amplitude conditions, but HFNet with better temporal resolution of atmospheric data is more effective at capturing extreme events.

The results of the study are interesting and indicate a way to achieve more justified and accurate HF sea level predictions. Before publishing, I recommend considering some comments given below.

Dear Reviewer,

Thank you for your constructive comments and the time you dedicated to reviewing our manuscript. Our responses to your comments are provided in blue text below. The indicated locations of the added or revised text refer to the manuscript (MS) with tracked changes.

General comments

1) Within the selected 8 model options (two models, two wind fields, two locations), the skill estimation is of crucial importance. Often, skill is estimated by comparing the observed and predicted time series at

the same times. Such an approach is not possible in the present study, because of missing HF forcing data. Instead, the prediction is oriented toward obtaining the "maximum absolute value of the HFO signal within each daily interval", named daily amplitude. The general background for such extreme-value skill metrics should be provided to help readers.

Yes, we agree. We added a justification of this approach in the revised manuscript (lines 196-201, please see below).

> HFO-A was selected as the prediction target because the magnitude of HFOs, rather than their exact phase, is of primary relevance for the associated coastal hazard. The phase-resolved HFO signal cannot be reliably predicted from the available atmospheric input fields, which do not explicitly resolve the short temporal scales and localised disturbances responsible for HFO generation. Daily maximum sea-level quantities have also been used as prediction targets in data-driven sea-level studies, particularly for modelling and reconstructing daily maximum storm surges from atmospheric predictors (e.g., Tadesse et al., 2020; Cid et al., 2017).

2) The performance metrics of model predictions are not very clear. RMSE, MAE, relative error, and bias are used in the common context, although they are not spelled out. Accuracy is explained only in the table heading (line 326) and is not easily understood. The metrics could be introduced in a separate section (for example, in methods), using formulae.

Thank you for pointing this out. In the revised MS, we explicitly defined the performance metrics used in the study in the Methods section (new subsection 3.5, Performance metrics, please see below) and reformulated the definition of accuracy for greater clarity.

3.5 Performance metrics

Model performance was evaluated using root-mean-square error (RMSE), mean absolute error (MAE), relative error, bias, and accuracy. For each experiment, RMSE, MAE, relative error, and bias were calculated as:

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i-y_i)^2} \quad (1)$$

$$\mathrm{MAE}=\frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i-y_i| \quad (2)$$

$$\text{Relative error}=\frac{100}{N}\sum_{i=1}^{N}\frac{|\hat{y}_i-y_i|}{y_i} \quad (3)$$

$$\mathrm{Bias}=\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i-y_i) \quad (4)$$

where y_i and $\hat{y}_i$ are the observed and predicted daily HFO-A, respectively, *i* denotes an individual sample, and *N* is the number of samples in the evaluated subset. RMSE, MAE, and Bias are expressed in centimetres, whereas Relative Error is expressed as a percentage. Accuracy was defined as the percentage of predictions differing from the corresponding observations by no more than one standard deviation of the observed HFO-A:

$$\text{Accuracy}=100\frac{N_{|\hat{y}-y|\leq\sigma}}{N} \quad (5)$$

where $N_{|\hat{y}-y|\leq\sigma}$ is the number of predictions satisfying this criterion and σ is the standard deviation of all observed HFO-A values in the test dataset. The same value of σ was used to calculate accuracy for all HFO-As, regular events, and extreme events. All metrics were calculated separately for all HFO-As (All), regular events (Reg), and extreme events (Ext), with the thresholds for regular and extreme HFO-As defined in Sect. 3.1. Each metric was calculated independently for each of the three experiments performed with different random initialisations of the model parameters, and the mean of the three resulting values is reported.

3) It would also be helpful to know whether the complex models outperform the low-level “persistence forecast”: the next forecasted values within the lead time are equal to the last observed values. Such a comparison could be rather interesting since the prediction of sub-hourly SL variations does not include any sub-hourly driving factors.

Thank you for this suggestion. We calculated the same performance metrics for a persistence-forecast baseline at both Bakar and Ploče, using the maximum HFO-A observed during the last completed 24 h interval as the prediction for the following 24 h interval. The results are discussed in the revised MS (lines 489-492, please see below), while the corresponding numerical values are provided below in Tables RC4 and RC5.

Table RC4: Performance metrics for the persistence forecast baseline at the Bakar station, reported for all

HFO-As (All), regular events (HFO-A < 20 cm; Reg), and extreme events (HFO-A ≥ 20 cm; Ext).

Model	Amplitude type	Events	RMSE (cm)	MAE (cm)	Relative error (%)	Bias (cm)	Accuracy (%)
persistence baseline	All	730	4.3	2.56	40.16	0.01	85.19
	Reg	715	3.6	2.3	39.88	0.25	86.83
	Ext	15	17.23	15.57	53.65	-11.39	6.67

Table RC5: Performance metrics for the persistence forecast baseline at the Ploče station, reported for all HFO-As (All), regular events (HFO-A < 15 cm; Reg), and extreme events (HFO-A ≥ 15 cm; Ext).

Model	Amplitude type	Events	RMSE (cm)	MAE (cm)	Relative error (%)	Bias (cm)	Accuracy (%)
persistence baseline	All	727	3.18	2.11	42.89	0.00	77.82
	Reg	716	2.95	2.01	42.79	0.14	78.88
	Ext	11	10.09	9.16	49.35	-9.16	9.09

The manuscript was modified as indicated below:

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. Comparison with a persistence-forecast baseline, which uses the maximum daily HFO-A from the preceding 24 hours as the forecast (not shown), showed that the DL models generally performed better at both stations. The DL models showed the clearest advantage over the persistence forecast for extreme events, with substantially lower errors and higher accuracy.

4) The manifold numerical values presented in Tables 2, 3, 4, and 5 are not easy to follow. It may be more informative to reduce the number of metrics (for example, MAE values align well with RMSE) and combine the ERA5 and CERRA tables for two locations into composite tables. Regarding graphics, perhaps a Taylor diagram could be useful for presenting many results in a compact manner.

Thank you for this suggestion. We considered different ways of presenting these results during preparation of the manuscript. We initially combined the ERA5 and CERRA results for each station in a single table, but found that the large number of values made the results more difficult to follow. We therefore separated them so that the results being compared and discussed in the text are presented next to each other in the tables. We also considered graphical presentation, but found that relatively small differences between some configurations were less clearly visible than in the tables. Finally, we prefer to retain the full set of performance metrics, as they provide complementary information on different aspects of model performance. For these reasons, we prefer to retain the current tabular presentation.

5) There are sub-hourly and/or spatially high-resolution atmospheric data that could be found and used in the update of the present study or in the subsequent forthcoming studies. Such additional data could include wind gusts, air pressure fluctuations, precipitation patterns from weather radars, current and wave patterns from coastal radars, and possibly more. The data extension possibilities are briefly mentioned in lines 563-566; I would prefer a bit more elaboration, also in the introduction.

In the revised Discussion (lines 654–681, please see below), we provided an overview of potentially relevant high-frequency observations, specifying the available stations, measured variables, temporal resolution and temporal coverage. One of key issues of these additional data is the current lack of a 20 year time series needed for consistent training. Following reviewer suggestion, we addressed the potential of such observations in the Introduction (lines 124-128, please see below).

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 and additional sources of information are required. High-frequency atmospheric observations, including measurements from automatic weather stations, offshore buoys, weather radars, and microbarographs, could provide complementary information on small-scale disturbances that is not captured by reanalysis products. Such observations are increasingly available in the Adriatic and may provide valuable additional inputs for future data-driven HFO forecasting systems (see Sect. 5). 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

I~~Accordingly, i~~mproving 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.~~additional atmospheric observations that better resolve the relevant forcing mechanisms (Šepić and Vilibić, 2011). In particular, high-frequency measurements of air pressure and wind could capture small, rapidly evolving disturbances that are not represented in the reanalysis fields. Considering observations with sampling intervals of 10 minutes or less and stations located in sectors from which HFO-generating atmospheric disturbances may approach Bakar and Ploče, several potentially useful data sources are available. The Croatian Meteorological and Hydrological Service (DHMZ) provides (i) 10-minute air-pressure and wind measurements from automatic weather stations (e.g., Sveti Ivan na Pučini, Porer, Pazin, Mali Lošinj, Split-Marjan, Lastovo, Hvar and Palagruža; DHMZ-Network, 2026), (ii) 10-minute meteorological and oceanographic measurements from offshore buoys (Kvarner, Blitvenica and Viški kanal; DHMZ-Met-Ocean, 2026; DHMZ-Network, 2026), and (iii) 5-minute horizontal and vertical reflectivity measurements from coastal weather radars (Goli, Debeljak and Uljenje; DHMZ-Radar, 2026). Of these, automatic weather stations currently offer the greatest potential for model training because of their longer records, whereas the buoy and radar networks have only been operational since 2022 (DHMZ-Buoys, 2022; DHMZ-Radar2022, 2022) and may become more useful as longer records accumulate. Additional high-frequency air-pressure observations are available from microbarographs. The Institute of Oceanography and Fisheries (IOR) operates a network of microbarographs providing 1-minute air-pressure measurements along the eastern and western Adriatic coasts (IOR-Faust, 2026). However, these records extend back only to 2017, and currently only the eastern-coast stations are active. A microbarograph was also installed at Bakar by GFZ in 2024 (GFZ-Bakar, 2026); while its record is currently too short for model training, it may be valuable for analysing recent HFO events. Other possible model developments include the incorporation of static parameters, such as bathymetry and land–sea masks, in spatial forecasts covering multiple coastal stretches. Such information could help account for local resonance properties and site-specific responses. Extending the spatial domain of the reanalysis inputs may also allow the model to better capture synoptic conditions associated with intense HFOs.

Technical comments

1. In line 14, sub-hourly SL oscillations are “reaching wave heights of several meters”, but in line 24, HF extremes amount to 60 cm in one location and 35 cm in another location. These statements could be rewritten for clarity, as a fast reading of the abstract suggests.

Done!

2. Figures 6-8, with many similar panels, are not easy to follow. Perhaps Figs 6 and 7, which differ only in geographical location, can be combined.

Thank you for this suggestion. We considered combining Figures 6 and 7; however, given the number of panels and the amount of information presented for each station, combining them would result in an overly dense figure and would not improve readability. Therefore, we prefer to retain the separate presentation, which we find provides the clearest representation of the results.

Thank you for your positive recommendation and constructive feedback. We appreciate your time and effort.