

Verification of an ensemble prediction system for storm surge forecast in the Adriatic Sea

Riccardo Mel · Piero Lionello

Received: 5 February 2014 / Accepted: 6 October 2014 / Published online: 21 October 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract In the Adriatic Sea, storm surges present a significant threat to Venice and to the flat coastal areas of the northern coast of the basin. Sea level forecast is of paramount importance for the management of daily activities and for operating the movable barriers that are presently being built for the protection of the city. In this paper, an EPS (ensemble prediction system) for operational forecasting of storm surge in the northern Adriatic Sea is presented and applied to a 3-month-long period (October–December 2010). The sea level EPS is based on the HYPSE (hydrostatic Padua Sea elevation) model, which is a standard single-layer nonlinear shallow water model, whose forcings (mean sea level pressure and surface wind fields) are provided by the ensemble members of the ECMWF (European Center for Medium-Range Weather Forecasts) EPS. Results are verified against observations at five tide gauges located along the Croatian and Italian coasts of the Adriatic Sea. Forecast uncertainty increases with the predicted value of the storm surge and with the forecast lead time. The EMF (ensemble mean forecast) provided by the EPS has a rms (root mean square) error lower than the DF (deterministic forecast), especially for short (up to 3 days) lead times. Uncertainty for short lead times of the forecast and for small storm surges is mainly caused by uncertainty of the

initial condition of the hydrodynamical model. Uncertainty for large lead times and large storm surges is mainly caused by uncertainty in the meteorological forcings. The EPS spread increases with the rms error of the forecast. For large lead times the EPS spread and the forecast error substantially coincide. However, the EPS spread in this study, which does not account for uncertainty in the initial condition, underestimates the error during the early part of the forecast and for small storm surge values. On the contrary, it overestimates the rms error for large surge values. The PF (probability forecast) of the EPS has a clear skill in predicting the actual probability distribution of sea level, and it outperforms simple “dressed” PF methods. A probability estimate based on the single DF is shown to be inadequate. However, a PF obtained with a prescribed Gaussian distribution and centered on the DF value performs very similarly to the EPS-based PF.

Keywords Venice surge forecast · Ensemble prediction system · Probabilistic forecast · Flood management

1 Introduction

Efficient managing of coastal defenses, such as movable dams, and delivery of flood warning to alert population are critically depending on the quality of the information that is delivered by a storm surge forecast system. A recent development is the extension of existing storm surge operational forecasts to ensemble prediction systems (EPS), whose main advantage is the production of information on forecast uncertainty (Flowerdew et al. 2012; De Vries 2009; Di Liberto et al. 2011; Mel and Lionello 2014; Mel et al. 2014). Storm surge ensemble predictions are already operationally used by the UK Environment Agency, for issuing coastal flood warnings in England and Wales and by the Storm Surge Warning Service (SVSD) of Rijkswaterstaat, the Public Works and

Responsible Editor: Kevin Horsburgh

This article is part of the Topical Collection on the *13th International Workshop on Wave Hindcasting and Forecasting in Banff, Alberta, Canada October 27–November 1, 2013*

R. Mel (✉)
DICEA (Dipartimento di Ingegneria Civile, Edile ed Ambientale),
University of Padova, Padova, Italy
e-mail: riccardo.mel@dicea.unipd.it

P. Lionello
DISTEBA (Dipartimento Scienze e Tecnologie Biologiche e
Ambientali), University of Salento, Lecce, Italy
e-mail: piero.lionello@unisalento.it

Water Management Authority in The Netherlands, for the Dutch coast.

In the Adriatic Sea, floods are a recurrent problem for the city of Venice. Their frequency has increased since the 1950s (Battistin and Canestrelli 2006; Lionello et al. 2012) due to increase in sea level (SL) relative to the land, which is caused by a combination of mean SL increase and local subsidence (see Lionello 2012 for a discussion). The hazard posed by the Adriatic Sea storm surges has been shown by the event of 4 November 1966, which produced severe damage and relevant economic losses (see De Zolt et al. 2006 for a description of its evolution).

A system of movable dams (called MOSE; *Modulo Sperimentale Elettromeccanico*, experimental electromechanic module) is presently been built across the three inlets of the Venetian Lagoon and will be operated to prevent the flooding of the city center during storm surges. An accurate forecast of SL and of its uncertainty is extremely useful for operating this movable dam system efficiently, which requires the decision on lifting the barriers to be taken about 8 h before water reaches 84 cm above the present mean SL (Eprim et al. 2005).

A tide forecast center (ICPSM—Istituto Centro Previsione e Segnalazione Maree) has been established by the town council of Venice in 1980 and operates a set of models for SL prediction (Massalin et al. 2007). These models are

- BIGSUMDP, which is the evolution (Tosoni and Canestrelli 2011) of a linear statistical autoregressive model (Tomasin 1972). BIGSUMDP is calibrated using observed sea level time series and predicts the water level in the lagoon using observed local SL and observed mean sea level pressure (MSLP) at stations along the coast of the Adriatic Sea.
- SHYFEM (shallow water hydrodynamic finite element model, Umgiesser et al. 2004), which is a hydrodynamical model based on the finite element method. SHYFEM integrates the shallow water equations and computes the evolution of current and sea level from a sequence of MSLP and surface wind atmospheric fields.
- HYPSEAM (hydrostatic Padua Sea elevation and adjoint model, Lionello et al. 2006), which is a hydrodynamical model based on the finite difference method and forced by the same meteorological fields as SHYFEM. HYPSEAM adopts an orthogonal grid and includes a data assimilation module based on the adjoint method. However, since this module is not used in this study, the acronym HYPSE is used in this manuscript.

In their present implementation at ICPSM, SHYFEM and HYPSEAM use the meteorological fields provided by the ECMWF prediction system.

This study describes the results of an operational implementation of an EPS (ensemble prediction system), which could be used as a further tool by ICPSM or other agencies for providing more information on SL evolution. This EPS follows the approach of Flowerdew et al. (2009, 2010, 2012), and it has previously been described in Mel and Lionello 2014, where it has been applied to ten individual storm surge events focusing on the prediction of the peak values. Mel and Lionello 2014 have shown that storm surge peaks correspond to maxima of uncertainty in the prediction (meaning that in correspondence with them, the likelihood of significant SL errors is largest), that such uncertainty increases linearly with the forecast lead time and it is linked to the uncertainty of the forcing meteorological fields. In relation to storm surge peak values, Mel and Lionello 2014 have further shown that the error of the EMF (ensemble mean forecast, meaning the mean of all members of the EPS) with respect to tide gauge observations is correlated with the EPS spread and that the EMF is more robust than a single deterministic forecast (DF) (meaning that its error is consistently smaller than the error of DF as the lead time of the forecast varies), though DF is based on meteorological forcings at higher resolution than EMF. Mel et al. 2014 have shown that it is possible to estimate forecast uncertainty via a linear combination of suitable meteorological variances, directly extracted from the EPS members, in order to reduce the computational cost for real-time application.

This new study contains analysis and statistics that are based on a 3-month-long period, during which EPS has been used imitating the operational practice. The analysis is not focused only on peak values but considers all hourly values. The aim is to show that the ensemble spread is a reliable indicator of the uncertainty associated with large surge events and that the EPS provides a skilled probabilistic forecast for the Adriatic Sea SL with a lead time sufficient for operating MOSE and warning the population. Further, the utility of complementing the hydrodynamic model single prediction (called “deterministic forecast”, DF, in the rest of this manuscript) with an EPS probabilistic forecast is investigated and the possibility of representation of the forecast uncertainty with algorithms that are computationally cheaper than the EPS is discussed.

Mel and Lionello (2014) discuss the dynamics of SL in the Adriatic Sea and of the meteorological forcings causing it. An essential point is the specificity of the Adriatic Sea SL dynamics, where SL peaks are caused not only by astronomical tide and storm surge but also by seiches (e.g., see Lionello et al. 2005), which are free oscillations of SL with fundamental periods of about 22 (main longitudinal seiche wave) and 11 h (transverse seiche wave). Seiches, after being triggered by a storm surge event, diminish slowly in amplitude, persisting for several cycles. A correct seiche forecast depends strongly on the correct timing and level of the initial storm

surge peak prediction, whose wrong forecast introduces in the simulation an error that will persist (and spoil the SL forecast) for several days.

The EPS is a consolidated tool for probabilistic weather prediction (Leutbecher and Palmer 2008). Its first implementation has been operational at ECMWF since 1992 (see Molteni et al. 1996). The original implementation of the EPS consists of a set of different forecasts based on a set of different atmospheric initial conditions, which are designed to represent the uncertainties inherent in the operational analysis. These initial differences are specified on the basis of the singular vector technique (Buizza and Palmer 1995) and are designed to amplify rapidly in time so that forecasts will deviate substantially from each other. This follows the basic idea of Lorenz (1965), who explored this effect for a simple model of weather (a two-dimensional convection system with three variables) with a crude perturbation design based on truncation error. The ensemble of different realizations allows the estimation of the probability distribution function of forecast states and provides a practical tool for estimating how initial uncertainties affect the forecast. The EPS approach has been successively extended to include also perturbed forecasts representing uncertainty on model parameters (Buizza et al. 1999). The approach of Flowerdew et al. (2009, 2010, 2012) uses a meteorological EPS (namely the resulting MSLP and surface wind fields) for forcing a storm surge model and producing an ensemble prediction of SL. In practice, each member of the EPS is used for obtaining a corresponding forecast of SL, from which a SL probabilistic prediction is computed.

The paper is organized in the following way. “Section 2” describes the shallow water model, the EPS method, the forecast experiment, data, and statistics used for forecast verification. “Section 3” describes the results of the EPS for an interesting event within the 3-month analyzed period. “Section 4” discusses how the ensemble spread represents the uncertainty of the forecast and its dependence on the lead time. “Section 5” discusses whether the EPS distribution describes correctly the possible conditions of SL in the Adriatic Sea. “Section 6” summarizes the conclusions of this study.

2 Data and methods

2.1 The hydrodynamical model

This study follows the practice of ICPSM (e.g., Massalin et al. 2007), which is to compute separately the astronomical tide and add it to the hydrodynamical model results for obtaining the actual SL. This is justified because of the modest tidal range on the Adriatic Sea (about 1 m in Venice, Trieste, and Rovinj and about half meter in Split and Dubrovnik), which makes nonlinear interactions with the surge negligible. All data presented here (both for observations and models)

consider exclusively the superposition of storm surge and seiches (without astronomical tide), which is called surge residual (SR) in this paper.

The SR EPS forecasts are carried out using the HYPSE model, which is HYPSEAM, without the data assimilation module. The version of HYPSE used adopts a latitude–longitude mesh grid of variable size, which has the minimum grid step (0.03°) in the northern part of the Adriatic Sea, from where grid step increases with a 1.01 factor in both latitude and longitude (in practice, resolution varies in the range from 3.3 to 7 km). This grid has been shown to produce more accurate results with respect to other grids (Lionello et al. 2005), and it has been used in Mel and Lionello 2014. This study does not explicitly account for the effects of changes of SL in the Ionian Sea on the Adriatic SL and of the total mass of water inside the Adriatic basin during each individual forecast. Further, it does not account for steric effects on SL. In fact, the model domain contains a unique open boundary (corresponding to the Otranto Strait connecting the Adriatic Sea to the Ionian Sea) across which sea level is kept fixed at its mean value during the forecast. This limitation is compensated by a bias removal technique (as described in Mel and Lionello 2014), which add to the SR prediction the effect of long term (several days to month) variability of the mean sea level in the Adriatic Sea. This correction is obtained by subtracting from the forecast the difference between the SR value at the beginning of the forecast and the value obtained by linear interpolation of the observed hourly SR data during the previous day. Figure 1 shows HYPSE domain and the position of tide gauges used in this study.

2.2 The EPS method for SR forecast

The SR-EPS uses the ECMWF meteorological EPS to force a hydrodynamic model and producing corresponding EPS forecast in the Adriatic Sea. Every 12 h, the 50 perturbed forecasts of the ECMWF-EPS are used for forcing HYPSE and

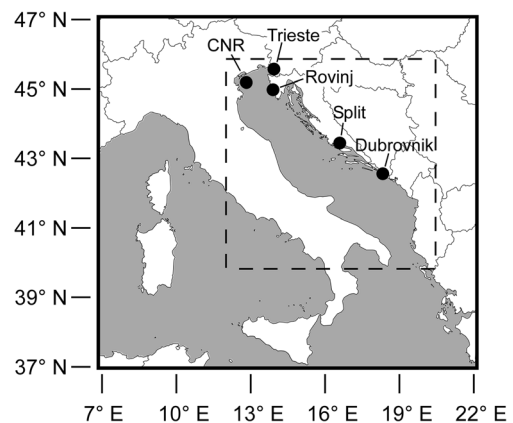


Fig. 1 The Adriatic Sea and the HYPSE domain with the locations of the tide gauges used in this study: the ISMAR-CNR platform, Trieste, Rovinj, Split, and Dubrovnik

producing a corresponding 50 SR perturbed forecasts, whose results differ only because the forcing meteorological fields are different. This approach is meant to represent the SR forecast uncertainty that is determined by errors or uncertainty in the meteorological forcings. It neglects modeling errors due to inaccuracy of the hydrodynamical model and errors caused by the uncertainty in the initial condition of the SR forecast. Modeling errors include insufficient spatial resolution, truncation errors in the dynamical equations, approximation errors to solve them, ad hoc parameterization, bias in the frequency of initialization, average, and coding errors, which produce uncertainty in the SR evolution. Errors in the ocean initial condition of the SR forecast may include wrong timing or amplitude of the storm surge and/or existing seiches. Because of the linear and weakly dissipative nature of the SR oscillations in the Adriatic Sea, this source of uncertainty has a peculiar oscillatory and slowly decaying behavior (see discussion in section 3).

2.3 The SR forecast

This study simulates the operational use of the SR EPS during the last 3 months (92 days) of the year 2010. Forcing of HYPSE is by three different sets of 3-hourly ECMWF 10-m wind and MSLP fields: the high resolution meteorological forecast (deterministic forecast, DF), the control run forecast (CRF, that differs from the DF only for its lower resolution), and the 50 ensemble members of the ECMWF EPS, which is a total of 52 sequences of meteorological fields representing 52 different evolutions of the weather. The resolution of DF fields is T1279 and resolution of both CRF and ECMWF EPS fields is T639. The 10-m wind and MSLP fields have been downloaded at 0.125° (DF) and 0.25° (CRF and EPS) and linearly interpolated to the HYPSE grid (which is the same for all simulations). The corresponding 52 SR simulations of HYPSE are denoted SR DF, CRF, and EPS. The mean of the 50 SR EPS simulations is called SR EMF.

This study is based on $92 \times 2 \times 52$ simulations of HYPSE. In the simulated operational prediction stream, the 52 HYPSE runs are carried out twice per day. They provide a set of 6-day forecasts forced by the 52 different ECMWF forecasts launched at 00 and 12UTC. Each HYPSE run is initialized by a 10-day analysis run (which is identical in all 52 simulations with the same initial date), where HYPSE is driven by the high resolution (T1279) ECMWF analysis so that the initial condition and the version of the HYPSE model are the same in all 52 SR predictions.

The EPS is the ideal tool for proving a probability forecast (PF) of occurrence of a predefined event. Given a threshold level h , the event H is defined as the SR exceeding h for at least one hourly step within a 12-h interval, where a set of 12 intervals is specified covering the whole time range from the beginning to the end of each forecast, which is 144 h long. The

probability $p_k(h)$ of the EPF (ensemble probability forecast) is given by the fraction of EPS members above the h threshold.

The SR predictions are complemented with a 92-day-long hindcast (initialized with a previous 10-day-long analysis) covering exactly the same period as the forecasts, in which the HYPSE model has been driven by the ECMWF analysis.

2.4 SR forecast validation

Hourly SR data computed from the time series of five tide gauges along the coast of the Adriatic Sea (Fig. 1) are available for the analysis of the quality of the SR prediction: the platform ISMAR-CNR, located 15 km offshore the Venice lagoon, and the tide gauges in the harbors of Trieste, Rovinj, Split and Dubrovnik. Instrumental errors of these hourly data are negligible and not considered in our analysis. Further, though it might be argued that actual peak values are underestimated in hourly observations, the smoothness of the model and observed time series indicate that this is not a relevant issue. Statistics include only time steps when observations at all tide gauges are available for all lead-time forecasts and include

- The SR EPS spread, which is defined as the standard deviation of the 50+1 ensemble members (including the CRF) around the EMF. It allows, for each time and location, the SR forecast uncertainty to be computed.
- The SR forecast rms (root mean square) error with respect to tide gauges, which is computed for DF, EMF, CRF, and for all the perturbed ensemble members, by comparing the SR forecast hourly time step with the corresponding observations.
- The SR forecast rms error with respect to the hindcast at the five tide gauges, that is model verification, that is carried out by sampling the hindcast.
- The Brier score (Brier 1950) to measure the accuracy of the probability estimated by the EPS. The Brier score $BS(h)$ is computed referring to a set of six threshold levels h ($h=0.0, 0.1, 0.2, 0.3, 0.4, 0.5$):

$$BS(h) = \frac{1}{N} \sum_{k=1}^N (p_k(h) - o_k(h))^2.$$

Here, N is the number of events, $p_k(h)$ and $o_k(h)$ are the probability of a surge higher than h according to the EPS and to observations, respectively, with $o_k(h)=1$ if the observed level is less or equal to h , $o_k(h)=0$ otherwise. $BS(h)$ varies from 0 (perfect prediction) to 1 (the prediction always fails)

- The Brier skill score, (BSS, Wilks 2006). The $BSS(h)$ measures the improvement of the EPS BS with respect to the $BS_{ref}(h)$ of unskilled standard forecast (DF), and it has a

range from $-\infty$ to 1 (positive values indicate that the forecast is more accurate than the unskilled standard forecast):

$$BSS(h) = 1 - \frac{BS(h)}{BS_{ref}(h)}$$

The $BS_{ref}(h)$ is obtained by substituting in the $BS(h)$ formula the forecast probability with the observed probability $\bar{o}(h)$, computed on the basis of a 30-year-long time series at the ISMAR-CNR platform (1981–2010) and 2-years long (2009–2010) at the other four gauges:

$$BS_{ref}(h) = \frac{1}{N} \sum_{k=1}^N (\bar{o}(h) - o_k(h))^2$$

However, other methods, which are less computationally expensive than the EPF, can be used for producing a PF by combining the DF with a prescribed (Gaussian) probability distribution to obtain a probability density function (PDF):

- DPF. The undressed probability forecast assumes a perfect forecast with no uncertainty so that the probability of H is 1.0 if it occurs in the DF, 0 otherwise (this can be considered a limit case in which a Gaussian with zero standard deviation is adopted).
- MDPF. The mean dressed PF assumes a Gaussian PDF centered on the DF with a prescribed standard deviation, which is the rms error of the DF computed at hourly step from 1 to 144 h lead time.
- MultiDPF. The multiple dressed PF differs from the PF, as it assumes a standard deviation which is the sum in quadrature of two terms: the mean overall DF rms error and a percentage of the SR predicted by the DF. This percentage has been assumed equal to 33 % (20 %), when verifying against observations (hindcast values). This is meant to describe the increase of uncertainty with the SR level.

3 Analysis of a storm surge event

This section analyzes the events which took place on the 9th and 10th of November 2010, during which a seiche and storm surge overlapped, causing two SR peaks on the 9th and 10th and a sequence of subsequent oscillations. The panels of Fig. 2 show the SR forecasts for different lead times. All panels consider the same overall time window (11 days long from the 4th to 15th of November) reporting the observed SR level at the ISMAR-CNR platform (thick blue line, identical in all panels). All forecasts in a single panel have the same initial date, from Fig. 2a showing the forecasts launched on the 4th of November at 00UT to Fig. 2f showing those launched on the 9th of November at

00UT (there is a 24-h step between forecasts in two successive panels). Each panel shows the 50 simulations of the EPS (E-members, thin gray lines), the DF (thick green line), the EMF (thick red line), the CRF (thick brown line), and the hindcast (thick black line). All forecasts in a single panel start from the same initial condition because of the common previous 10-day analysis.

The main weather impulse producing the storm surge in the Adriatic Sea occurred on the 9th of November. The seiche triggered by the first storm surge peak returned amplified on the 10th of November and was followed by a sequence of seiches with a period of about 22 h. The days before the events were characterized by a slow steady decrease in MSLP along the entire Adriatic Sea, during which pressure was almost uniform in the north and south Adriatic (slightly higher in the south than in the north). Then, in the 48 h before the surge peak, MSLP dropped at the north Adriatic coast and the pressure gradient determined the Sirocco wind (blowing from south-east), which accumulated water towards the closed end of the basin and produced the first storm surge peak on the 9th. The situation triggered the formation of the longitudinal Adriatic 22-h seiche, whose presence is evident during the following days. On November 10, the highest SR event in this time window (62 cm) was caused by the superposition of persisting Sirocco wind with the seiche. The DF, CRF, and EMF are similar. The hindcast performs better than the forecasts, especially for simulations launched at the beginning of the considered time window, possibly suggesting errors in the meteorological prediction of the conditions leading to the onset of the storm surge.

The panels report that the EPF reached the 50-cm SR level. This value has been selected to represent a SR threshold that in combination with the peaks of the astronomical tide would have required to lift the MOSE barriers for preventing the flooding of Venice. In this case, the confidence of the prediction increased as the lead time of the forecast decreased and reaching the threshold was certain in the forecast issued about 2 days in advance.

This event has been used for showing how seiches characterize the effect of uncertainties in the initial condition of the forecast. A set of 50 different initial conditions has been generated using the EPS launched on the 9th of November (panel 2f). After 36 h, on the 10th of November at 12UT (approximately in correspondence with the maximum SR level), the state of the circulation and SR fields differ appreciably among the ensemble members. These 50 different circulations and SR fields are used as initial conditions of 50 simulations driven by the same sequence of meteorological fields, which have been provided by the ECMWF reanalysis. Therefore, the behavior of the ensemble shows how the differences due to the ocean initial condition evolve under the action of a common meteorological forcing. Figure 3 shows at the five tide gauge stations (ISMAR-CNR, Trieste, Rovinj, Split and Dubrovnik) the evolution of the ensemble spread,

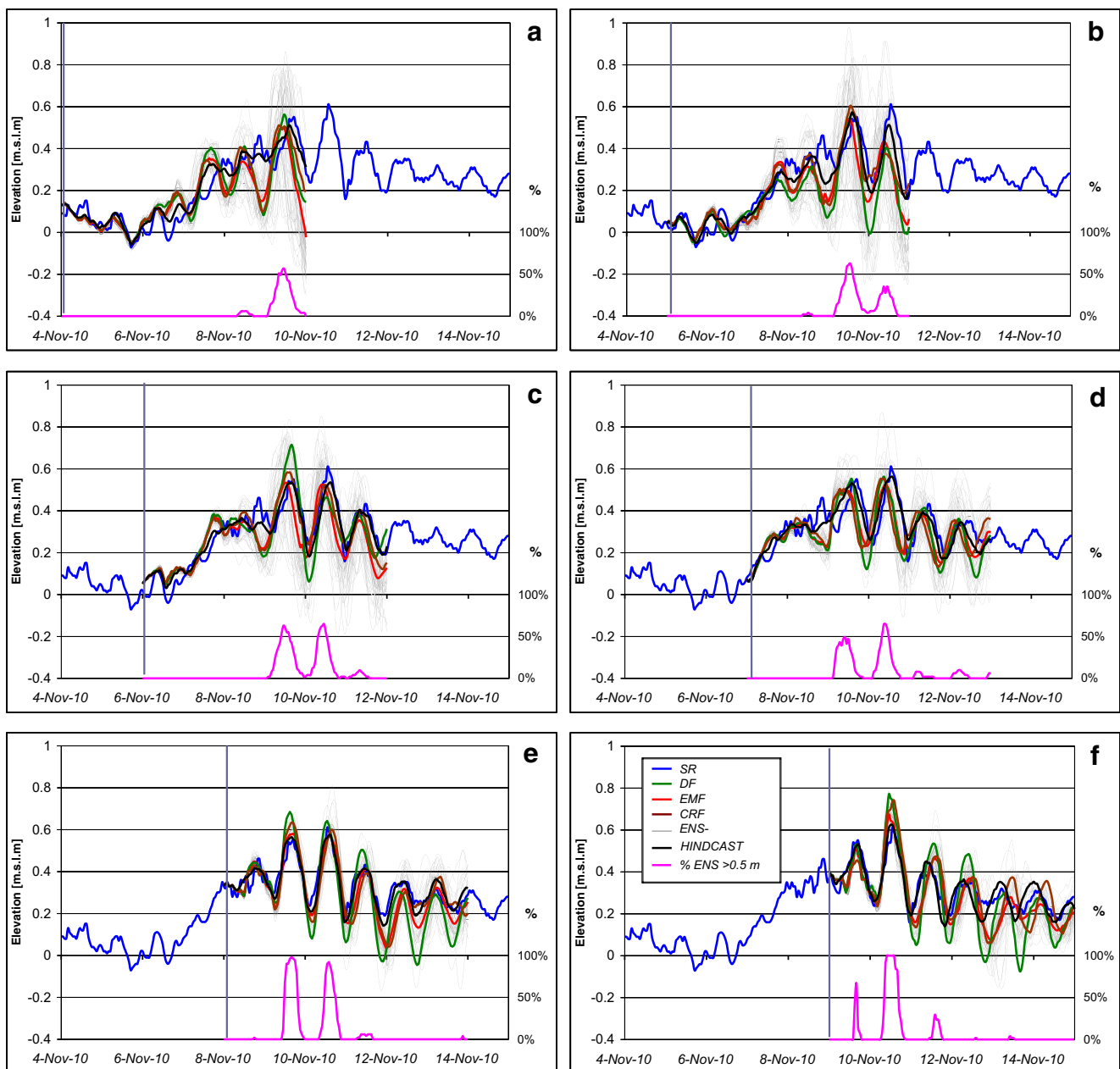


Fig. 2 Forecasts of the event from the 9th to the 10th of November 2010 at the ISMAR-CNR platform with different lead times. Each panel considers the same overall time window (11 days long from the 4th to the 15th of November) reporting the observed SR level (values in meters, *left side of panels*) at the ISMAR-CNR platform (*thick blue line*, identical in all panels). Panels report a set of forecast with different lead times at 24-h intervals: from panel **a**, showing the forecasts launched on the 4th of

November at 00UT to panel **f**, showing those launched on the 9th of November at 00UT. Each panel shows the 50 simulations of the EPS (ENS-MEMBERS, *thin gray lines*), the DF (*thick green line*), the EMF (*thick red line*), the CRF (*thick brown line*), and the hindcast (*thick black line*). The *pink line* shows the probability of exceeding the 50-cm SR level (values on the *right side of panels*)

whose decay (with a characteristic time that can be visually estimated about 3–4 days) is strongly modulated by oscillations, which correspond to the main Adriatic seiche. Note that switching off completely the meteorological forcing would have produced very similar results as Fig. 3 is concerned. Therefore, this figure shows the relevance of the seiches in the Adriatic Sea, whose enduring dynamics that are rather insensitive to small changes in the meteorological forcing.

4 EPS spread and error statistics

In this section, the link between EPS spread and rms errors of DF, EMF, and CRF at the tide gauges located along the Adriatic Sea is discussed. The analysis considers data aggregated over all five tide gauges. However, a similar behavior can be seen if tide gauges were analyzed separately. The aim is to establish the relation linking EPS spread to the forecast

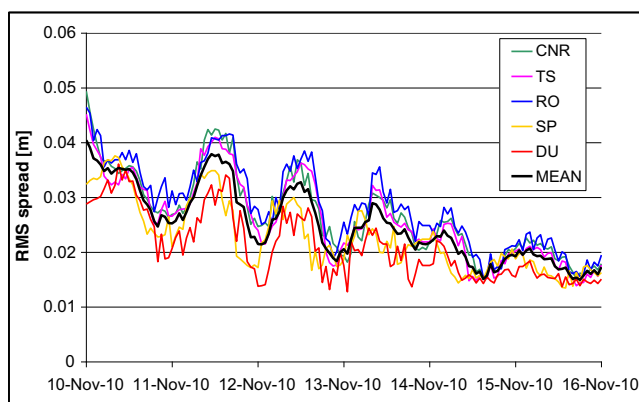


Fig. 3 Evolution of the ensemble spread caused by different ocean initial conditions under the action of a common meteorological forcing. The figure shows the spread at the gauge stations ISMAR-CNR (*CNR*), Trieste (*TS*), Rovinj (*RO*), Split (*SP*), Dubrovnik (*DU*), and their *mean*

uncertainty, to analyze how it varies with forecast lead time and with the SR predicted level.

4.1 SR-EPS spread and rms errors as a function of the lead time

The spread among EPS simulations represents a measure of the uncertainty of prediction. It is expected that the EPS spread increases with the forecast lead time and that cases with large spread are those when the EMF, DF, and CRF errors are likely to be large.

Figure 4 shows the evolution of the SR-EPS forecast spread and rms errors as a function of the lead time. The error behavior is shown for DF, EMF, CRF, and for the average of all the ensemble members. The left panel (Fig. 4a) refers to the verification of predictions against the observed data at tide gauges, the right panel (Fig. 4b), against the corresponding hindcast values at those locations.

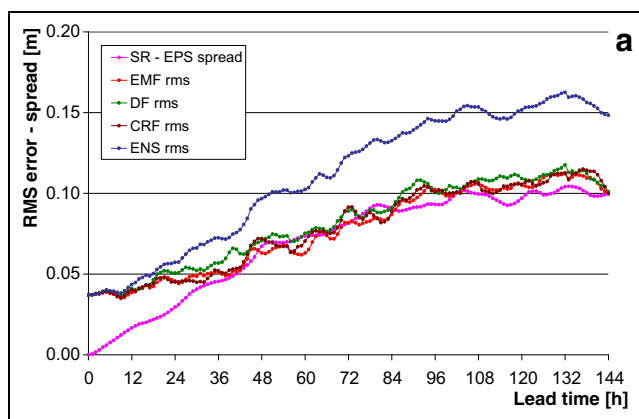


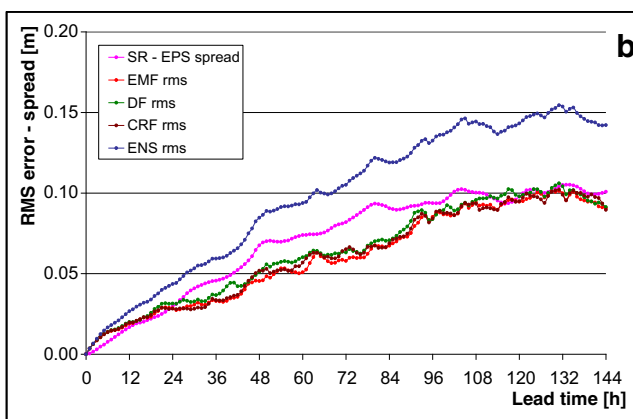
Fig. 4 *Left panel a* Time evolution of the SR-EPS spread (m, pink line) and rms error (m) of ensemble mean forecast (EMF, red line), deterministic forecast (DF, green line), control run forecast (CRF, blue line), and all ensemble members (ENS, blue line) with respect to observations as

In both panels, rms errors and spread increase substantially linearly with lead time up to 100 h, reaching a value of approximately 10 cm (15 cm for the ensemble average rms) and stop growing for larger lead time, possibly because of the limited amplitude of the SR signal, whose standard deviation lies in the range 13–15 cm and limits the maximum value of the rms error.

Generally, the EMF presents a slightly lower error than DF and CRF, but the difference is not significant. CRF and DF present very similar rms errors, even though DF is based on meteorological data at a higher resolution than CRF and it would be expected to provide a more realistic simulation. This may be due to the “double penalty” effect, where point-wise comparisons favor smooth forecasts over sharper but slightly misplaced detail (Flowerdew et al. 2010).

The rms error with respect to observations (Fig. 4a) is nonzero (4 cm) at lead time zero for all forecasts because of the error in the initial condition of the hydrodynamical model, which is an important contribution, being approximately 40 % of the rms error value at the end of the forecast. For lead times larger than 36 h, the left and right panels of Fig. 4 present very similar rms values, suggesting that errors of the meteorological forcings are the main source of uncertainty for large lead times, while the atmospheric initial condition error dominates at short lead times.

The SR-EPS spread, which is obviously zero at the beginning of the forecast, initially grows faster than the rms errors with respect to observations, and it reaches their value after about 36 h (Fig. 4a). Therefore, the SR-EPS spread, as computed by this EPS configuration, is not a reliable representation of the forecast error for short lead times. However, over a large range of lead times (24–108 h), the SR-EPS spread matches well the rms error of the EMF, showing that ensemble members are sampling well the distribution of possible outcomes and for large lead time, the SR-EPS spread provides an acceptable description of the forecast uncertainty.



function of the lead time (h). The average of the values of individual gauges is considered. *Right panel b* shows the same quantities, except that the values of the hindcast at the tide gauges are considered for rms computation.

The SR-EPS spread is larger than the EMF rms with respect to the hindcast (Fig. 4b) for lead times from 36 to 96 h, showing that in this time range uncertainty is not fully related to the meteorological fields, but, possibly, it is partially associated with seiches keeping the memory of previous uncertainty.

4.2 SR-EPS spread and EMF

Figure 5 shows rms spread and rms errors binned by EMF values aggregated over all five tide gauges and considering all forecast lead times. The left panel (Fig. 5a) considers rms with respect to observations, the right panel (Fig. 5b) with respect to the corresponding hindcast values. Figure 5 reports also the percent of data in each bin and shows the asymmetry of positive and negative anomalies (Table 1 shows the actual values). Small (large) negative anomalies are more (less) frequent than positive anomalies, which corresponds to the asymmetry of the conditions leading to these two opposite conditions (see Conte and Lionello 2013 for a climatological analysis along the Mediterranean coastline).

Results show that the highest SR forecast values have the largest uncertainty and are most likely to be appreciably wrong. However, percent-wise EPS spread is smaller for large than for small SR anomalies, suggesting a robust prediction of large storm surges.

For large SR EMF anomalies, EMF rms error is lower than DF and CRF rms error, showing a clear improvement and robustness of the EMF with respect to the traditional techniques. For small anomalies, rms errors in Fig. 5b are much smaller (about 50 %) than those in Fig. 5a, showing that a substantial fraction of them is, likely, caused by uncertainty in the initial condition and not in the meteorological evolution. However, even for small anomalies, EPS spread remains non negligible (about 5 cm), as the seiches in the Adriatic Sea keeps for a long time the memory of uncertainty previously

introduced in the system (see Mel and Lionello 2014 and Fig. 2 of this paper).

4.3 SR-EPS spread and rms error of SR predictions

Figure 6 shows the relation between the EPS spread and the rms error. When considering the observations (Fig. 6a), the results show that the EPS spread is linked to the EMF rms error except for very small errors. This is because in the initial period of the forecast EMF errors are small, but not nil, while the corresponding EPS spread is zero. In other words, a small spread cannot be associated to a vanishing error when the error is in the initial part of the run. On the contrary, the EPS spread increases faster than the rms errors so that the largest errors of EMF (and the DF and CRF as well) are appreciably smaller than the corresponding EPS spread. This suggests that the EPS overestimates the uncertainty associated with large departure of individual ensemble members from the EMF.

When errors are computed with respect to the hindcast (Fig. 6b), the EPS spread is linked to them by an approximately linear relation up to values of about 8 cm. In this range, the EPS spread realistically reproduces uncertainty in the surge that is produced by errors in the meteorological forcing. However, for large values (14 cm), the spread overestimates the uncertainty of the EMF (and of the DF and CRF as well) and such overestimate is larger than in Fig. 5a. This larger discrepancy between SR-EPS spread and rms error is not surprising, as verification against hindcast necessarily underestimates actual rms errors.

5 Surge residual Brier skill score

Figure 7 shows the BSS computed on the basis of the EPS as a function of the lead time for six different thresholds h . The left

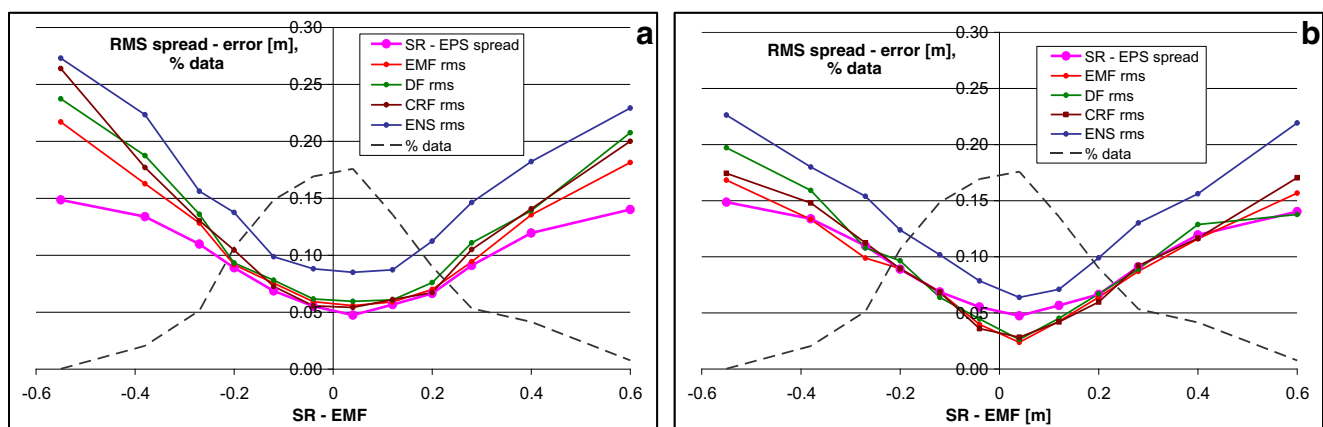


Fig. 5 Left panel **a** Dependence of SR-EPS spread (m, pink line) and rms (m) of ensemble mean forecast (EMF, red line), deterministic forecast (DF, green line), control run forecast (CTR, blue line), and all ensemble members (E, brown line) as function of the EMF results (m). The gray

dashed line represents the data percentage in each bin (see Table 1 for normalization). The left panel **a** refers to the rms of the observations and the right panel **b** to the corresponding hindcast values

Table 1 Number of data and percent for the bins used in Fig. 5 and represented by the gray line

Range	Events	Percentage of data
−0.80 to −0.50	23	0.02
−0.50 to −0.32	2668	2.04
−0.32 to −0.24	6700	5.13
−0.24 to −0.16	13,929	10.67
−0.16 to −0.08	19,368	14.84
−0.08 to −0.00	22,077	16.92
0.00 to 0.08	22,943	17.58
0.08 to 0.16	17,709	13.57
0.16 to 0.24	11,731	8.99
0.24 to 0.32	6956	5.33
0.32 to 0.50	5405	4.14
0.50 to 0.80	991	0.76

panel (7a) shows the verification against observed values, the right panel (7b), against the hindcast. Data are aggregated considering all tide gauges.

Figure 7a shows that BSS remains positive for very large lead times and diminishes with the threshold. Therefore, the skill of the PF decreases for increasing thresholds but it remains positive for 4 days, seven for the highest considered threshold (50 cm). The values of BSS referring to the hindcast are larger than those referring to the observations and remain positive for all lead times that have been considered in this study (>6 days). This emphasizes the relevance of the initial condition uncertainty as a factor decreasing the skill of the prediction and that seiches, when they are introduced in the forecast, can affect for a long time the results of the SR forecast. However, when considering large thresholds ($h=40$ and 50 cm), the differences between verification against observation (7a) and hindcast (7b) decreases, suggesting that for

large SR values, the uncertainty in the meteorological forcing can affect the results of the SR forecast for a long time.

The BSS has been computed using the EPS results and three simple methods (UDPF, MDPF, MultiDPF) for producing a probabilistic forecast (see “Section 2” for their description). Figure 8 shows the BSS for the thresholds $h=0$ cm and $h=40$ cm (Fig. 8a considers the observed tide gauge data, Fig. 8b, the corresponding hindcast values). The BSS of the EPF systematically outperforms those of the other PF methods. Particularly, the BSS of the UDPF, which does not account for any error in the forecast, performs poorly. In Fig. 8a, even for $h=0$, the value for which the BSS of all other methods remains consistently above 0.2 for the whole 6-day-long period, the BSS of UDPF drops below 0.2 already after 48 h. Considering the $h=40$ cm threshold, the UDPF BSS drops below 0 after 60 h. This is a strong indication that a probability estimate based on a plain DF approach cannot deliver reliable results. The MultiDPF, though its results are marginally less accurate than those of the EPF, has a high BSS and could be used as an alternative to the much more expensive EPF. Finally, Fig. 8 shows that the BSS verified against the hindcast is consistently higher for all lead times and for all methods due to the exclusion of the wrong initial conditions error.

6 Conclusions

An EPS forecasting system has been implemented in the Adriatic Sea, using the ECMWF EPS for the wind and MSLP forcing fields for a hydrodynamical SR model. The analysis of the results is focused on the 3-month period October–December 2010 during which the operational use of the EPS has been simulated.

In the example discussed in “Section 3,” the first peak was predicted with good accuracy already with a 6-day lead time,

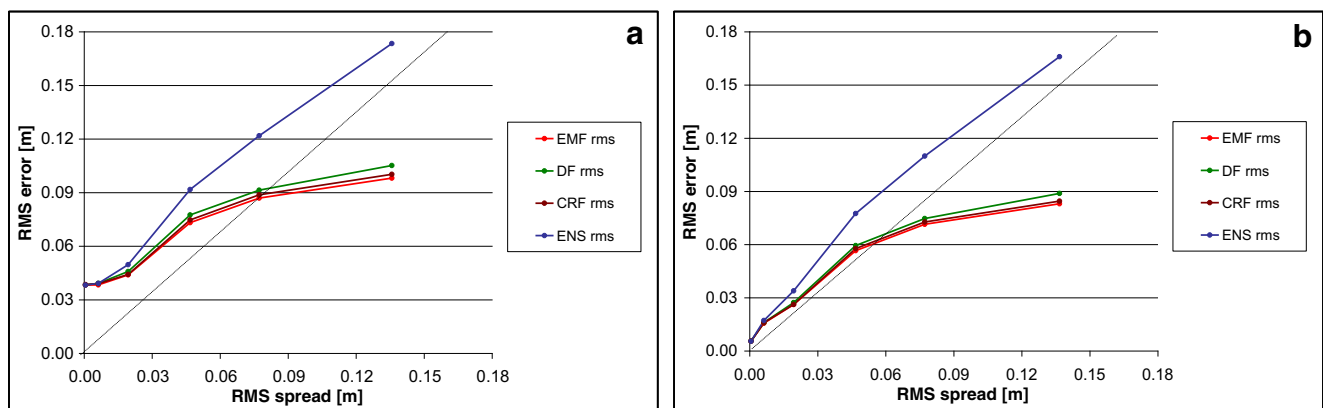


Fig. 6 rms errors (m) as a function of the SR-EPS spread (c) for ensemble mean forecast (EMF, red line), deterministic forecast (DF, green line), control run forecast (CRF, brown line), and all ensemble members

(E, brown line). The left panel a refers to the rms of the observations and the right panel b to the corresponding hindcast values

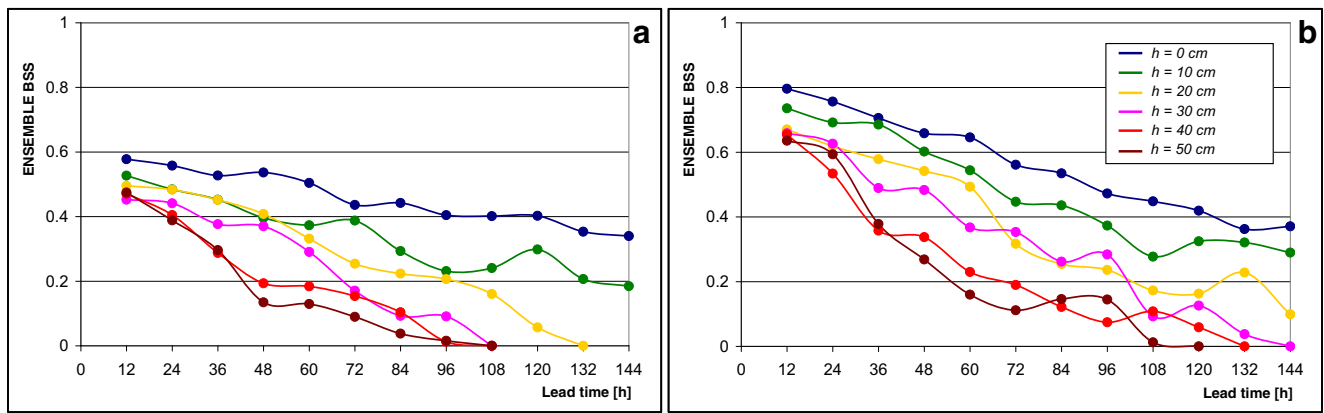


Fig. 7 Brier skill score of the EPS as a function of the lead time for different surge residual threshold h (blue line $h=0$ cm, green line $h=10$ cm, yellow line $h=20$ cm, pink line $h=30$ cm, red line $h=40$ cm, brown

line $h=50$ cm). The left panel **a** shows the BSS computed using observations, the right panel **b** using the hindcast

while a comparable accuracy for the second peak was achieved with a 3-day lead time. In any case, the EPS spread included the observed values at all lead times and would have allowed a warning of the occurrence of SR levels above critical thresholds (here 50 cm has been adopted) several days in advance.

The forecast errors increase with the lead time. The EMF has a rms error lower than DF, especially for short (up to 3 days) lead times. However, differences are not large, and the main advantage of EMF appears to be related to a more robust prediction of the peak values (Mel and Lionello 2014) or large SR values and not to an overall substantial reduction of the

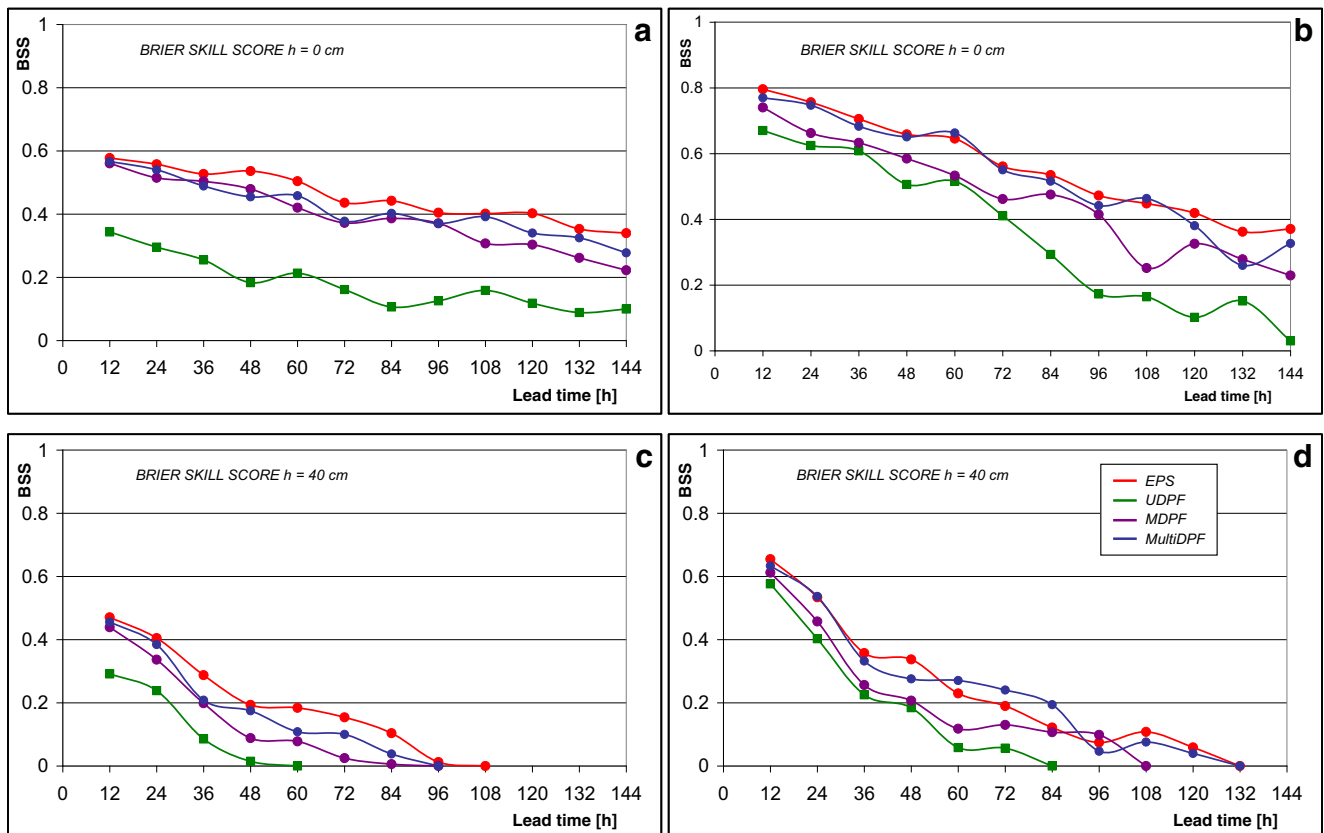


Fig. 8 Brier skill score as a function of lead time for the three different PF methods (red line EPS, green line UDPF, purple line MDPF, blue line MultiDPF). In the left/right panels, the BSS is computed with respect to

observations/hindcasts. Top/bottom line considers $h=0$ cm/ $h=40$ cm. (Panel **a** observations, $h=0$ cm. Panel **b** hindcast $h=0$ cm. Panel **c** observations $h=40$ cm. Panel **d** hindcast $h=40$ cm)

rms with respect to DF. In general, in absolute terms, the uncertainty of the forecast increases with the predicted SR value, but as percent of the predicted SR, it is lower for large than for small SR. (this is reflected in Fig. 4).

Results show that uncertainty for short lead times (up to 36 h) of the forecast and for small SR values is caused by uncertainty of the initial condition of the hydrodynamical model. This suggests that inserting in a prediction system a data assimilation procedure such as that proposed by Lionello et al. 2006, it is essential for a reliable short lead time forecast. Uncertainty for large lead times and large SR values is mainly caused by uncertainty on the meteorological forcings.

The EPS spread is demonstrated to be linked to the rms error of the forecast, and it increases with the rms error. For large lead times, the EPS spread and the forecast error substantially coincides. However, the EPS spread in this study, which does not account for uncertainty in the initial condition, underestimates the error during the early part of the forecast (up to 36 h) and for small SR values. On the contrary, it overestimates the rms error for large SR values. This is probably because the EPS spread underestimates the uncertainty due to the meteorological forcings when large SR anomalies (both positive and negative) are predicted. The importance of the uncertainty on the initial condition suggests that it would be interesting to combine the EPS with an ensemble data assimilation procedure capable of producing a set of different initial conditions for the hydrodynamical model. This might improve the information on the uncertainty of the prediction for lead times up to 36 h, which are operationally crucial, and it appears to be a priority for future developments of an operational prediction system.

The EPF has a clear skill in predicting the actual probability distribution of SR, while a probability estimate based on a single DF is shown to be inadequate. However, a PF obtained with a prescribed Gaussian centered on the DF value accounting for the dependence of the DF rms on the predicted SR values and lead times performs very similarly to the EPS-based PF. Therefore, MultiDPF can be considered a practical computationally cheap alternative to EPF, though, however, EPF outperforms all other PF methods.

The progress of this research will regard the study of the initial conditions, uncertainties, and the sensibility of the forecast of them. The future development of the EPS is to estimate via a linear combination of suitable meteorological variances the uncertainty affecting storm surge prediction.

Acknowledgments The hourly SL observations in CNR-PLATFORM and Venice “Punta Salute” were provided by I.C.P.S.M. of the Venice City Council, observations in Trieste by Dr. Fabio Raicich of ISMAR-CNR (Italy), observations in Rovinj, Split, Dubrovnik by Dr. Nenad Leder of the Croatian Hydrographic Institute.

References

- Battistin D, Canestrelli P (2006) 1872–2004 La serie storica delle maree a Venezia, Internal report of Comune di Venezia - Istituzione CPSM. http://93.62.201.235/maree/DOCUMENTI/La_serie_storica_delle_maree_a_Venezia_1872-2004_web_ridotto.pdf
- Brier GW (1950) Verification of forecasts expressed in terms of probabilities. *Mon Weather Rev* 78:1–3
- Buizza R, Palmer TN (1995) The singular-vector structure of the atmospheric general circulation. *J Atmos Sci* 52:1434–1456
- Buizza R, Miller M, Palmer TN (1999) Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Q J R Meteorol Soc* 125:2887–2908. doi:10.1002/qj.49712556006
- Conte D, Lionello P (2013) Characteristics of large positive and negative surges in the Mediterranean Sea and their attenuation in future climate scenarios. *Glob Planet Chang* 111:159–173. doi:10.1016/j.gloplacha.2013.09.006, ISSN 0921-8181
- de Vries JW (2009) Probability forecasts for water levels at the coast of The Netherlands. *Mar Geod* 32(2):100–107. doi:10.1080/01490410902869185
- De Zolt S, Lionello P, Malguzzi P, Nuhu A, Tomasin A (2006) The disastrous storm of 4 November 1966 on Italy. *Nat Hazards Earth Syst Sci* 6:861–879
- Di Liberto T, Brian AC, Nickitas G, Blumberg AF, Taylor AA (2011) Verification of a multimodel storm surge ensemble around New York City and Long Island for the cool season. *Weather Forecast* 26:922–939. doi:10.1175/WAF-D-10-05055.1
- Eprim Y, Donato MD, Cecconi G (2005) Gates strategies and storm surge forecasting system developed for the Venice flood management. In: Fletcher C, Spencer T (eds) *Venice and its lagoon*, State of Knowledge. Cambridge University Press, Cambridge UK, pp 267–277
- Flowerdew J, Horsburgh K, Kevin C, Mylne K (2009) Ensemble forecasting of storm surges. *Mar Geod* 32(2):91–99. doi:10.1080/01490410902869151
- Flowerdew J, Horsburgh K, Wilson C, Mylne K (2010) Development and evaluation of an ensemble forecasting system for coastal storm surges. *Q J R Meteorol Soc* 136:1444–1456. doi:10.1002/qj.648
- Flowerdew J, Mylne K, Jones C, Titley H (2012) Extending the forecast range of the UK storm surge ensemble. *Q J R Meteorol Soc*. doi:10.1002/qj.1950
- Leutbecher M, Palmer TN (2008) Ensemble forecasting. *J Comput Phys* 227:3515–3539
- Lionello P (2012) The climate of the Venetian and North Adriatic region: variability, trends and future change. *Phys Chem Earth* 40–41:1–8
- Lionello P, Mufato R, Tomasin A (2005) Effect of sea level rise on the dynamics of free and forced oscillations of the Adriatic Sea. *Clim Res* 29:23–39
- Lionello P, Sanna A, Elvini E, Mufato R (2006) A data assimilation procedure for operational prediction of storm surge in the northern Adriatic Sea. *Cont Shelf Res* 26:539–553
- Lionello P, Cavaleri L, Nissen KM, Pino C, Raicich F, Ulbrich U (2012) Severe marine storms in the Northern Adriatic: characteristics and trends. *Phys Chem Earth* 40–41:93–105. doi:10.1016/j.pce.2010.10.002
- Lorenz EN (1965) A study of the predictability of a 28-variable atmospheric model. *Tellus* 17:321–333
- Massalin A, Zampato L, Papa A, Canestrelli P (2007) Data monitoring and sea level forecasting in the Venice Lagoon: the ICPSM’s activity. *Boll Geofis Teor Appl* 48:241–257
- Mel R, Lionello P (2014) Storm surge ensemble prediction for the city of Venice. *Weather Forecast* 29:1044–1057. doi:10.1175/WAF-D-13-00117.1

- Mel R, Viero DP, Carniello L, Defina A, D'Alpaos L (2014) Simplified methods for real-time prediction of storm surge uncertainty: the city of Venice case study. *Adv Water Resour* 71:177–185
- Molteni F, Buizza R, Palmer TN, Petroliagis T (1996) The ECMWF ensemble prediction system: methodology and validation. *Q J R Meteorol Soc* 122:73–119
- Tomasin A (1972) Auto-regressive prediction of sea level in the northern Adriatic. *Riv Ital Geofis XXI*:211–214
- Tosoni A, Canestrelli P (2011) Il modello stocastico per la previsione di marea a Venezia. *Atti Ist Veneto Sci Lett Arti Tomo CLXIX*(2010–2011):21–42
- Umgiesser G, MelakuCanu D, Cucco A, Solidoro C (2004) A finite element model for the Venice Lagoon. Development, set up, calibration and validation. *J Mar Syst* 51:123–145
- Wilks D S (2006) *Statistical methods in the atmospheric sciences*, 2nd edn. Academic Press, 627 pp