Weather and Climate Extremes 48 (2025) 100775

Contents lists available at ScienceDirect

Weather and Climate Extremes

journal homepage: www.elsevier.com/locate/wace

Synoptic-based model for reconstructing and forecasting high-frequency sea-level extremes in the Mediterranean

P. Zemunik Selak [a,*], I. Vilibić [b,c], C. Denamiel [b,c], P. Pranić [a]

[a] Institute of Oceanography and Fisheries, Split, Croatia
[b] Ruđer Bošković Institute, Division for Marine and Environmental Research, Zagreb, Croatia
[c] Institute for Adriatic Crops and Karst Reclamation, Split, Croatia

ARTICLE INFO

Keywords:
Sea level extremes
Meteotsunami
Atmosphere-ocean connection
Synoptic index

ABSTRACT

This paper evaluates the performance of a synoptic index-based model designed to predict extreme non-seismic sea-level oscillations at tsunami timescales (NSLOTTs) across 32 tide-gauge stations in the Mediterranean Sea, where NSLOTTs can contribute up to 50 % of the total sea-level range. The model employs percentile-determined threshold exceedance criteria to define extreme NSLOTT events. A part of the time series containing half of extreme NSLOTT events is used for model training, while the rest is used for performance assessing. The baseline model integrates seven synoptic variables previously identified for a known NSLOTT hotspot and available within atmospheric reanalysis products. Various model configurations and modifications were tested to evaluate adaptability and robustness in forecasting and detecting extreme NSLOTT events. Results indicate that the model success in forecasting extreme events slightly outweighs its success in detecting observed extreme events. For stations where the baseline model performs well, this proficiency remains consistent across different configurations. However, the uncertainty in model performance is greater for these stations compared to those with poorer performance, which show minimal improvement despite configuration adjustments. Sub-basin analysis reveals that tide-gauge stations located in the eastern Adriatic Sea exhibit the best performance on average. These findings provide valuable insights for optimizing the model setup, enhancing its predictive capabilities, and improving its application in projecting extreme NSLOTT events in future climates. Ultimately, this work may contribute to coastal hazard and flooding mitigation, as well as resilience-building efforts, where extreme NSLOTT events could play a substantial role.

1. Introduction

The Mediterranean Sea, surrounded by diverse coastal landscapes and characterized by complex atmospheric dynamics, is a region prone to atmospherically-influenced sea-level oscillations. These oscillations are unified by the name NSLOTT (Non-seismic Sea-Level Oscillations at Tsunami Timescales) (Vilibić and Šepić, 2017; Williams et al., 2021; Zemunik et al., 2022a), and their extreme manifestations are known as meteorological tsunamis (e.g. Monserrat et al., 2006; Rabinovich, 2020). Meteotsunamis in the Mediterranean regions are often triggered by intense, fast-travelling atmospheric disturbances that may accompany thunderstorms, squalls, frontal zones, or other. These disturbances induce rapid changes in sea level and generate long waves with tsunami-like characteristics through the resonant transfer of energy towards the sea via Proudman (1929) or Greenspan (1956) resonance (e.g., Belušić and Strelec Mahović, 2009). The combination of favourable geographic features, including narrow and shallow coastal areas, amplifies the incoming waves, leading to the formation of meteotsunamis (Pattiaratchi and Wijeratne, 2015; Vilibić et al., 2021).

The Mediterranean region boasts a well-developed tide gauge network with a long tradition of monitoring sea-level (Pérez Gómez et al., 2022). This network provides valuable data on sea-level oscillations, including these at minute timescales, allowing for the detection and analysis of NSLOTT and meteotsunami events. Following the recent development of the Minute Sea-Level Analysis dataset (MISELA; Zemunik et al., 2021a,b), which encompasses global minute NSLOTTs records, a new opportunity emerged for systematic research of the NSLOTT signal. This dataset, combined with global reanalysis products like ERA5, enables the exploration of its relationship with atmospheric processes.

* Corresponding author.
E-mail address: zemunik@izor.hr (P. Zemunik Selak).

https://doi.org/10.1016/j.wace.2025.100775
Received 16 August 2024; Received in revised form 7 April 2025; Accepted 5 May 2025
Available online 6 May 2025

The connection between NSLOTTs and specific atmospheric processes was first established several decades ago and remains a subject of ongoing research. Ramis and Jansà (1983) were among the pioneers who observed specific synoptic conditions during the occurrence of the Balearic meteotsunami (known as rissaga), including: (i) a weak surface depression (ii) overtopped by warm African air in the low troposphere, (iii) creating the temperature inversion that separates two air masses, with (iv) pronounced vertical wind shear across the fast-moving and unstable layer in the mid-troposphere. The persistence of similar synoptic situation has been observed in many Mediterranean meteotsunami case studies (e.g., Jansà et al., 2007; Zemunik et al., 2021c; Pupić Vurilj et al., 2023), systematic regional studies (e.g., Šepić et al., 2009; Vilibić and Šepić, 2009; Šepić et al., 2015b) and recently in global studies (Vilibić and Šepić, 2017; Zemunik et al., 2022b). Moreover, the world's only long-functioning meteotsunami warning system, implemented and maintained by State Meteorological Agency for Ciutadella, relies on qualitative assessment of synoptic conditions and comparison to known setups occurring during destructive events (Jansà and Ramis, 2021).

The first attempt to quantify meteotsunamigenic synoptic patterns was developed for the rissaga "hotspot" Ciutadella by Šepić et al. (2016), who demonstrated that favourable synoptic conditions are crucial but insufficient for meteotsunami occurrence. The problem lies in the unpredictability of mesoscale convective systems and internal gravity waves that are generating meteotsunamigenic atmospheric disturbances, with time-space variability over hours and tens of kilometres not capturable by too-coarse atmospheric reanalysis or operational products. Later, Vich and Romero (2021) utilized neural networks with synoptic variables from radio-sounding measurements (temperature, wind, relative humidity) as inputs to identify and forecast meteotsunamis. The authors showed that the estimates of rissagas produced by neural networks were fairly similar to those of computationally expensive atmospheric-ocean numerical simulations (Renault et al., 2011). Furthermore, Zemunik et al. (2022b) developed a site-specific synoptic index for over 300 locations worldwide, which exhibited the highest correlations with NSLOTTs series in the Mediterranean, suggesting that it is particularly well-suited for this region.

This study – presenting a follow-up to the work by Zemunik et al. (2022b) in which the synoptic index was constructed – aims to investigate the sensitivity of the index-based model to various modifications in model configuration and NSLOTT series. Therefore, the purpose of this research is to test the model adaptability to a diverse range of NSLOTT data recorded on tide gauges equipped with different sensors, placed in different positions within or outside of harbours, and under the influence of various bathymetry and topography effects. By systematically testing the model under different conditions, we aim to gain insights into the robustness and limitations of the model, which may guide future improvements with the ultimate goal of developing a reliable and adaptable tool for predicting and monitoring extreme NSLOTT events.

The structure of this article is as follows. Section 2 presents the data and methods used in the analyses, describes the synoptic index-based model and setups for analyses of the model performance under various configurations. Section 3 showcases the performance of the baseline model in comparison with the performance of sensitivity variations. Finally, Section 4 provides a detailed discussion of the results and possible model applications, while the main conclusions are summarized in the last section.

2. Data and methods

2.1. The MISELA dataset

The presented analyses are based on sea-level data from 32 tide-gauge stations located in the Mediterranean Sea (Fig. 1), of which 31 stations belong to the MISELA dataset (Zemunik et al., 2021a, 2021b), while data from the Ciutadella station are provided by the Spanish National Geographic Institute.

The MISELA dataset comprises a high-frequency component (<2 h) of quality-checked global sea-level measurements sampled at 1-min resolution, making it appropriate for research of short-period processes and oscillations in sea-level. As seismic tsunamis are extraordinary events that can bias analyses and statistics, their records are not included in the dataset. Data within the MISELA dataset originate from three distinct sources: (1) the Intergovernmental Oceanographic Commission (IOC) Sea Level Station Monitoring Facility (SLSMF), (2) the Institute of Oceanography and Fisheries (IOF), and (3) the Finnish Meteorological Institute. In this study, we used 27 SLSMF and 4 IOF stations. The creation of the MISELA dataset involved several steps performed on the raw data to obtain a research-quality dataset (Zemunik et al., 2021a): (i) removal of tidal signal, (ii) automatic spike detection and removal procedures, (iii) visual inspection of detided series to remove spurious data, (iv) exclusion of records related to seismic tsunamis, and (v) digital filtering of the series to extract high-frequency sea-level oscillations with periods up to 2 h. This procedure ensured that the MISELA dataset is well-suited for investigating short-period sea-level oscillations with atmospheric origins, such as meteorological tsunamis, and more generalized NSLOTTs.

The 31 stations from the MISELA dataset used in this study have series up to June 2018 at the latest. However, for the purpose of this research, the series from stations that remain active afterwards have been extended until June 2022 to obtain more robust analyses. Raw sea level data were retrieved from the SLSMF for the period from June 2018 to June 2022, and the MISELA quality-control procedures were strictly

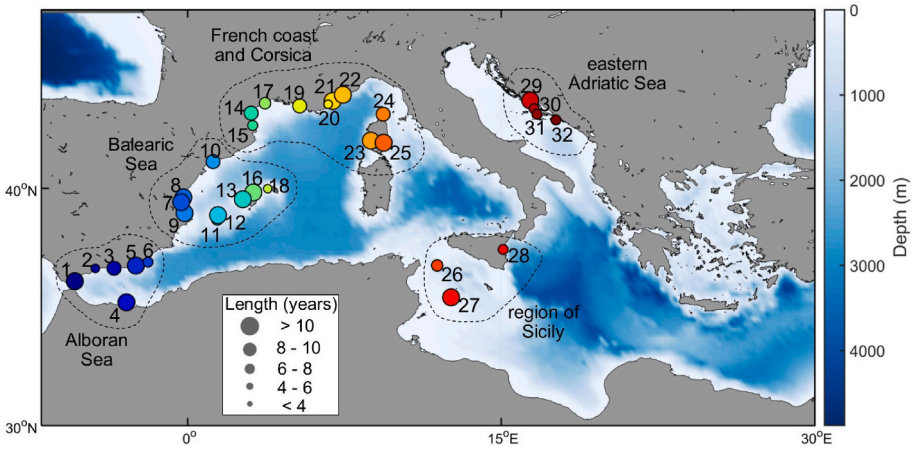

Fig. 1. Map of the stations in the Mediterranean Sea with the denoted regions of Alboran Sea, Balearic Sea, French coast and Corsica, Region of Sicily and Eastern Adriatic Sea.

followed. The data have been filtered to retain only the high-frequency component of the signal and merged with the MISELA dataset. Additionally, the records from the Split station in the MISELA dataset starts in 2017, but older data (since 2009) have been recovered and merged with the original series. For the Ciutadella station, over 3 years of data entered the analyses, commencing in October 2012. In total, 32 stations that are used in the study have series ranging from 2 to 14 years in length. The locations of these stations are shown in Fig. 1.

To determine the ranges of NSLOTTs, the methodology described by Zemunik et al. (2022a, b) was employed. Specifically, the range of high-frequency sea-level oscillations is calculated in each minute as the difference between the upper and lower envelope of the series – smoothed curves that wrap the local extremes and encapsulate the signal. Interpolated data, hour-long data adjacent to data gaps and short data sequences (<6 h), where envelope estimates are not reliable, are discarded.

2.2. Synoptic index-based model

The synoptic index-based model, based on methodology developed by Šepić et al. (2016), was constructed to connect NSLOTTs with synoptic patterns available through atmospheric reanalysis products.

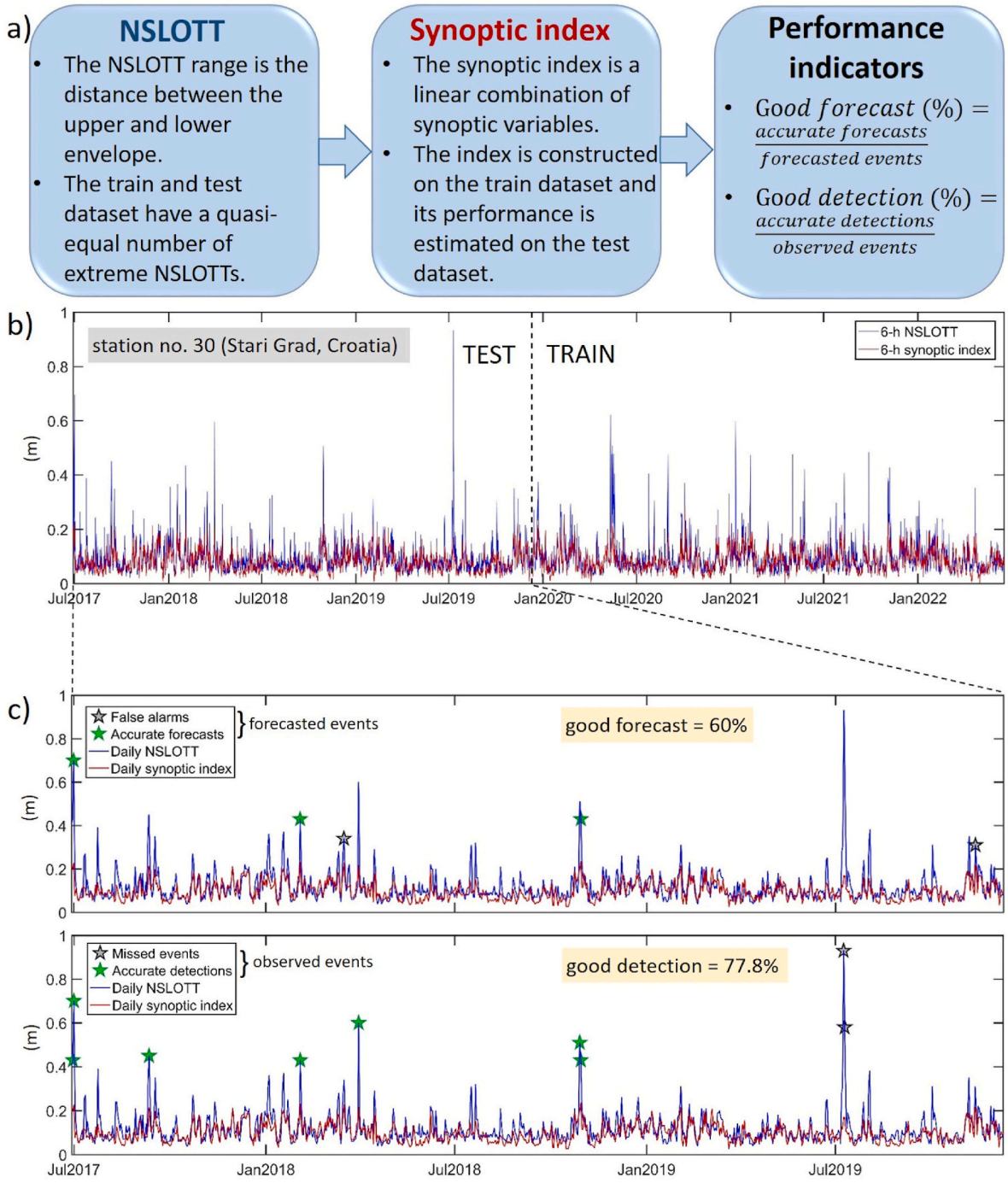

Fig. 2. Diagram of construction of the synoptic index-based model and calculation of its performance, b) 6-h time series of NSLOTTs and synoptic index at the station Stari Grad (Croatia, station 30), c) daily time series of NSLOTTs and synoptic index with indicated accurate forecasts, false alarms, accurate detections and missed events.

Establishing such a connection will enable us to (1) forecast moderate and extreme NSLOTT events, and eventually meteotsunamis, using operational weather forecast products (such as operational forecast of the European Centre for Medium-Range Weather Forecast, ECMWF, www.ecmwf.int), and to (2) project NSLOTTs and meteotsunamis in the future climate. This is significant as climate models are still far from reproducing mesoscale atmospheric processes that generate NSLOTTs but can reproduce synoptic patterns (Vilibić et al., 2018).

The NSLOTT daily series were first partitioned into two parts – a training dataset, upon which the model is constructed and a testing dataset, used for validation of the model – ensuring approximately equal distribution of NSLOTT extreme events in both sets (extreme events are defined as NSLOTTs higher than 99th percentile of the entire series at a tide gauge station). The daily series were created by selecting the maximum value within a day if more than 50 % of minute NSLOTT values are present in a day. The series were split at a midpoint between the two central extreme events, assigning the training dataset to the right (more recent) side and the testing dataset to the left (older) side of the series (an example can be found in Fig. 2).

The synoptic index-based model was constructed on the training dataset as a linear combination of specific synoptic variables, extracted from the ECMWF ERA5 reanalysis dataset (Hersbach et al., 2020, https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5). Following the construction of the NSLOTT index by Zemunik et al. (2022b), seven variables were chosen for the model: vertical profiles of (1) wind speed and (2) relative humidity in the mid-troposphere; (3) surface wind speed; and vertical gradients of (4) temperature in the low-troposphere at the ERA5 grid point closest to the corresponding station, as well as horizontal gradients of (5) mean sea-level pressure, (6) temperature at 850 hPa and (7) geopotential at 550 hPa in a rectangular area spreading 10° around the grid point closest to the station. To match the reanalysis data, subsampled series of 6-h NSLOTTs were created by selecting the highest value every 6 h centred over the middle of the interval. Then, correlations between NSLOTTs and each synoptic variable were computed.

Prior to their introduction to the model, the synoptic variables were normalized by subtracting the median and dividing by the median absolute deviation (MAD) in the 1979–2022 period. The model was built individually for each station, with variables selected one-by-one based on their highest correlation with NSLOTTs and lowest correlation with the previously selected variables in the model. The final number of variables constituting the model was determined by the overall correlation between the model and NSLOTTs. Specifically, if the newly added variable did not significantly increase such correlation, it was discarded. Detailed information on how synoptic index-based model was constructed may be found in Zemunik et al. (2022b, chapter 2.3).

2.3. Performance indicators

To assess the model performance, two indicators have been introduced: (1) *good forecast* and (2) *good detection*. *Good forecast* describes how many extreme NSLOTT events that are forecasted by the model are actual extreme events (observed in the NSLOTTs). It is calculated through the following steps separately for each station: (i) determining the 99th percentile of the daily synoptic index in the training dataset, (ii) detecting synoptic indices during the testing period that surpass the threshold given in step (i) (hereafter referred to as *forecasted events*), (iii) determining the 98th percentile of the entire daily dataset of NSLOTTs (the 98th percentile is chosen to accommodate uncertainty of the forecast), (iv) identifying NSLOTTs during forecasted events that exceed the threshold given in step (iii) (hereafter referred to as *accurate forecasts*). Forecasted events during which NSLOTTs do not exceed threshold given in step (iii) are considered as *false alarms*. The parameter *good forecast* is then computed as the ratio of accurate forecasts to forecasted events, representing a percentage in which NSLOTTs were forecasted successfully.

The second parameter, *good detection*, quantifies how many of observed (real) extreme events are forecasted by the model. It is calculated analogously to the first parameter, as follows: (i) determining 99th percentile of the entire daily dataset of NSLOTTs, (ii) detecting NSLOTTs during the testing period that surpass the threshold given in step (i) (hereafter referred to as *observed events*), (iii) determining 98th percentile of the daily synoptic index in the training dataset, (iv) identifying the index values during observed events that exceed the threshold given in step (iii) (hereafter referred to as *accurate detections*). Observed events during which index values do not exceed threshold given in step (iii) are considered as *missed events*. Finally, the parameter *good detection* is computed as the ratio of accurate detections to observed events and represents the percentage of the extreme NSLOTT events that the model successfully detected. It may be important to note that the number of observed events depends on the length of the series, and varies from 3 to 33 across stations. Stations with longer data series generally have a higher number of events, as they contain more days exceeding the threshold. The number of forecasted events, however, does not increase linearly with the length of the series, since only indices surpassing the training-defined threshold are identified in the testing period. This number varies from 1 to 40 across stations. Fig. 2 depicts the diagram of construction of synoptic index-based model and calculation of its performance, as well as time series of NSLOTTs and synoptic index at the station Stari Grad (Croatia, station 30), with indicated accurate forecasts, false alarms, accurate detections and missed events.

2.4. Sensitivity studies setup

To estimate the sensitivity of the synoptic index-based model to variations in its construction process, several analyses were carried out. First, the model was tested using different reanalyses for the extraction of synoptic variables on which it was trained. Three reanalyses were used here: ERA5 (as a baseline reanalysis), ERA-Interim (ERA-I, https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-interim, Dee et al., 2011) coming from the European Centre for Medium-range Weather Forecast (ECMWF), and NCEP/DOE Reanalysis II (NCEP, https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html, Kanamitsu et al., 2002) from the National Centers for Environmental Prediction and the National Center for Atmospheric Research. These reanalyses differ in spatial resolution – 0.25° for ERA5, 0.75° for ERA-I and 2.5° for NCEP – as well as model physics, parameterizations and data assimilation techniques. Since the ERA-I reanalysis stops in August 2019, the NSLOTT series over common interval (until August 2019) was used for this sensitivity analysis. Additionally, a single split step was chosen at each station to ensure equal training and testing sets introduced to the model. The major idea behind this sensitivity analysis was to determine if the model benefits from using more advanced and higher-resolution synoptic variables compared to those with lower resolution.

To identify the most relevant synoptic variables describing the atmospheric setup during NSLOTTs, two tests were conducted on the choice of model variables. One test considered only mid-troposphere variables, while the other included all non-surface variables. This test intends to quantify if the model performs similarly without surface variables, which typically exhibit more complex synoptic patterns and dynamics due to surface and orographic influences, making them less reproducible by climate models. Further, the influence of training and testing dataset choices was assessed by swapping these sets, setting the training set to an older time period and the testing set to a more recent period. This analysis aimed to see if the model performance is invariant to the choice of the data used for training and testing.

The model performance was also examined under various temporal resolutions, generating a series of NSLOTTs and synoptic variables with time steps of 1 h, 2 h, 3 h, 6 h, 12 h and 24 h. The NSLOTT series were constructed by selecting the maximum 1-min value in the selected data window, while the synoptic variables were taken at exact times centred

over the interval. This analysis aimed to understand the impact of temporal resolution on model efficiency, i.e., to determine if low-resolution products (e.g., daily) yield similar model performance or not.

To evaluate the sensitivity of the model to data gaps, random gaps were artificially introduced into the NSLOTT time series, creating scenarios with data gaps of 20 %, 30 %, 40 %, and 50 %. For this sensitivity study, only stations with good data coverage (gaps less than 20 %) were used. The question here is whether a large number of data gaps in the series lowers model performance. Next, for stations with the longest series, the model sensitivity to data length was examined by shortening these series by two and 4 years at the end, to see if the length of the series is introducing instabilities in the model solutions.

Lastly, the sensitivity of the model to different tresholds that determine accurate forecasts and detections was tested by adjusting the second percentile in the calculations of *good forecast* and *good detection.* The original second percentile is set to the 98th percentile (see chapter 2.3, step (iii) in both indicators), and in this sensitivity test, it was changed to the 97th, 96th and 95th percentile.

In summary, these comprehensive analyses contribute to a thorough understanding of the model robustness and its performance under diverse configurations. The characteristics of all sensitivity analyses is listed in Table 1.

Table 1
Sensitivity variations of the model and their characteristics.

Abbrev.	Type of sensitivity	Short description	Change from the baseline model	BASELINE MODEL CHARaCTERISTICS
REA	Varying in the model setup (configuration)	Changing reanalysis product utilized in the model	• ERA-I • NCEP	Reanalysis product: • ERA5
VAR		Changing synoptic variables utilized in the model	• 5 non-surface synoptic variables • 2 mid-troposphere synoptic variables	Synoptic variables: • 7 synoptic variables
SWA		Swapping of training and testing datasets	• Training on the left side of the series, testing on the right side	Training and testing datasets: • Training on the right side of the series, testing on the left side
RES	Varying the NSLOTT series	Changing temporal resolution of the NSLOTT series	• 1 h • 2 h • 3 h • 12 h • 24 h	Temporal resolution: • 6 h
GAP		Changing the amount of gaps in the NSLOTT series (tested only for 14 stations with less than 20 % of data gaps)	• 20 % • 30 % • 40 % • 50 %	Amount of data gaps: • Original amount of gaps in series (varying from 3 % to 55 % across stations)
LEN		Changing the length of the NSLOTT series (tested only for 10 stations with longest series)	• Series shortened by 2 y • Series shortened by 4 y	Length of time series: • Original length of series (varying from 2 to 14 y across stations)
THR	Varying the estimation of model performance	Changing percentile threshold that determine accurate forecasts and accurate detections	• 97 p • 96 p • 95 p	Percentile threshold: • 98 p

3. Performance of the synoptic index-based model

3.1. Baseline model

The baseline model was constructed using the ERA5 reanalysis with a temporal resolution of 6 h. Fig. 3a illustrates the correlation between the model and the NSLOTTs for the entire period of NSLOTTs. The correlations vary across the stations, ranging between 0.45 and 0.74, indicating moderate to high correlation levels. Moreover, the correlations appear to be lowest in the Western Mediterranean, along the coasts of Spain and France, with values mostly up to 0.55. Exceptions are observed in the Balearic Islands where correlations may surpass 0.6. In the Eastern Mediterranean, including the coasts of Corsica, Sicily, and the Adriatic Sea, the correlations are highest, surpassing 0.6 at almost all stations and reaching up to 0.74 (stations 24 and 26).

Fig. 3b and c shows that stations with higher *good forecast* values typically exhibit higher *good detection* values as well. For example, stations located in the Adriatic Sea (stations 29–32) have some of the highest values for both *good forecast* and *good detection*, whereas stations on the coast of Spain (stations 5–10) exhibit poor model performance, with both values below 20 %. However, some stations exhibit a high *good forecast* value conjoined with a low *good detection* value (stations 22, 29). Approximately half of the stations show higher *good forecast* than *good detection*, indicating fewer false alarms than missed extreme events. Conversely, for the remaining stations, the situation is reversed, with a higher success rate in detecting (recognizing) real events than in accurately forecasting events. Although the median values for overall *good forecast* and *good detection* are similar (around 20 %), *good forecast* values may surpass 80 %, indicating quite good model performance at these stations.

3.2. Individual sensitivity studies

When changing reanalyses upon which the synoptic index-based model is trained (experiment REA in Table 1, Fig. 4), the index built on the NCEP reanalysis generally yields to slightly lower values of both *good forecast* and *good detection*. The median values of *good forecast* and *good detection* for the NCEP reanalysis are 25 % and 19.4 %, respectively, compared to 29.4 % and 27.3 %, for the ERA5 reanalysis. However, this underperformance does not seem substantial enough to provide a general conclusion about the best choice of reanalysis (there are some exceptions on which the model performance differs a lot, such as stations 26 and 29). For stations with lower model performance (such as stations number 7–17 in the Central Mediterranean), the difference in using various reanalyses is minor. However, for stations with higher performance indicators (such as stations 21, 22, 24 in the Ligurian Sea and 29–32 in the Adriatic Sea), the variations are more pronounced. For instance, *good forecast* values in Vela Luka (Croatia, station 31) vary from 60 % for ERA5 to 100 % for NCEP, which contrasts with the general underperformance by NCEP. *Good detection* shows the opposite – the highest values for ERA5 and the lowest for NCEP. Generally, the model with ERA-I performs between ERA5 and NCEP.

The next sensitivity analysis involved constructing the model with

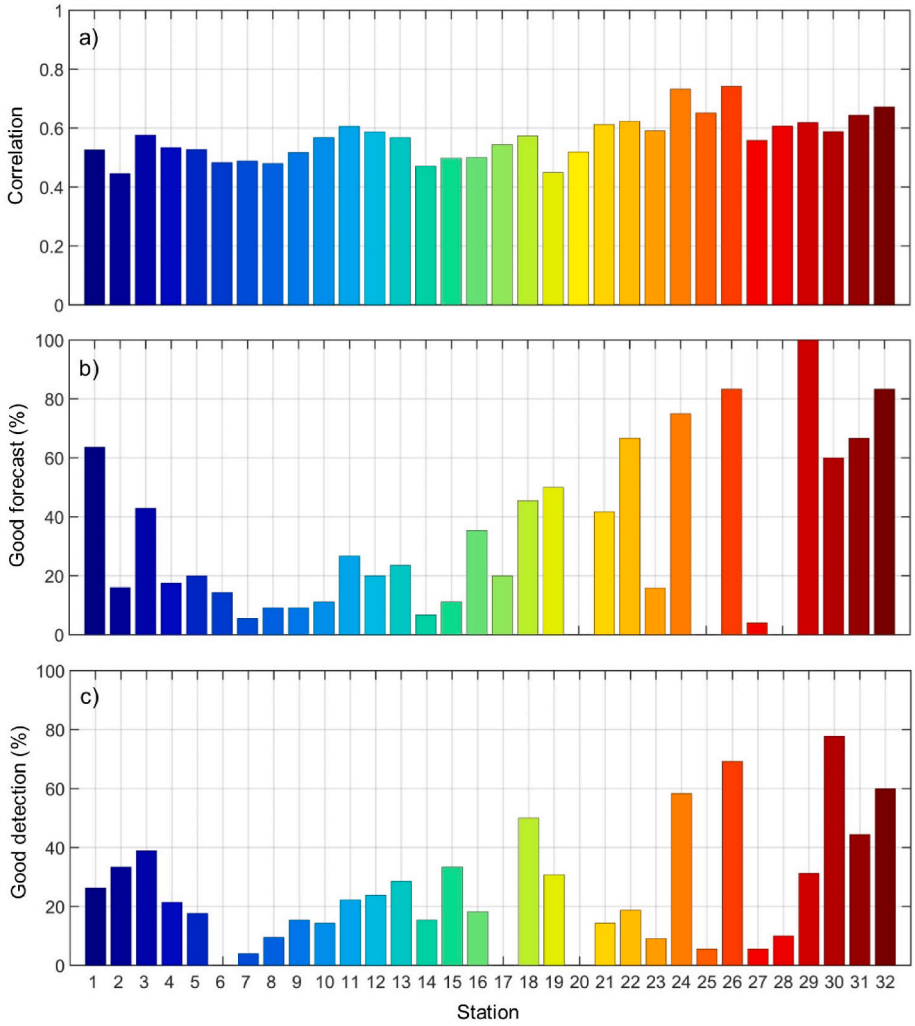

Fig. 3. Bar plot of the a) baseline model-NSLOTTs correlations, b) *good forecast* and c) *good detection*.

different atmospheric synoptic variables (experiment VAR in Table 1, Fig. 5). The model exhibits slightly lower performance when utilizing only the two mid-troposphere variables (with a median of *good forecast* and *good detection* at 16 %). This suggests that although these variables are the most relevant for occurrence of extreme NSLOTT events (e.g., Zemunik et al., 2022b), the model performes better when other synoptic variables are introduced. The success rate of the model constructed with all synoptic variables and with non-surface variables is very similar, with quasi-equal median values for both parameters (around 20 %). However, a substantial decrease in model performance without surface variables may occur at some stations (e.g., stations 24, 26 and 29).

The next analysis documents sensitivity of the model to the choice of training and testing datasets (experiment SWA in Table 1; Fig. S1). In most cases, the model performs better when the training dataset is longer than the testing dataset (not shown). This may be due to grouping of NSLOTT extreme values (>99th p) over certain time intervals, rather then being homogeneously distributed over the whole time series. In other words, the baseline model yields better outcome when extremes are more frequent in the latter part of the series (e.g., stations 1, 3, 19, 32). Conversely, swapping the training and testing datasets provides better performance if the extremes are more frequent at the beginning of the series (e.g., stations 8, 10, 28).

Furthermore, three analyses were conducted to test the model sensitivity to different characteristics of the NSLOTT series. Changes in temporal resolution resulted in little or no change in the model performance (experiment RES in Table 1, Fig. S2). For stations with higher performance indicators, the highest values were frequently obtained in the series with higher resolution (1, 2 or 3 h; e.g., stations 22, 24, 26, 31, 32). However, this difference does not seem substantial. Even lower model performance sensitivity was observed for varying amounts of data gaps in the NSLOTT series (experiment GAP in Table 1; Fig. S3). This can be attributed to the model construction process, where 6-h NSLOTT series on which the baseline model is built, are created with a criterion of containing at least 50 % of the minute data in each window to calculate the maximal value and assign it to the 6-h series. Thus, if the maximal value from the window is not covered by the added gap, the final 6-h series remains unchanged, and the index is consequently identical.

The sensitivity of the synoptic index-based model to the length of the NSLOTT series (experiment LEN in Table 1; Fig. S4) at the ten longest stations, each having over 11 years of data, generally does not change the performance indicators. However, at a few stations, a bigger difference is noticeable. For instance, the station in Solenzara (France, station 25) shows the highest discrepancy in model performance, with 0 % of *good forecast* for the full time series and time series shortened by 4 years, but 100 % of *good forecast* for the time series shortened by 2 years. In the latter case, only one event is forecasted (i.e. the index in testing set surpass the 99th percentile calculated on the training set), and it is successfully recognized as an extreme event (i.e., the corresponding NSLOTT surpasses the 98th percentile calculated on the entire set), resulting in 100 % of *good forecast*.

The last sensitivity analysis evaluated the model performance under

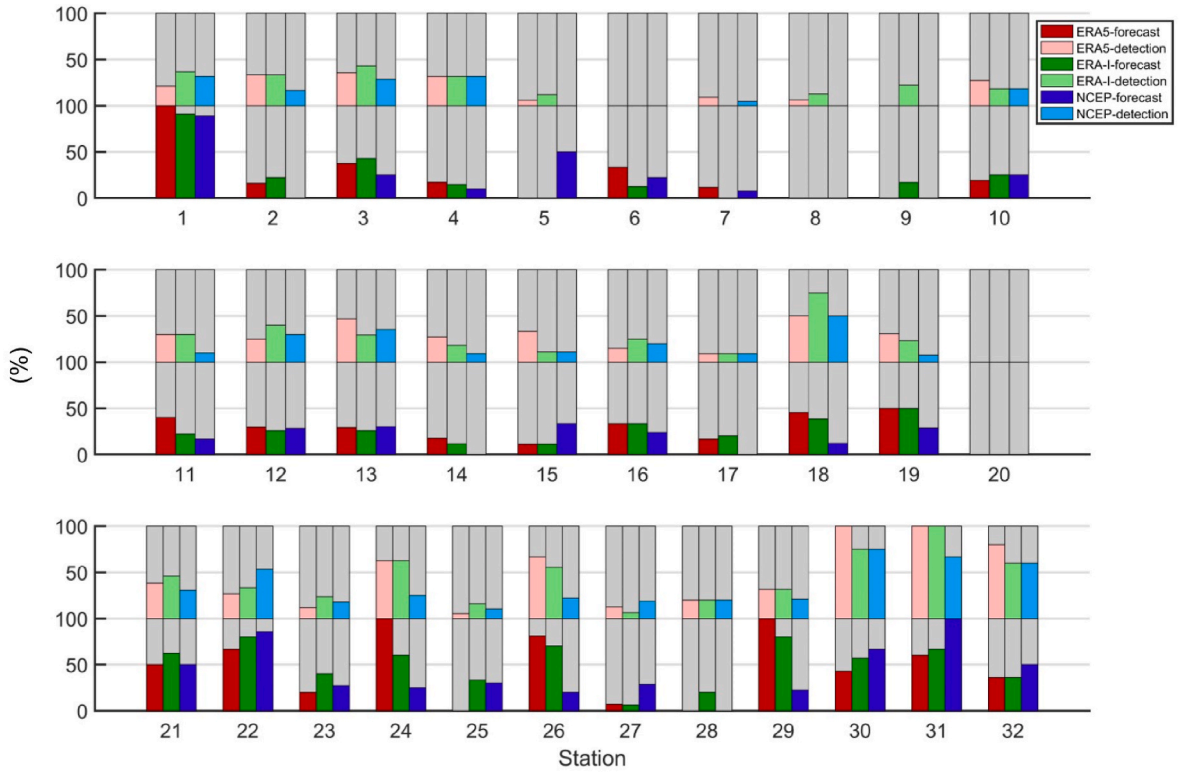

Fig. 4. Bar plot of *good forecast* and *good detection* estimated with the model based on ERA5, ERA-Interim and NCEP reanalyses (experiment REA).

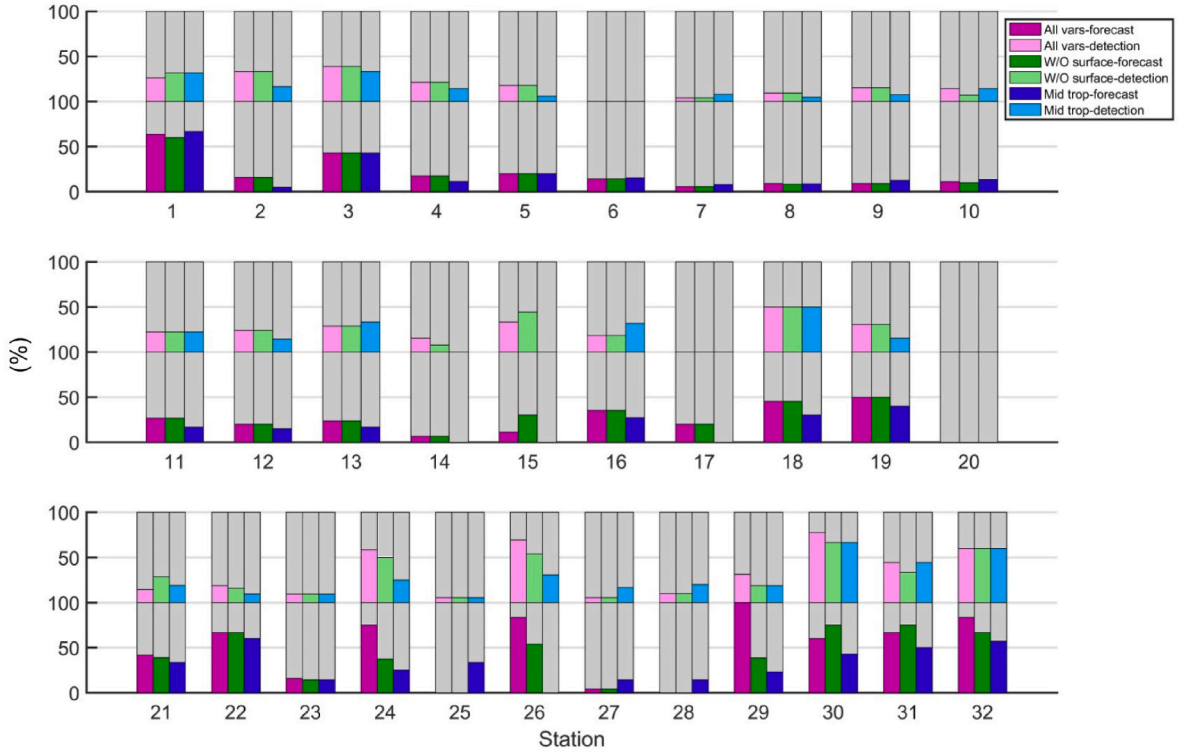

Fig. 5. Bar plot of *good forecast* and *good detection* estimated with the model based on all synoptic variables, only non-surface synoptic variables and only mid-troposphere synoptic variables (experiment VAR). Note that the model based on all synoptic variables is the baseline model.

different thresholds that classify events as extremes (experiment THR in Table 1; Fig. S5). Expectedly, the model performance improves as the threshold percentile is decreased, i.e., if the model allows for higher uncertainty of its performance. As the criterion for classifying an event as extreme becomes less strict, more events are categorized as extremes, resulting in a clear increase in the model success rate at all stations. The

median increase from the baseline model in *good forecast* across all stations is 11.1 %, 17.4 % and 24.5 % in case of the 97th, 96th and 95th percentile, respectively, while for *good detection* it is somewhat lower, at 7.4 %, 15.1 % and 16.7 %, respectively.

3.3. Composite analysis

After analyzing the performance of each model variation, composite analyses were conducted to compare the skills of the models trained under different configurations. Fig. 6 presents the range of model-NSLOTTs correlations at each station (Fig. 6a) and the probability density function for each sensitivity variation (Fig. 6b). At certain stations (e.g., stations 1, 3, 5, 11, 18), the correlation between the model and NSLOTTs remains relatively stable for different sensitivity studies, indicating that the overall agreement between NSLOTT and synoptic variables is invariant to changes in input variables. This, however, does not seem to be the case for some stations with highest correlations, such as stations 24 and 26, suggesting that the quality of the relationship between NSLOTT and synoptic variables may be affected by certain choices. In general, correlations are the lowest (median value 0.48) when building the model with just two mid-troposphere synoptic variables (Fig. 6b), while the median of correlations ranges between 0.55 and 0.60 for the rest of the sensitivity choices. Some sensitivity variations display narrower distributions, indicating more consistent correlations across different stations, while for other variations, the distributions are wider. Roughly, for variations in temporal resolution, the probability density function becomes narrower as the resolution decreases. Sensitivity variations related to data gaps and the length of the time series are not included in this analysis as they were conducted only on selected stations and could give inaccurate perception of probability density functions.

The box plot of *good forecast* and *good detection* for each station, along with PDFs for each sensitivity variation, are shown in Fig. 7a and b. Expectedly, the deviations of *good forecast* and *good detection* from their median values at each station are much larger compared to deviations in median correlations (Fig. 7a and b and Fig. 6a). This is because these statistics are based on extreme NSLOTT events only, which are much harder to connect to synoptic papers (Šepić et al., 2016; Zemunik et al., 2022b) compared to the entire series. For approximately half of the stations, the distance between minimum and maximum value of *good forecast* exceeds 50 %, suggesting that the variations in model configuration strongly influence the quality of the model. Overall, the values of

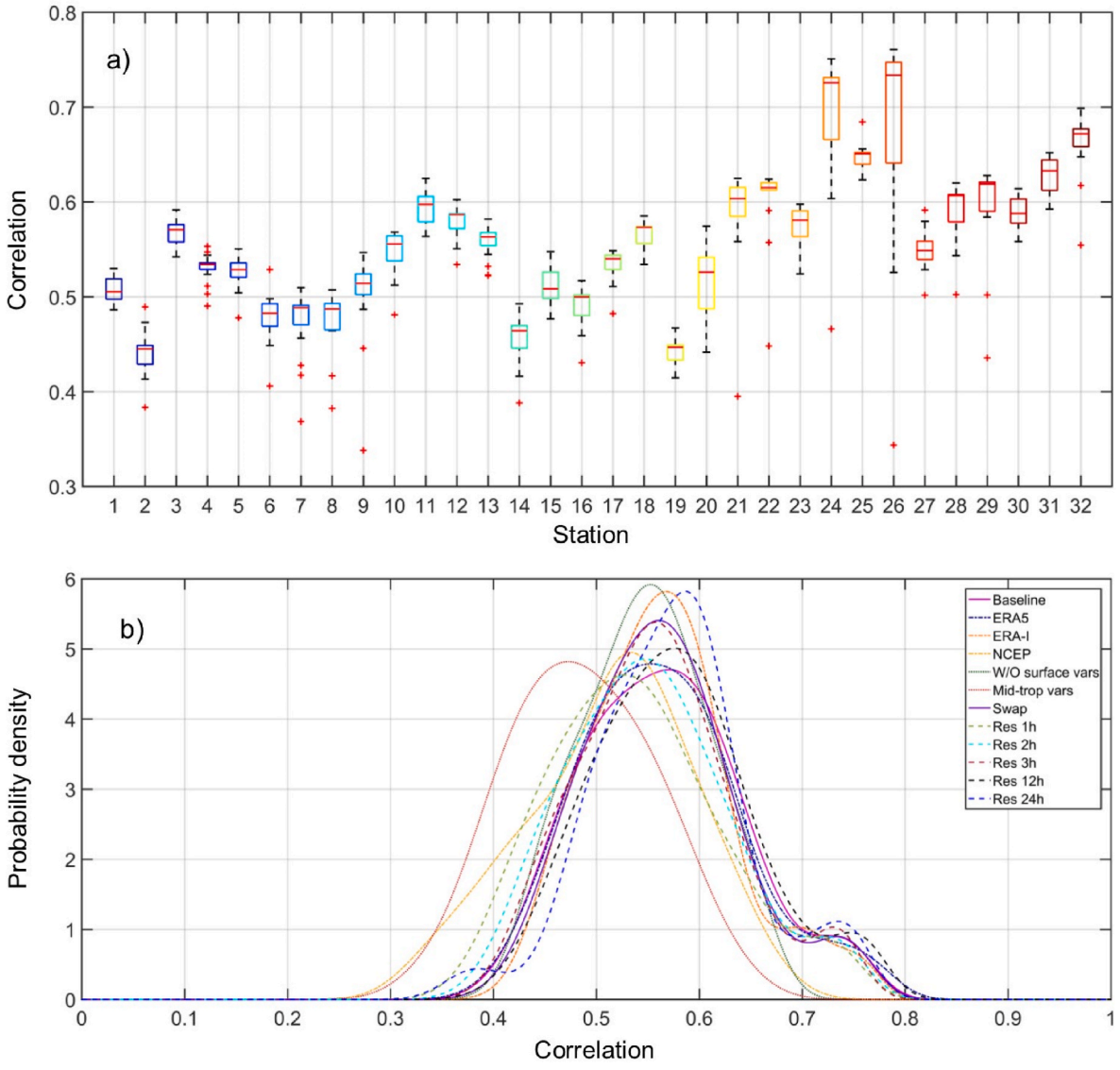

Fig. 6. a) Box plot of model-NSLOTT correlations for all sensitivity variations (except experiment THR) at each station, and b) potential density function of correlations computed across all stations for each sensitivity variation (except experiments GAP, LEN and THR).

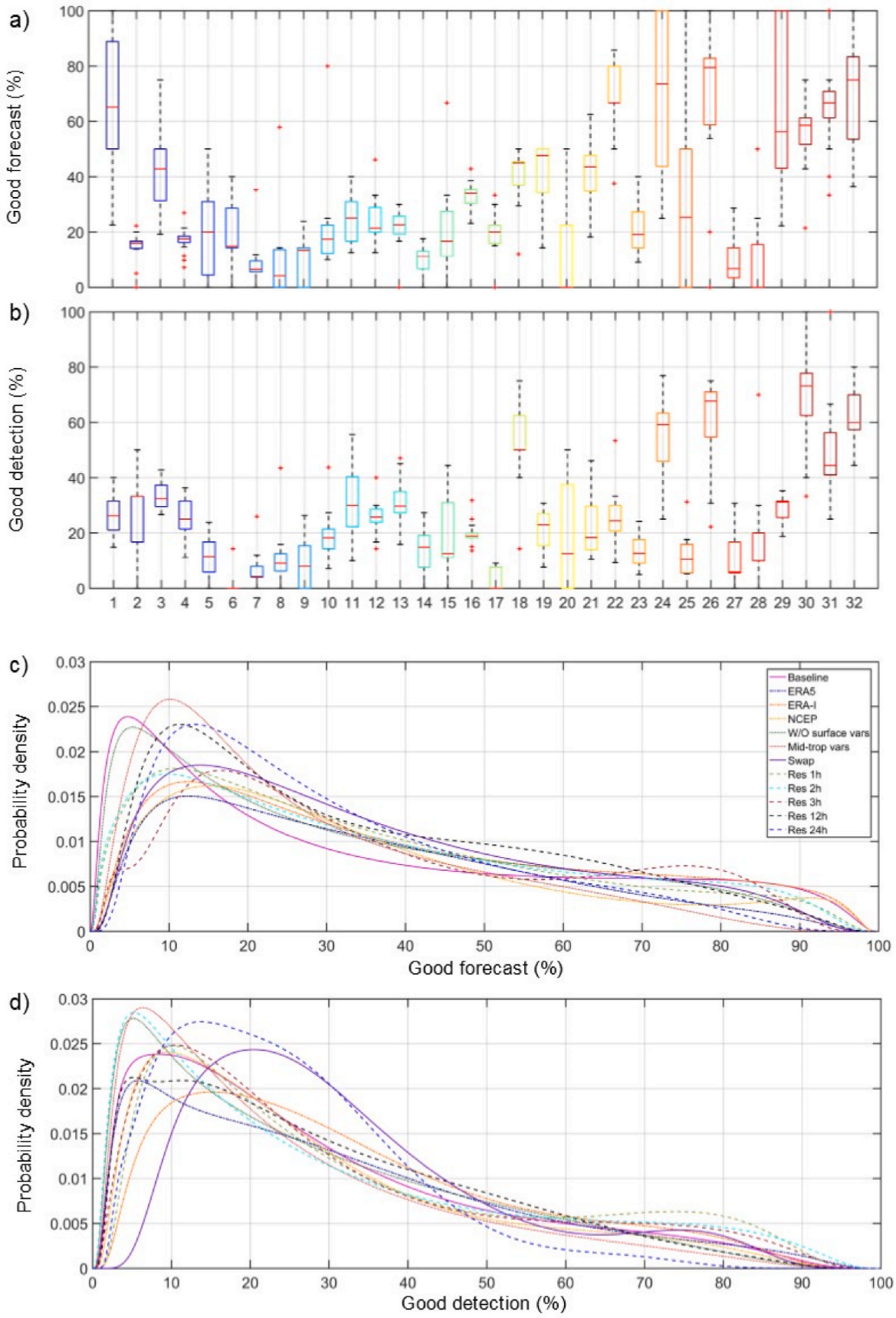

Fig. 7. Box plot of a) *good forecast* and b) *good detection* for all sensitivity variations (except experiment THR) at each station, and potential density function of c) *good forecast* and d) *good detection* computed across all stations for each sensitivity variation (except experiments GAP, LEN and THR).

good forecast are higher than of *good detection*. Moreover, it is evident that the stations with higher success in forecasting also demonstrate higher success in detecting events. When comparing probability density functions of *good forecast* and *good detection* for different sensitivity variations (Fig. 7c and d), a varying spread of the distribution, median values and tail weight are observed. The model constructed solely with mid-troposphere variables again shows the lowest median among all models. A heavier tail in the distribution, i.e., higher number of stations with the highest values of performance indicators, is observed for models with higher temporal resolutions. These models also show higher median values of *good forecast* (31.7 % for 1-h model and 30 % for 2-h model) compared to models with lower resolutions (23.4 % for 12-h model and 25 % for 24-h model).

Fig. 8 presents the probability density functions of model-to-NSLOTT bias for three different parts of their distributions: (i) up to the 95th percentile, (ii) between the 95th and 99th percentile, and (iii) over the 99th percentile. The bias of the first group is roughly spread around zero for all sensitivity variations, with a relatively weak tendency of the models to overestimate or underestimate NSLOTT values by up to 1 cm. However, the models tend to consistently underestimate NSLOTT values when considering the extreme values. For NSLOTTs between 95th and 99th percentile, the model shows the median underestimation of 4.1 cm,

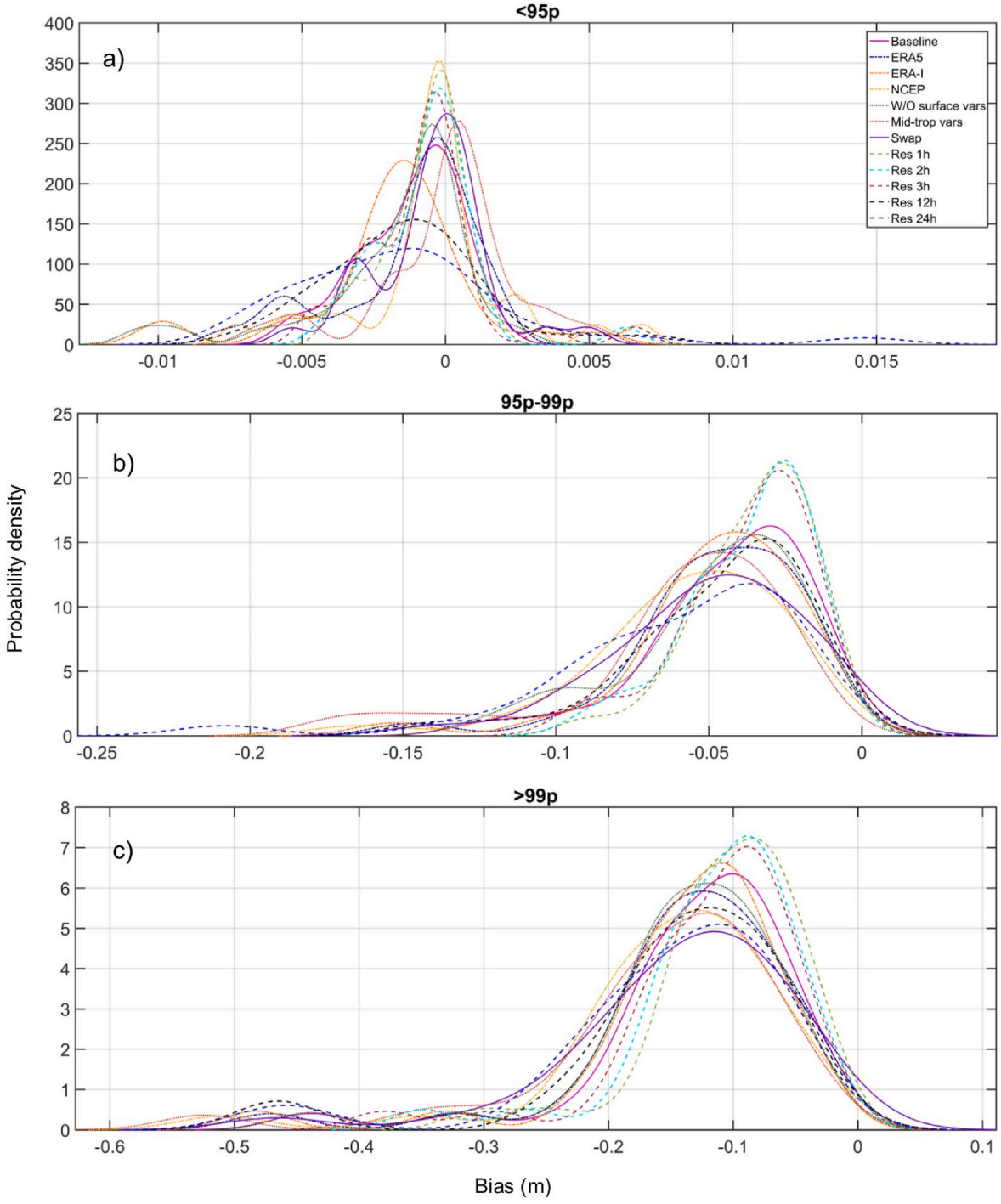

Fig. 8. Potential density function of model-NSLOTT bias computed across all stations for each sensitivity variation (except experiments GAP, LEN and THR) for the bias a) up to 95th percentile, b) between 95th and 99th percentile, and c) over 99th percentile of series.

while the median bias value is 11.7 cm for the NSLOTTs at the very tail of the distribution (>99th percentile). The median bias value decreases with an increase in temporal resolution, and the distribution becomes narrower. Additionally, the bias increases when fewer synoptic variables are included in the model, with the median value reaching up to 13.4 cm for the model that was constructed solely with mid-troposphere variables.

Furthermore, the stations were grouped into five different regions to detect distinct behaviours of the model in different parts of the Mediterranean (Fig. 1). The regions considered are: (i) Alboran Sea, (ii) Balearic Sea, (iii) French coast and Corsica, (iv) region of Sicily, and (v) eastern Adriatic Sea, each consisting of 3–10 stations. The region of Sicily, with only 3 stations included, exhibits the lowest median value of *good forecast* (11.1 %) and the second-lowest of *good detection* (18.2 %) (Fig. 9). The performance of the model is clearly better for the eastern Adriatic Sea region compared to other regions, with a median of *good forecast* of 64.6 % and a median of *good detection* of 50 %. The regions of the Alboran and Balearic Sea demonstrate fairly similar distributions for both parameters, especially for *good detection*, with quasi-equal values at the 25th, 50th and 75th percentiles. In contrast, the French coast and Corsica region, which includes the highest number of stations (10 stations), has a wider spread in values.

The stations were also subjectively categorized based on their geographical characteristics, such as the local bathymetry characteristics around a tide gauge, the properties of the shelf stretching in front of the coast, and including the orientation of the harbour. With regard to the local bathymetry characteristics around a tide gauge, two groups were formed: (i) *open-coast stations,* including stations that are situated outside harbours and not protected by islands, and (ii) *protected stations,* encompassing stations that are inside protected harbours or bays. Stations with just a few kilometres of shelf extending in front of the coast were classified into the *narrow shelf* group*,* while those with a long shelf off the coast, of up to a few tens of kilometres, are assigned to the *wide shelf* group. Lastly, the groups defining the major orientation of the bay or harbour in which the tide gauge is located are: (i) 0–90°, (ii) 90–180°, (iii) 180–270°, and (iv) 270–360°. Open-coast stations exhibit somewhat higher success in forecasting extreme events compared to the protected stations, with a median difference of approximately 15 %, while the average success in detecting extremes is nearly equal for both groups (Fig. 10). A similar trend is observed for the second categorization, where stations with the narrow shelf show a slightly higher median of *good forecast* (around 4 %), but the median values of *good detection* are fairly similar. The orientation of the harbour of bay presumably plays a significant role in the performance of the model. The highest results are obtained for stations located in harbours open towards the northwest (270–360°), with at least twice higher the median of both parameters compared to the other three groups.

4. Discussion

The aim of this paper was to examine the sensitivity of the synoptic index which quantifies the connection of NSLOTTs with atmospheric synoptic patterns. Such an index was initially developed for the meteotsunami 'hotspot' Ciutadella (Šepić et al., 2016), and later expanded for the entire globe (Zemunik et al., 2022b). Still, there are many questions related to this approach, like (i) if using limited set of prescribed synoptic variables is an appropriate way to maximize the connections, or (ii) are the estimates sensitive to products on which synoptic-based index is built, or (iii) which level of quality of NSLOTTs should be reached to get invariant estimates, and others. To tackle some of these questions, this study evaluated the synoptic index-based model performance under various modifications to its construction process, like using different reanalysis products, alterations in the NSLOTT time series, and adjustments to the thresholds determining extreme events. The analysis showed that horizontal resolutions of different atmospheric reanalyses (e.g., ERA5 vs. ERA-I) for input synoptic variables do not influence the model performance. A slight underperformance when using NCAR reanalysis might be a result of different choices (assimilated data, assimilation schemes, physics) during the construction of the reanalysis with respect to the ERA-I and ERA5 reanalysis which are both created by ECMWF. Therefore, the synoptic patterns associated with extreme NSLOTT values can be reproduced by coarser resolutions (O (100 km)) and does not require high-resolution atmospheric products (O (10 km)), given that the correlation between large-scale atmospheric patterns and extreme NSLOTTs is established at first place.

Among all studied sensitivity variations, the poorest performance is observed when only two synoptic variables - wind speed and relative humidity in the mid-troposphere - are utilized in the model. These

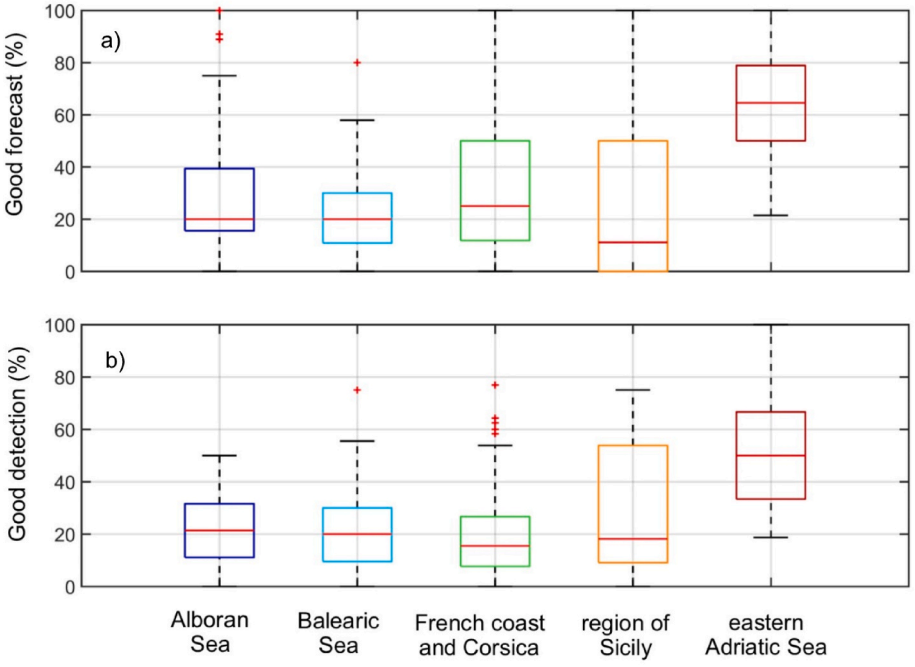

Fig. 9. Box plot of a) *good forecast* and b) *good detection* for all sensitivity variations (except experiment THR) in each region.

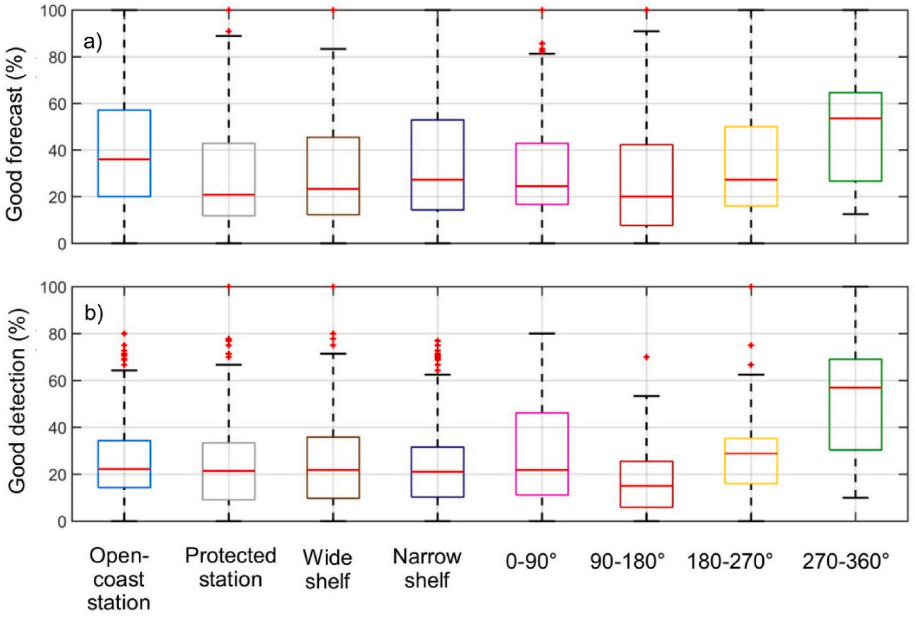

Fig. 10. Box plot of a) *good forecast* and b) *good detection* for all sensitivity variations (except experiment THR) in each group.

variables describe a specific mid-tropospheric setup observed commonly during meteotsunamis in the Mediterranean Sea (Šepić et al., 2015b), and are found to have globally the highest correlations with NSLOTTs (Vilibić and Šepić, 2017; Zemunik et al., 2022b). Precisely, meteotsunamigenic synoptic patterns in the Mediterranean are characterized by strong jet streams in an upper dynamically unstable layer, that acts as a lid to the lower stable layer with embedded atmospheric disturbance, keeping the energy of a disturbance towards the surface. This mechanism, known as wave-ducting (Lindzen and Tung, 1976; Monserrat and Thorpe, 1996), is shown to be crucial for meteotsunami formation (Zhang et al., 2003; Šepić et al., 2015a). However, our analysis show that using only these two variables – which are the easiest to reproduce, e.g., by climate models as being far from boundaries and discontinuities and as mostly resembling geostrophic balance – is not enough to get the best possible connection between NSLOTT and synoptic-based index. Incorporating additional synoptic variables into the model allows for a more detailed presentation of the overall synoptic situation, capturing other important processes, such as inflow of warm African air in the low troposphere and presence of weak surface cyclones, that often accompany Mediterranean meteotsunamis (Jansà et al., 2007; Šepić et al., 2015b; Pupić Vurilj et al., 2023).

Furthermore, the synoptic index-based model demonstrated consistency in performance, regardless of alterations in characteristics of the NSLOTT series. This robustness underscores the adaptability and versatility of the model to accommodate diverse data lengths, temporal resolutions and data gap amounts. However, different choices of the training and testing datasets revealed an impact on the model performance, in particular when the series are considered to be relatively short. This is generally the case for minute sea level series coming from the MISELA dataset in comparison to hourly sea level dataset frequently spanning several decades to a century, like the GESLA (Global Extreme Sea-Level Analysis) dataset (Haigh et al., 2023). Unfortunately, sea-level networks that have 1-min or higher resolutions are strongly dependent on the agency maintaining such a network, resulting in a great amount of poorly performing tide gauges that are not appropriate for research purposes (Zemunik et al., 2021b). Further, not all existing minute sea level datasets in the Mediterranean is available for the research. Conclusively, establishing global standards for sea-level monitoring at minute timescales and obeying FAIR (Findability, Accessibility, Interoperability, and Reuse; Wilkinson et al., 2016) principles in data sharing is a must for achieving better reproducibility of extreme NSLOTT events and meteotsunamis, including the creation of synoptic-index based or other models which connects synoptic patterns with NSLOTTs.

The model demonstrates higher performance at certain stations, while exhibiting lower performance at others, despite adjustments in its configuration. The varying performance of stations could be attributed to local atmospheric and oceanographic conditions, particularly in the Adriatic Sea. Meteotsunamigenic disturbances in this region often originate or being amplified southwest to west from meteotsunami hotspots, i.e., over the Apennines, where they may generate intense internal gravity waves that propagate over the Adriatic. A short distance from Apennines to the eastern Adriatic coast (~200–250 km) likely reduces energy dissipation and signal modulation. In contrast, regions like the Balearic Islands are much farther from major topographic features, such as the Atlas Mountains (~800 km from Menorca), leading to greater energy loss or modulation before disturbances reach tide gauges, thus introducing an additional uncertainty in forecast and detection analyses. Another limitation of this study is the absence of tide gauge data from the eastern Mediterranean, which may affect the generalizability of our findings across the entire region. The analyses were based solely on tide gauge records from the MISELA dataset, extended by approximately six years, without incorporating additional stations, apart from Ciutadella. In addition to extending of data series which are relatively short to develop a methodology for rare events, the future studies should include more stations to enhance spatial coverage and improve the robustness of analyses, ultimately strengthening the models aplicability.

Expectedly, the model exhibits higher performance when the criterion for classifying an event as extreme is relaxed. However, for site-specific research, one should be cautious about setting the threshold for distinguishing extreme events from non-extreme and moderate events (Denamiel et al., 2019a,b). This is especially important in warning systems for extreme sea levels, where each harbour has its own critical value of sea level that leads to flooding. In the absence of such criteria, we used different percentiles to define the threshold values and to evaluate the model performance. Namely, the model will fail if the model-reproduced extreme NSLOTT is slightly below the threshold, e.g., at 98.9 percentile, while the observed extreme NSLOTT is slightly above the threshold, e.g., at 99.1 percentile. Instructively, such a situation should be treated as a successful reproduction or forecast of the extreme NSLOTT event as the two values are almost the same. That was the exact reason for lowering the second percentile to 98 and even more in

sensitivity studies, to accommodate the uncertainty of the model performance.

Existing warning systems for meteotsunamis primarily rely on real-time monitoring of atmospheric conditions, and their comparison to historical patterns to issue alerts to at-risk coastal areas (Jansà and Ramis, 2021). While these systems have undoubtedly reduced damage and saved lives, they often face challenges in accurately predicting the onset and intensity of meteotsunamis due to the complex interactions between meteorological phenomena and coastal topography. In this study, we did not perform an assessment of applicability of our methodology in eventual warning systems, which are heavily dependent on success rates for good forecasts, i.e., quite large number of false alarms would discriminate its applicability, as known for a long time for tsunami early warning systems (Igarashi et al., 2011). Indeed, such testing might be a part of the future research. Integrating predictive modelling techniques, such as those based on specific combination of synoptic variables, holds immense promise for enhancing the effectiveness of warning systems. By leveraging models trained on historical data, it may be possible to forecast the likelihood and severity of meteotsunami events with greater accuracy and lead time (Vich and Romero, 2021). However, to create a more comprehensive product and transcend purely meteorological approach, a warning system should also engage other aspects, such as: (i) detection of tsunamigenic atmospheric disturbances recorded on microbarographs and/or radio-soundings, (ii) detection and tracking of indicative NSLOTTs on neighbouring tide gauges, and (iii) numerical simulation of NSLOTTs based on coupling of atmosphere-ocean numerical models (Vilibić et al., 2016). To this date, there is no warning system for meteotsunamis that encompasses all of these aspects.

Another application of the index-based model could be for climate projections of NSLOTTs and meteorological tsunamis. Recent climate studies have shed light on their increasing frequency and potential for significant coastal hazards (Bechle et al., 2016; Calafat and Marcos, 2020; Ruić et al., 2023). Utilizing predictive models based on synoptic variables might bring a significant advancement in climate projections of NSLOTTs, which are valuable for coastal communities and policy-makers seeking to develop adaptation strategies and improve coastal resilience. Denamiel et al. (2023) propose deriving future projections of meteotsunami hazard by using synoptic index to extract potential extreme episodes, which are further investigated through downscaling from an ensemble of global climate models. This approach should overcome the issue of large computing resources of long-term high--resolution climate models, by optimizing extractions to preselected episodes only.

Indeed, being built on multi-regression analysis, this model may further be improved by using pattern recognition and machine learning algorithms (Cheng et al., 2014; Leuenberger and Kanevski, 2015; Yu and Ma, 2021). Modern techniques might capture complex relationships and nonlinear dynamics of synoptic patterns, enhancing their predictive capabilities. Moreover, evaluating the model performance in this study is presumably oversimplified by the definitions of *good forecast* and *good detection*. Both parameters use a binary approach where events are classified as either non-extreme or extreme, based on a predefined threshold. Adopting a more nuanced approach that considers gradation of events could address this limitation. One such classification is proposed in the recent study by Lewis et al. (2024), which presents a novel five-level meteotsunami intensity index based on various wave characteristics and effects on the shoreline.

5. Conclusions

After the creation of the synoptic index by Zemunik et al. (2022b), this study further investigates the ability of the model, based on the synoptic patterns, to forecast and detect extreme NSLOTTs by varying its configuration. The key factors that influence model efficiency have been extensively analysed, leading to the following conclusions:

1. Stations that exhibit good performance for the baseline model, generally maintain good performance regardless of configuration changes, suggesting that the model performance heavily depends on the pre-established connection between the considered synoptic variables and extreme NSLOTTs.
2. Conversely, stations exhibiting poorer performance in the baseline model typically do not show significant improvement despite configuration adjustments.
3. Stations that exhibit good performance for the baseline model show larger fluctuations in performance when varying model configuration, compared to those with poorer model performance.
4. The models success in forecasting extreme events is slightly higher than its success in detecting extreme events, indicating a higher tendency to miss observed extremes rather than to produce false alarms. Additionally, stations excelling in forecasting extreme NSLOTTs also excel in detecting them.
5. The best performance is achieved for the stations in the eastern Adriatic Sea, suggesting that the model is most effectively tailored for extreme NSLOTT occurrences in this region of the Mediterranean.

In summary, this study not only dissects the important factors that determine the efficiency and success of the synoptic index-based model, but also underscores the necessity for continuous refinement and validation of predictive models in addressing the evolving challenges posed by extreme NSLOTTs. As climate change exacerbates the frequency and intensity of coastal hazards, the need for accurate and reliable prediction models for all extremes is of the outmost importance for coastal communities.

CRediT authorship contribution statement

P. Zemunik Selak: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **I. Vilibić:** Writing – review & editing, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **C. Denamiel:** Writing – review & editing, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **P. Pranić:** Writing – review & editing, Software, Methodology, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The work has been conducted with the support of the Croatian Science Foundation projects GLOMETS (Grant IP-2022-10-3064) and C3PO (Grant IP-2022-9139), HORIZON EUROHPC JU project ChEESE-2P (Grant 101093038), as well as the project KLIMADRIA funded by the European Union - NextGenerationEU. We are grateful to the hundreds of data providers and thousands of researchers, engineers and technicians engaged in the maintenance of tide-gauge stations used in this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wace.2025.100775.

Data availability

Data will be made available on request.

References

Bechle, A.J., Wu, C.H., Kristovich, D.A.R., Anderson, E.J., Schwab, D.J., Rabinovich, A. B., 2016. Meteotsunamis in the laurentian great lakes. Sci. Rep. 6, 37832. https://doi.org/10.1038/srep37832.

Belušić, D., Strelec Mahović, N., 2009. Detecting and following atmospheric disturbances with a potential to generate meteotsunamis in the Adriatic. Phys. Chem. Earth 34, 918–927. https://doi.org/10.1016/j.pce.2009.08.009.

Calafat, F.M., Marcos, M., 2020. Probabilistic reanalysis of storm surge extremes in Europe. Proceedings of the National Academy of Sciences of the United States of America 117, 1877–1883. https://doi.org/10.1073/pnas.1913049117.

Cheng, L., AghaKouchak, A., Gilleland, E., Katz, R.W., 2014. Non-stationary extreme value analysis in a changing climate. Clim. Change 127, 353–369. https://doi.org/10.1007/s10584-014-1254-5.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137, 553–597. https://doi.org/10.1002/qj.828.

Denamiel, C., Šepić, J., Huan, X., Bolzer, C., Vilibić, I., 2019a. Stochastic surrogate model for meteotsunami early warning system in the eastern Adriatic Sea. J. Geophys. Res.: Oceans 124, 8485–8499. https://doi.org/10.1029/2019JC015574.

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

Denamiel, C., Belušić, D., Zemunik, P., Vilibić, I., 2023. Climate projections of meteotsunami hazards. Front. Mar. Sci. 10, 1167863. https://doi.org/10.3389/fmars.2023.1167863.

Greenspan, H.P., 1956. The generation of edge waves by moving pressure distributions. J. Fluid Mech. 1, 574–592. https://doi.org/10.1017/S002211205600038X.

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

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., Thépaut, J.-N., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146, 1999–2049. https://doi.org/10.1002/qj.3803.

Igarashi, Y., Kong, L., Yamamoto, M., McCreery, C.S., 2011. Anatomy of historical tsunamis: lessons learned for tsunami warning. Pure Appl. Geophys. 168, 2043–2063. https://doi.org/10.1007/s00024-011-0287-1.

Jansà, A., Monserrat, S., Gomis, D., 2007. The rissaga of 15 June 2006 in Ciutadella (Menorca), a meteorological tsunami. Adv. Geosci. 12, 1–4. https://doi.org/10.5194/adgeo-12-1-2007.

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

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J.J., Fiorino, M., Potter, G. L., 2002. NCEP–DOE AMIP-II reanalysis (R-2). Bull. Am. Meteorol. Soc. 83, 1631–1644. https://doi.org/10.1175/BAMS-83-11-1631.

Leuenberger, M., Kanevski, M., 2015. Extreme Learning Machines for spatial environmental data. Comput. Geosci. 85 (Part B), 64–73. https://doi.org/10.1016/j.cageo.2015.06.020.

Lewis, C., Smyth, T., Neumann, J., Cloke, H., 2024. Proposal for a new meteotsunami intensity index. Nat. Hazards Earth Syst. Sci. 24, 121–131. https://doi.org/10.5194/nhess-24-121-2024.

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

Monserrat, S., Thorpe, A.J., 1996. Use of ducting theory in an observed case of gravity waves. J. Atmos. Sci. 53, 1724–1736. https://doi.org/10.1175/1520-0469(1996)053<1724:UODTIA>2.0.CO;2.

Monserrat, S., Vilibić, I., Rabinovich, A.B., 2006. Meteotsunamis: atmospherically induced destructive ocean waves in the tsunami frequency band. Nat. Hazards Earth Syst. Sci. 6 (6), 1035–1051. https://doi.org/10.5194/nhess-6-1035-2006.

Pattiaratchi, C.B., Wijeratne, E.M.S., 2015. Are meteotsunamis an underrated hazard? Philosophical Transactions of the Royal Society A 373, 20140377. https://doi.org/10.1098/rsta.2014.0377.

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.Á., 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., Zodiatis, G., 2022. Coastal sea level monitoring in the Mediterranean and Black seas. Ocean Sci. 18, 997–1053. https://doi.org/10.5194/os-18-997-2022.

Proudman, J., 1929. The effects on the sea of changes in atmospheric pressure. Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society 2 (4), 197–209. https://doi.org/10.1111/j.1365-246x.1929.tb05408.x.

Pupić Vurilj, M., Brnac, T., Ruić, K., Šepić, J., Balić, M., 2023. Mediterranean meteotsunamis of May 2021 and June 2022: observations, data analysis and synoptic background. Geofizika 40, 179–205. https://doi.org/10.15233/gfz.2023.40.8.

Rabinovich, A.B., 2020. Twenty-seven years of progress in the science of meteorological tsunamis following the 1992 Daytona Beach event. Pure Appl. Geophys. 177, 1193–1230. https://doi.org/10.1007/s00024-019-02349-3.

Ramis, C., Jansà, A., 1983. Condiciones meteorológicas simultáneas a la aparición de oscilaciones del nivel del mar de amplitud extraordinaria en el Mediterráneo occidental. Rev. Geofisc. 39, 35–42 (in Spanish).

Renault, L., Vizoso, G., Jansà, A., Wilkin, J., Tintoré, J., 2011. Toward the predictability of meteotsunamis in the Balearic Sea using regional nested atmosphere and ocean models. Geophys. Res. Lett. 38, L10601. https://doi.org/10.1029/2011gl047361.

Ruić, K., Šepić, J., Mlinar, M., Međugorec, I., 2023. Contribution of high-frequency (T < 2 h) sea level oscillations to the Adriatic sea level maxima. Nat. Hazards 116, 1–31. https://doi.org/10.1007/s11069-023-05834-0.

Šepić, J., Vilibić, I., Lafon, A., Macheboeuf, L., Ivanović, Z., 2015b. High-frequency sea level oscillations in the Mediterranean and their connection to synoptic patterns. Prog. Oceanogr. 137, 284–298. https://doi.org/10.1016/j.pocean.2015.07.005.

Šepić, J., Vilibić, I., Monserrat, S., 2009. Teleconnections between the adriatic and the balearic meteotsunamis. Phys. Chem. Earth 34, 928–937. https://doi.org/10.1016/j.pce.2009.08.007.

Šepić, J., Vilibić, I., Monserrat, S., 2016. Quantifying the probability of meteotsunami occurrence from synoptic atmospheric patterns. Geophys. Res. Lett. 43, 10377–10384. https://doi.org/10.1002/2016GL070754.

Šepić, J., Vilibić, I., Rabinovich, A.B., Monserrat, S., 2015a. Widespread tsunami-like waves of 23-27 June in the Mediterranean and Black Seas generated by high-altitude atmospheric forcing. Sci. Rep. 5 (1), 11682. https://doi.org/10.1038/srep11682.

Vich, M.-M., Romero, R., 2021. Forecasting meteotsunamis with neural networks: the case of Ciutadella harbour (Balearic Islands). Nat. Hazards 106, 1299–1314. https://doi.org/10.1007/s11069-020-04041-5.

Vilibić, I., Rabinovich, A.B., Anderson, E.J., 2021. Special issue on the global perspective on meteotsunami science: editorial. Nat. Hazards 106, 1087–1104. https://doi.org/10.1007/s11069-021-04679-9.

Vilibić, I., Šepić, J., 2009. Destructive meteotsunamis along the eastern Adriatic coast: overview. Phys. Chem. Earth 34, 904–917. https://doi.org/10.1016/j.pce.2009.08.004.

Vilibić, I., Šepić, J., 2017. Global mapping of nonseismic sea level oscillations at tsunami timescales. Sci. Rep. 7, 40818. https://doi.org/10.1038/srep40818.

Vilibić, I., Šepić, J., Dunić, N., Sevault, F., Monserrat, S., Jordà, G., 2018. Proxy-based assessment of strength and frequency of meteotsunamis in future climate. Geophys. Res. Lett. 45, 10501–10508. https://doi.org/10.1029/2018GL079566.

Vilibić, I., Šepić, J., Rabinovich, A.B., Monserrat, S., 2016. Modern approaches in meteotsunami research and early warning. Front. Mar. Sci. 3, 57. https://doi.org/10.3389/fmars.2016.00057.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3, 160018. https://doi.org/10.1038/sdata.2016.18.

Williams, D.A., Schultz, D.M., Horsburgh, K.J., Hughes, C.W., 2021. An 8-yr meteotsunami climatology across Northwest Europe: 2010–17. J. Phys. Oceanogr. 1145–1161. https://doi.org/10.1175/JPO-D-20-0175.1.

Yu, S., Ma, J., 2021. Deep learning for geophysics: current and future trends. Rev. Geophys. 59 (3). https://doi.org/10.1029/2021RG000742.

Zemunik, P., Bonanno, A., Mazzola, S., Giacalone, G., Fontana, I., Genovese, S., Basilone, G., Candela, J., Šepić, J., Vilibić, I., Aronica, S., 2021c. Observing meteotsunamis ("Marrobbio") on the southwestern coast of Sicily. Nat. Hazards 106, 1337–1363. https://doi.org/10.1007/s11069-020-04303-2.

Zemunik, P., Denamiel, C., Šepić, J., Vilibić, I., 2022a. High-frequency sea-level analysis: global distributions. Global Planet. Change 210, 103775. https://doi.org/10.1016/j.gloplacha.2022.103775.

Zemunik, P., Denamiel, C., Williams, J., Vilibić, I., 2022b. High-frequency sea-level extremes: global correlations to synoptic atmospheric patterns. Weather Clim. Extrem. 38, 100516. https://doi.org/10.1016/j.wace.2022.100516.

Zemunik, P., Vilibić, I., Šepić, J., Pellikka, H., Ćatipović, L., 2021a. MISELA: minute Sea-Level analysis. Marine Data Archive. https://doi.org/10.14284/456.

Zemunik, P., Vilibić, I., Šepić, J., Pellikka, H., Ćatipović, L., 2021b. MISELA: 1-minute sea-level analysis global dataset. Earth Syst. Sci. Data 13, 4121–4132. https://doi.org/10.5194/essd-13-4121-2021.

Zhang, F., Koch, S.E., Kaplan, M.L., 2003. Numerical simulations of a large-amplitude mesoscale gravity event. Meteorol. Atmos. Phys. 84, 199–216. https://doi.org/10.1007/s00703-002-0594-2.