ORIGINAL ARTICLE



A forecasting and warning system of storm surge events along the Norwegian coast

Nils Melsom Kristensen¹ · Lars Petter Røed^{1,2} · Øyvind Sætra¹

Received: 31 January 2022 / Accepted: 2 May 2022 © The Author(s) 2022

Abstract

We consider the Norwegian Meteorological Institute's system regarding the daily forecasts of water level, and warnings of possible dangerous water level events along the Norwegian coast. The system consists of three parts, the production of water level forecasts, a decision support system, and a system for dissemination of warnings to key users and the general public. Included is a brief description of the forecasting model and the parallel ensemble prediction system, and an assessment of the forecasts produced by them. Based on the assessment of the production models for a three year period we find that they provide sufficiently trustworthy forecasts of water levels for the purpose at hand. Also included is a description of the web based decision support system for issuing warnings. The decision support system was for instance used during the extreme weather event "Elsa" in February 2020, and was found to be an efficient tool both to monitor the event in its early stages and to expedite warnings to key users and the general public.

Keywords Storm surge · Water level · Ocean modeling · Probabilistic forecasting

Abbreviations

2D	Two-dimensional
3D	Three-dimensional
API	Application programming interface
ECMWF	European centre for medium-range weather forecasts
ECMWF-ENS	ECMWF ensemble prediction system
ECOM3D	The estuarine and coastal ocean circulation model 3D
EPS	Ensemble prediction system
FFT	Fast fourier transform
HAT	Highest astronomical tide
LAT	Lowest astronomical tide
M2	M2 tidal component
M4	M4 tidal component

Nils Melsom Kristensen nilsmk@met.no

¹ Department for Ocean and Ice, The Norwegian Meteorological Institute, PO Box 43, 0313 Blindern, Oslo, Norway

² Department of Geosciences, University of Oslo, PO Box 1047, 0316 Blindern, Oslo, Norway

M6	M6 tidal component
MAE	Mean absolute error
Max_T	Maximum total water level
Max _s	Maximum storm surge
ME	Mean Error
MET Norway	Norwegian Meteorological Institute
Min_T	Minimum total water level
Min _s	Minimum storm surge
MNOK	Million Norwegian Kroner
NNT	Norwegian normal time
NOK	Norwegian Kroner
RMSE	Root mean square error
ROMS	Regional ocean modeling system
UK	United Kingdom
UTC	Coordinated universal time

1 Introduction

Flooding due to coastal storm surges presents a significant threat to life and property world wide. This is also true in the North Sea and along the entire coast of Norway. For instance the North Sea storm surge of 31 January 1953 caused the loss of 307 lives in East Anglia, UK [1] and a further 1836 fatalities in the Netherlands [2]. An historic example for Norway is a catastrophic surge that occurred at a low island on the western coast of Norway [3]. In the spring of 1670 48 people out of a population of 50 were reported to be killed during a storm which obviously has flooded the island.

More recent examples for Norway includes the storm surge by an extreme weather event in October 1987, and the two extreme weather events "Dagmar" (December 2011) and "Elsa" (February 2020). While the 1987 event mostly affected Norway's capital Oslo and the nearby town of Drammen, the two others hit the western coast of Norway. All storms caused record high water levels, but caused no loss of life. However, the damage to properties due the storm surge alone was estimated to about 130 MNOK in 1987, to record high 400 MNOK in 2011 and to about 100 MNOK in 2020 according to statistics from Finans Norge (Fig. 1). In fact since 1980 flooding caused by storm surges have caused damages estimated to 1000 MNOK in Norway alone.

Today it is common to mitigate this threat by providing advance warning of extreme water levels, so that protective actions can be taken. Already in the 1950s the Norwegian Meteorological Institute (MET Norway) started to analyze extreme storm surge conditions events along the coasts of Norway. At that time they also developed empirical methods to forecast possible damaging events [4]. Later [3, 5] showed that extreme surge conditions along the western coast of Norway usually occurred during south westerly winds, that is, with a wind direction parallel to the coast. By using a simple barotropic storm surge model [5] they showed that severe damage occurs only if the storm at the same time generates high seas. This situation usually happens if the south westerly wind shifts to a strong westerly wind before the amplitude of the surge has decayed, a not at all uncommon situation along the western coast of Norway, and which was demonstrated by the storm "Dagmar" in 2011.



Damages due to storm surges alone 1980-2021

Fig. 1 Damage to property due to storm surges alone from 1980-2021 (in 1000 NOK). Note the years 1987, 2011 and 2020. Source: Finans Norge (https://www.finansnorge.no/statistikk/skadeforsikring/naturskade statistikk-nask/)

The numerical storm surge model developed by [5] was implemented at MET Norway in the early 1980s as a tool to possibly forecast dangerous storm surge events. At the same time also a prediction and warning system to provide advance warning of dangerous water level events was put into action. At the core of this system was the numerical model [5], a model that in the late 1980s and early 1990s was replaced by the so called ECOM3D model [6, 7]. As is well known changes in the day to day water level is caused mainly by the astronomical tides (henceforth tides) caused by the gravitational pull from astronomical objects of which the forces of the moon and the sun are the most important ones. The storm surge, which is the water level rise and fall due to variations in the atmospheric forcing due to wind and pressure, may sometimes work to lower the high tides. Since variations in the storm surge follows the weather pattern, which generally has a longer time scale than the tidal cycle, the extreme storm surge events commonly occurs jointly with high tides.

At the core of the present day warning system at MET Norway is currently the barotropic version of the ocean model ROMS [8, 9]. Although a well calibrated and validated numerical forecasting model will always be the backbone of any such system, the surrounding infrastructure, such as the flow of real-time observations, protocols for dissemination of forecasts to relevant key personnel and authorities responsible for safeguarding life and property, as well as informing the general public, is of equal importance. Hence, we give an holistic account of today's daily forecasting of storm surges at MET Norway and the warning system of possible dangerous water-level events. Included is the observation system and its data flow, the present numerical storm surge model and the treatment of tides, an assessment of the performance of today's storm surge model, the real-time use of observations for the correction of model forecasts, the uncertainty estimates using ensemble predictions, and the procedures for issuing emergency warnings during extreme events.

While Sect. 2 gives a short overview of the forecast system and the data flow, Sect. 3 gives a description of the observational network and how the data are used in real-time to correct the model predictions. Section 4 presents the numerical storm surge model and the ensemble prediction system. Section 5 gives a summary of the performance of the deterministic and probabilistic (ensemble) forecasts, while Sect. 6 describes the dissemination



system, the decision support system, and the process of issuing extreme water-level warnings. Finally, Sect. 7 provides a summary together with some final remarks.

2 Overview of the model systems and the data flow

The present system at MET Norway to predict and warn about extreme water levels is targeted to 23 permanent stations for water level observations. 22 of them are along the Norwegian coast while one is located on the arctic island of Svalbard (Fig. 2). Observations from these stations are continuously transferred to MET Norway in near-real time and utilized for post-processing to adjust and improve the forecasts (Fig. 3). The tide gauges are operated by the Norwegian Mapping Authorities and the data is freely available online through an API¹.

¹ https://api.sehavniva.no/tideapi_no.html.



Fig.3 Schematic description of the production chain and work flow regarding forecasting and warning of possible dangerous water level events in Norway. The system is in large divided into two parts: The production of the forecast and the dissemination and warning system. Each part consist of a number of key components. For the production part the backbone is the ROMS ocean model used for creating the deterministic and probabilistic forecasts. These use atmospheric forcing and water level input at open boundaries (through the inverse barometer effect) from the ECMWF atmospheric model. The model forecasts are in turn post-processed and adjusted by adding astronomical corrections (tides) and correcting for differences between forecast and observations from the Norwegian Mapping Authorities (from the SeHavnivå API at https://api.sehavniva.no/tideapi_no.html). For the dissemination and warning system part the backbone of the system is the Forecasting Center at MET Norway itself (that is the forecaster on duty). The main tool here is the decision support system dashboard as described in Sect. 6.1. The dissemination of every day forecasts of water level to the general public is handled by web visualization on the SeHavnivå web page at http://sehavniva.no. If a warning of high water level is issued by the forecaster on duty, this is communicated to subscription and key users via email, and also communicated to the general public via media and the MET Norway web page https://twitter.com/meteorologene.

As mentioned (Sect. 1) the total observed water level is usually divided into two components, the astronomical tide and the storm surge component. While the storm surge component is commonly forecasted by a numerical model the tidal component is estimated from harmonic analysis of long time-series of observations (often 30 years or more). To obtain the storm surge component the astronomical tides found by harmonic analysis is subtracted from the observed total water level. However, these two components are non-linearly dependent and can not be completely separated. The observed tide will therefore always be subject to some contamination by weather effects, even based on analysis of very long time-series. Ideally, the forecasting model should therefore contain and forecast the combined water level due to both tides and weather effects, as demonstrated in the work by [10].

Astronomical corrections, where tidal predictions based on observations are added to the pure storm surge signal, have traditionally been shown to increase the accuracy of the total water level signal [10]. Consequently the operational storm surge model is first run producing forecasts with a lead time of 5 days. This applies both to the deterministic and probabilistic model component (for details see Sect. 4). Because of the shortcomings of tides in the model, the system runs without this component, predicting only the atmospheric contribution to the water level variations, that is, the storm surge. The astronomical tides are then added at each targeted stations based on predictions from harmonic analysis of the observations. Finally, the latest available observations are used to adjust the forecasts for possible discrepancies before they are disseminated via the Forecasting Center at MET Norway to the official web-site for water level forecasts operated by the Norwegian Mapping Authorities² as illustrated by Fig. 3 (for further details see Sect. 5).

At the Forecasting Center, forecaster are on duty 24/7 monitoring the forecasts via the decision support system (as described in Sect. 6.1). When certain alert levels for the predicted total water level is exceeded, specific to each station, the forecaster issues warnings that are disseminated to local responsible authorities and through national media (Sect. 6).

3 Observations of water level and their usage

We use water level observations downloaded from the Norwegian Mapping Authorities API every day for several purposes. One is for validation and optimization of the model forecast, another for real-time correction or adjustment of the model predictions, and last but not least by the forecasters to ensure the validity of the forecast.

With regard to the tidal component we observe from Table 1 that the contribution from the Highest Astronomical Tide (HAT) to the maximum total water level Max_T generally increase as we move from south to north along the western Norwegian coast, that is, from Stavanger to Vardø, while the maximum storm surge contribution (Max_S) is relatively unchanged (with a few exceptions inside long, narrow fjords, e.g., Narvik).

In contrast we note that at all stations east of Stavanger, that is, in the Skagerrak area, the maximum storm surge contribution is of the same order of magnitude or larger than the tidal contribution. The low tidal amplitudes in this area are caused by the closeness to the amphidromic point positioned in the North Sea near the village of Egersund. The storm surge signal on the other hand is relatively large due to several factors, which may mainly be attributed to the shape of the Skagerrak area. For instance Kelvin waves formed in the North Sea easily gets "trapped" in the Skagerrak, and the fact that this area is usually located to the south of the path of incoming low pressure systems (storms) that cross southern Norway. The latter often cause south westerly winds that transport water into the Skagerrak which then becomes trapped inside it as long as the wind continues.

Consequently, by comparing HAT to Max_S in Table 1, we notice that there is a separation line somewhere between Stavanger and Bergen where the contributions from the astronomical tides and the storm surges are of similar order of magnitude. We also observe that regarding the Skagerrak stations the difference between HAT and Lowest Astronomical Tide (LAT) is typically around 60 cm, whereas the stations along the western coast, and further north, has a maximum tidal range typically around 200 cm. As a result, the threshold values shown in three last columns of Table 1 will typically require a higher contribution from storm surge in the Skagerrak area than along the western coast of Norway to be exceeded.

Since we have access to the water level observations in near-real time we also use them to correct our model predictions regarding the storm surge at the 23 water level stations. This may of course be done in several ways, but we have opted to use a simple "weighted differences correction". We compare the forecast and observations for the last five days, calculate the differences for each hour and weight them so that the oldest data is given zero weight, and the newest data the highest weight. The weighted differences is in turn averaged over the five day period and subtracted from the forecast for each station. This

² http://sehavniva.no

Station	Min _T	Min _s	Max_T	Max _s	LAT	HAT	Yellow	Orange	Red
Viker	-101	-82	121	126	-31	29	100	121	139
Oscarsborg	-104	-80	136	139	-35	33	104	124	142
Oslo	-109	-86	143	151	-35	34	107	131	159
Helgeroa	-81	-66	116	118	-29	28	90	109	126
Tregde	-54	-52	90	89	-25	24	70	84	95
Stavanger	-61	-56	112	88	-44	46	80	91	101
Bergen	-106	-40	148	82	-90	89	108	120	139
Måløy	-128	-71	172	88	-114	111	136	148	158
Ålesund	-139	-52	171	90	-120	119	143	159	174
Kristiansund	-144	-48	179	83	-128	132	152	167	180
Heimdal	-183	-73	194	81	-145	151	166	181	199
Trondheim	-176	-53	220	77	-165	184	192	208	220
Rørvik	-164	-54	198	169	-151	154	174	192	208
Bodø	-174	-86	213	91	-166	167	190	209	229
Kabelvåg	-179	-91	226	100	-173	179	206	227	246
Narvik	-189	-211	237	214	-184	196	218	240	266
Harstad	-146	-89	176	80	-135	133	148	163	180
Andenes	-137	-80	172	83	-130	126	149	168	184
Tromsø	-168	-45	198	108	-162	155	174	190	203
Hammerfest	-165	-130	196	83	-168	157	172	188	201
Honningsvåg	-166	-45	194	167	-162	153	170	186	201
Vardø	-195	-149	208	177	-192	180	190	206	223
Ny-Ålesund	-110	-107	98	65	-92	92	104	115	123

Table 1 Minimum and maximum total water level (Min_T, Max_T) and storm surge (Min_S, Max_S) for Norwegian tidal gauge stations.

In addition Lowest and Highest Astronomical Tides (LAT, HAT) are shown. The Yellow, Orange and Red columns show the various threshold levels. We note that the threshold levels for each station are unique. The thresholds for each station was chosen based on the corresponding damage potential due to high water level for that station, and are selected in agreement with local authorities. Typically the Yellow level is close to a water level event with a return period of 2 years, Orange level an event with a 10 year return period and the Red level a return period of approximately 25 years. All units are centimeter, and all reference levels are the mean sea level. The statistics are based on observations of total water level and tidal predictions downloaded from the Norwegian Mapping Authorities API every day for the period 2018-01-01 00UTC to 2021-01-01 00UTC (3 full years). The storm surge component has been calculated by subtracting the predicted tides from the observed total water level

method is rerun every 30 minutes, essentially giving an updated forecast every 30 minutes. This has a large impact on reducing errors in the short term forecast (+0 to +24 h), but also reduce the error on longer lead times of up to 5 days.

4 The storm surge model and the ensemble prediction system

The present model used to predict water levels in Norwegian waters, and their extremes due to storm surges alone, is the ROMS ocean model (version 3.5) [8, 9]. Due to the fact that the storm surge is mainly a barotropic phenomenon, ROMS' ability to run in

barotropic (2D) mode as a single layer model is utilized. This results in a simpler model setup, and even more important, requires less computational power and therefore has shorter computational wall-time compared to a full baroclinic (3D) setup. In turn, these choices are important for our capability to, in addition to the single deterministic model, run a big (probabilistic) ensemble prediction system (EPS). At this point, we would like to emphasize that we have two systems, the deterministic and the probabilistic (or EPS), but the core of both systems is the same dynamical model (for governing equations, see Appendix 1). The EPS, either as an ensemble mean or by the use of an individual member such as the control run, may be used as a deterministic forecast. To avoid any confusion in this regard, we henceforth refer to the term "deterministic model or forecast" as being the model or forecast forced with the highest resolution atmospheric model.

Each model run is initialized from a previous run starting 24 h before the analysis time. This results in a period of 24 h where the model is forced by analyzed atmospheric forcing (wind and pressure) to ensure a best possible initial condition for the forecast. The forecast length produced by both systems is +120 h (i.e. 5 days) from the initialization time. The open boundary conditions are formulated using the Chapman condition for two-dimensional momentum [11] and the Flather condition [12] for free surface. Due to the lack of a, preferably global, outer model to prescribe realistic storm surge at the open boundaries of our domain, the values for two-dimensional momentum $\overline{u}, \overline{v}$ and sea surface deviation ζ are all set to zero using the option of prescribing analytical values in the ROMS source code. We do, however, add the inverted barometer effect ζ^{IB} to the analytic surface deviation at the boundaries to balance the solution in accord with the formula given by [13], that is,

$$\zeta^{IB} = \frac{1}{\rho_0 g} (\overline{p_a} - p_a) \tag{1}$$

where p_a is the air pressure, $\overline{p_a} = 1013, 25$ hPa is the global time mean air pressure, g is the gravitational acceleration and ρ_0 is the density of sea water. The model is forced at the surface boundary by atmospheric pressure and surface momentum fluxes calculated using 10 meter winds taken from an atmospheric model and by use of the Charnock relation [14]. The bottom friction is quadratic, with a drag coefficient of $2.5 \cdot 10^{-3}$. The horizontal grid resolution of the model is 4 km, and the length of the barotropic time-step is set to 10 seconds. The model area cover the North Sea, Nordic Sea and Barents Sea (Fig. 4). The model bathymetry is inherited from a legacy model at MET Norway [7].

4.1 The deterministic system

Our deterministic storm surge model is run twice per day (at 00 and 12 UTC). The highest resolution deterministic model from the European Centre for Medium-Range Weather Forecasts (ECMWF) available to us is used as atmospheric forcing. This setup has been used at MET Norway for decades and was used long before MET Norway switched to ROMS as the modeling tool [7], and thus is the "oldest" component of our system. The resolution of the ECMWF atmospheric model in use today is about 16 km in Norwegian waters and is delivered every 6 h with a temporal resolution of 3 h. The ROMS model utilize a simple linear interpolation to provide atmospheric forcing at every model time step and grid point.



Fig.4 The model domain used in the MET Norway storm surge model. The domain has a polar stereographic projection covering the area from Bretagne to Novaya Zemlya and between Norway and Greenland. Note that the Baltic Sea is not included (masked out) in the model domain. The horizontal model resolution is 4 km. The colors indicate bottom depth in meters, and the contour lines are plotted at 100, 500, 1000, 2000 and 3000 meters

4.2 The ensemble prediction system

Following modern weather forecasting also ocean weather forecasting is moving from just trying to deterministically predict what will happen, into forecasting the uncertainty estimates of the predictions. In order to provide the uncertainty estimates of our forecast, we run a storm surge Ensemble Prediction System (EPS) twice per day (at 06 and 18 UTC). It utilizes all 50+1 members (control + 50 perturbed members) of the ECMWF ensemble prediction system (ECMWF-ENS, [15] and [16]) to provide uncertainties and probabilistic forecasts of storm surge. The ECMWF-ENS is tuned to provide good spread from about 3-5 days lead time [16], given that at this point in time the combination of uncertain initial conditions, and the non-linear nature of the atmospheric models, typically start to produce large errors in deterministic forecasts.

All 51 members of our storm surge EPS are initialized from the same initial state as the deterministic system, and are run concurrently as soon as the ECMWF-ENS model output is available to us. This way of initializing the storm surge EPS has the obvious unrealistic effect that the EPS has predicted zero spread, and hence zero uncertainty, at +0 h lead time, and in turn will probably add to the low ensemble spread at short lead times.

5 Model performance

As alluded to above (Sect. 4) the storm surge forecasting system at MET Norway consist of two almost independent parts; the deterministic system and the ensemble prediction system.

We therefore present an evaluation of the model performance in two parts, one for each system, in the following subsections. In both parts, the differences between model and observations are calculated by subtracting the observations from the model results. The



Fig. 5 Deterministic storm surge forecast statistics showing the ME (top panel), MAE (middle panel) and RMSE (bottom panel) as functions of lead time for the years 2018, 2019 and 2020. The dotted lines show the result when we focus on the larger storm surges, that is, when we neglect, or mask out, events with observed storm surges less than 25 cm (approximately one standard deviation). Lead time in hours is shown along the horizontal axes, while the respective errors in centimeters are shown along the vertical axes

time period used consist of the three full years 2018, 2019 and 2020. Recall that observations here entails the observed tide gauge water levels from which is subtracted the astronomical tides computed by use of harmonic analysis.

5.1 Assessment of the deterministic forecasts

There are many ways to evaluate the performance of the model. We choose to focus on the Mean (or Bias), Mean Absolute and Root Mean Square Errors (ME, MAE and RMSE) as a function of lead time.

By looking at Fig. 5, which shows the statistics for all stations together for each of the three years 2018, 2019, and 2020, we observe that the ME is typically very small for all lead times. The largest deviations is seen towards the end of the five day forecast with an ME of about -1 cm. Regarding the MAE and the RMSE, the errors grow almost linearly with forecast lead time, starting at approximately 3 and 5 cm at +0 h, and ending at 8 and 10 cm at +120 h. Figure 5 also presents the masked data, that is, the statistics when we mask out all the events with an observed surge with an absolute value less than 25 cm (approximately one standard deviation for storm surge for all of the stations). By doing this we get an indication of how well the larger storm surge are predicted compared to the smaller ones. In comparison with the unmasked data we do see that the absolute value of the errors are somewhat larger. This might lead one to speculate whether this indicates that the model forecasts the larger surges better than the smaller ones. To verify whether this is true, we have calculated the relative error as shown by Fig. 6. We use the same criterion for masking of values as in Fig. 5. It is clear that even if the relative error grows with forecast lead time, there is no significant difference between the masked and unmasked data. Thus, we conclude that our model forecasts events with a larger storm surge with the same level of confidence as the smaller ones.



Fig. 6 Relative error for the deterministic model forecast averaged over all stations and all the years 2018, 2019 and 2020. The dotted line indicates the relative error when we only consider data for when the observed surge has an absolute value of more than 25 cm. Lead time in hours is shown along the horizontal axes, while the relative errors in percent are shown along the vertical axes

We also note by looking at Fig. 5 that there is a year to year difference in the absolute errors with the year 2020 showing the largest absolute errors. The reason for this is that the number of large storm surge events was particularly high during 2020 with record breaking water level height at some stations. Finally, we would like to emphasize that there are local differences between the stations (not shown). Even though the sources of errors in the forecast, as visualized by Fig. 5, may be due to errors in the storm surge model itself, we nonetheless believe that most of the error may be attributed to loss in the atmospheric predictability with lead time.

Taking a closer look at Fig. 5 we see evidence of oscillations with a period of around 12 h, which are most revealing when examining the ME. The oscillations vary in amplitude for different stations (not shown), but we note that the amplitude is very small with a maximum around 1 cm for the stations we have studied. This signal may be a result of contamination of the tidal component as discussed in Sect. 2, and hence that the residual tides has not been removed from the observations. To further investigate this, we did a Fourier Transform (using FFT) of the difference between the modeled and observed storm surge as displayed by Fig. 7 (top panel) for all 23 stations. Quite visibly we see that there are signals with periods at exactly 12 hours as well as around the M2 frequency. We also note the signals around the M4 and M6 frequencies. In the next two panels in Fig. 7 we present the FFT of the modeled and the observed, respectively, surge for all 23 stations. Based on all three panels of Fig. 7, we conclude that there is indeed a tidal signal left in the observed surge, and that there is a signal with period of exactly 12 h in the model. At present we can not explain the latter. However, we speculated that it could be resonance in the model due to topographic effects and/or grid size, as it is evident when studying the frequency spectra for each station (not shown here), that the signal is only visible for stations typically located inside long, narrow fjords, or surrounded by complex land mask topography. Another possible explanation is that the 12 h signal could be an effect created by the model restart every 12 h. We therefore did a test where we compared results from a continuous run, with one from a run that was restarted every 12 h. The frequency spectra for both models were the same. We therefore conclude that the model restart does not explain the signal with a 12 h period.

5.2 Assessment of the ensemble forecasts

Before we perform a probabilistic evaluation of the ensemble prediction system, we start by verifying the deterministic behaviour of the ensemble mean and the control



Fig. 7 Frequency spectra of (top panel) the mean error (the difference between the simulated storm surge and the observed storm surge), (middle panel) the simulated storm surge, and (bottom panel) the observed storm surge (the observed total water level minus the predicted tides). Vertical coloured lines mark the periods of (from left to right) M2 (12.42 h, red), 12 h (blue), M4 (6.21 h, cyan) and M6 (4.14 h, magenta)

members, the latter being the member forced with the atmospheric control forecast initiated with the unperturbed analysis. Figure 8 shows the ME, MAE and RMSE for ensemble mean and the control as well as the deterministic model for comparison. The results are very similar to those in Fig. 5, and show that the ensemble mean forecast overall has smaller MAE and RMSE than the deterministic forecast at longer lead times, but the differences is only about 1 cm at +120 h forecast lead time. Although the ensemble mean and control members are not directly used by the decision support system as described in Sect. 6, we include them in the validation to further show that the ensemble prediction system has a similar behaviour and error statistics as the deterministic model. By examining the results in Fig. 8 one could also argue that the forecasting and decision support system should utilize the ensemble mean rather than the deterministic forecast at longer lead times. This has been considered, but not yet implemented in the current system.

Description Springer



Fig.8 ME, MAE and RMSE comparison between deterministic, ensemble control member and ensemble mean averaged over all coastal stations and the years 2018, 2019 and 2020 as a function of lead time. The dotted lines, in the legend denoted as "masked", are the statistics focusing on the larger storm surges as explained by Fig. 5. Lead time in hours is shown along the horizontal axes, while the respective errors in centimeters are shown along the vertical axes

For a perfect ensemble prediction system, the observations and the ensemble members should all be random draws from the same probability distributions. To evaluate the ensemble spread we have created rank histograms [17] (see Fig. 10 and 9). Here, the observation has been given the rank it would have if it was part of the ensemble when sorting all members in increasing order. If the above hypothesis is correct, the observation ranks will be evenly distributed between 1 and the size of the ensemble plus one (in this case 52). This will produce flat rank histograms [17]. A too low spread would produce a "U-shaped" rank histogram where the frequencies of the observation rank is increased for both the highest and lowest ranks. In the opposite event, an ensemble with too much spread, the rank histogram would have an "inverted U-shape", with increased frequencies near the center of the rank histogram. As demonstrated by [18], errors in the observations may artificially result in U-shaped histograms for a ensemble system which has a perfect spread. They propose to compensate for this by adding normally distributed noise to the ensemble members, with a standard deviation given by these observation errors. The observation errors associated with the instruments at the water-level stations are generally less than 1 cm. However, an additional "observation error" is introduced when the pure storm surge signal is calculated by subtracting the tide predictions from the total water level. Based on oscillations on the tidal M2 frequency observed in verification graphs, such as in Fig. 5, we estimate the effective observation error to be 3 cm. Accordingly, normally distributed noise with this standard deviation has been added to the ensemble members.

Figure 9 shows the results for 3 selected stations from day 1 to day 5 forecast range, Oslo, Stavanger and Rørvik, and shows the geographical differences in the ensemble spread. The result for all stations gives a clear U-shaped histogram, indicating too little spread. However, the U-shape is more pronounced for Oslo than at Stavanger and Rørvik, and in general we see that stations located in open areas along the coast, like in this case Rørvik, have a more realistic spread than those located inside fjords and bays. Also, we see a clear tendency that the ranks are a bit more evenly distributed as the



Fig. 9 Rank histograms for a few selected stations, with a normal distributed random error (ϵ) with a standard deviation of 3 cm added to the ensemble members. The stations are, from top to bottom, Oslo, Stavanger and Rørvik. And from left to right forecast lead times of +24, +48, +72, +96 and +120 h. These stations are selected since they represent different geographical areas. Observation rank is given along the x-axis and normalized frequencies along the y-axis. The red horizontal line indicate the frequencies for an unbiased ensemble



Fig. 10 Rank histograms for all 23 stations combined, with a normal distributed random error (ϵ) with a standard deviation of 3 cm added to the ensemble members. Forecast lead times of +24, +48, +72, +96 and +120 h are shown. In general, we can see that there is too little spread in the ensemble, as we see that the observations tend to have low or high rank. Observation rank is given along the x-axis and normalized frequencies along the y-axis. The red horizontal line indicate the frequencies for an unbiased ensemble



Fig. 11 Reliability diagrams for three different forecast lead times (+72h, +96h and +120h) for all stations combined. The grey numbers indicate the number of observations for each of the forecast frequencies in the figure. Diagrams in the left column show forecast reliability for observed storm surge for a threshold above 0.25 meters and the right column above 0.5 meters. The horizontal axis is the forecast probability between 0 and 1 and the vertical axis indicates the observed frequency for each of the forecast probabilities (with a step resolution of 0.1). A perfect ensemble forecast would have all data along the red diagonal line, that is, when, e.g., the model predict a 0.5 (50%) chance of exceeding the threshold, this would be true for 50% of the observations, and the observed frequency would also be 0.5

forecast lead time increase. This applies to all stations. Considering that the ECMWF atmospheric ensemble is targeted to longer forecast ranges this is to be expected.

The ranks for all stations combined is shown by Fig. 10. We clearly see the tendency for the EPS to have too little spread. However, the frequency of the ranks near the edge of the histogram are generally less than 1/3 above the frequencies for the more flat part near the center.

The EPS is used to forecast the probability that an event will occur. A tool commonly used for verification of forecasted probabilities is reliability diagrams [18, 19]. In this case, an event would be for the storm surge to exceed certain thresholds. To evaluate if the forecasted probabilities from our system are consistent with the observed statistical distribution, we have created reliability diagrams for the probabilities that the water level exceed 25 cm and 50 cm, respectively (Fig. 11). We have calculated the reliability for the lead times of +72, +96 and +120 h. All stations are combined in one plot. Note that when we increase the threshold, the number of events (data points) decrease dramatically, and hence we do not see it fit to increase the threshold beyond 50 cm (approximately two standards deviations). The analysis is done according to the method described in [20]. For a given event, the forecast probabilities are split into discrete classes (bins) ranging from zero to one. For each probability class, the fraction of times the event is observed, defined as the observed frequency, is plotted against the corresponding forecast probability. For a perfectly reliable forecasting system, these points lie on the diagonal line. The probabilities have been calculated in the simplest way possible, by dividing the number of ensembles that forecast an event with the total number of ensemble members. We claim that the results in Fig. 11 is fairly close to the diagonal, although in some cases the lines are slightly chopped. The latter is most probably caused by insufficient number of cases. We also see a clear tendency for all lead times, and for both choices of threshold, that the low probabilities occur more often in the observations than in the forecast, and the high probabilities tend to occur more often in the forecast than the observations, yielding S-shaped curves shown by [18] to be characteristic for cases with too low ensemble spread, like the U-shaped rank histograms in Fig. 10.

6 The decision support system and dissemination of forecasts

The storm surge forecast data is made freely available to the public via the MET Norway API³ on a daily basis. The forecasts of storm surge, tides and total water level, together with real-time observations and other statistics are also made freely available to the public in a user friendly fashion through the Norwegian Mapping Authorities web page SeHavnivå⁴ (loosely translated to English as "Look at the water level").

MET Norway have three different alert levels for protective action when issuing warnings of possible dangerous weather events; yellow, orange and red. Red alert events are also called "extreme weather events". The word "extreme", in this context, refer both to the rarity and the potential damage to life, property and infrastructure. After the decision to issue a warning has been made, there are three different "paths" for the warning as shown in Fig. 3; subscription users, key users and the general public.

Subscription users are users that for some reason or another are interested in forecasts of high water level events that may have other alert levels (e.g., marine industry, port authorities or media) than those defined for key users. During a MET Norway red alert, a warning will be sent directly to key users such as the county governor and the Norwegian rescue services. It is the responsibility of MET Norway to issue these warnings and describe the damage potential and severity they represent, but it is the county governors

³ https://api.met.no/weatherapi/tidalwater/1.1/documentation

⁴ https://www.kartverket.no/sehavniva/

Color	Meaning				
Grey	Missing observations				
Green	All ensemble members more than 5 cm below yellow criterion				
Light green	At least one ensemble member less than 5 cm below yellow criterion				
Orange	At least one ensemble member above orange criterion				
Red	At least one ensemble member above red criterion				

 Table 2
 Color codes from the water level dashboard, and their meaning.

The alert level criteria correspond to the various threshold values in the columns Yellow, Orange and Red of Table 1 and are unique and different for each station

who are accountable and therefore has to decide which protective actions or measurements are needed for the given event. In addition, all yellow, orange and red warnings are communicated to the general public via weather forecast in TV and radio, and also via the MET Norway web page⁵.

6.1 The decision support software

The decision support software consist of a dashboard on a internal web page available to the forecaster on duty (Fig. 12). The dashboard has one row per tide gauge station, and one column per day. Each of these cells or boxes are automatically given an alert color based on the forecast of the highest total water level for that day. The colors relates to the different threshold levels for protective action given by Table 1, and their meaning is listed in Table 2.

By clicking on the different colored boxes, the forecaster gets access to the detailed forecasts, including probabilities of exceeding the different alert levels, for each station taken from the EPS system. The forecaster also gets all the latest observations of water level and forecast validation statistics for the last 7 days to quickly assess if the model performance is as expected for the given event. This also enables the forecasters to add value to their forecast through subjective analysis.

Although the storm surge forecasting system itself is fully automated, we emphasize that the actual written warnings that are issued when alert levels are exceeded, are processed and prepared by human forecasters. The design of the storm surge dashboard has been optimized with this in mind, so as to reduce the workload of the forecaster. By taking a quick glance at the dashboard the forecaster is able to expeditiously decide whether further investigations are needed or not.

6.2 Example of an extreme water level event

The extreme weather event named "Elsa" [21] hit the western coast of Norway during the night between Monday the 10th and Tuesday the 11th of February 2020 and resulted in record high water levels (see Fig. 14). A series of low pressure centers hit the southern part of Norway in the period from the evening of February 9th until the 12th. The combination of very low pressure (as low as 945 hPa), and persistent westerly winds over many days,

⁵ https://met.no



Fig. 12 The decision support software dashboard during the extreme weather event "Elsa" in February 2020. The meaning of the different colors is explained in Table 2. There is one row for each station, going from south to north clockwise along the Norwegian coast, and one column per day (5 in total). The color code for a given station for a given day, is set based on the highest water level for that day. By clicking on the colored boxes, the forecaster gets access to detailed forecasts and the latest observations for that station (see Fig. 13). The dashboard is available as an internal web page for the forecaster on duty

resulted in record high water levels on the western coast of Norway. In fact return values for the water level was typically in the range of 200 - 1000 years for all stations on the west coast between Stavanger and Stad. At the Måløy station a new record of 172 cm above mean sea level was recorded, up 5 cm from the old 1993 record. Just north of Stad the station at Ålesund recorded the fourth highest water level during the time of measurements (175 cm, highest ever was 184 cm on January 12 1993). For the Bergen and Stavanger station, the water level was 1 and 2 cm, respectively, below the station records. The second highest water level ever recorded at those stations during a period of more than 100 years.

The event was well predicted by the storm surge forecasting system up to 5 days in advance. The precision and skill of the forecasting system was crucial to stakeholders for making the right decisions. Even though some damage was reported after the episode had passed, the consequences was reduced since the county governors were able to take the needed measures in due time before the event occurred.

The dashboard of the decision support software, as seen by the forecasters two days prior to the extreme weather event "Elsa", is shown by Fig. 12. We also, as displayed by Fig. 13, show an example of the more detailed information available to the forecaster inside of the dashboard at the station Måløy. As is evident the water level this far north is dominated by the tides as shown by the green curve (astronomical tide), the thick black line (observed total water level) and the blue line (total water level forecast). The weather contribution to the total water level, or the storm surge, is the burgundy line. The top right table repeats the information of the graph by showing the numbers. The bottom right table shows the highest and lowest total water level for each forecast day, and the difference between the forecasted maximum value for that day and each of the criteria. The "view mode" depicted in this figure show the results provided by deterministic model. Further view modes, e.g., results from the EPS model, may be selected by the user from this view



Fig. 13 Details from the Måløy station in the decision support software during the "Elsa" extreme weather event. Also shown are the added observations for the entire period. The top left panel shows the forecast water level due to storm surge in burgundy, the tidal prediction in green and the forecasted total water level including both tides and storm surge in blue. The observation of total water level is shown in black. Also shown are the three different criteria as yellow, orange and red horizontal lines, respectively. Time (as Norwegian Normal Time, NNT, is given along the x-axis and water level in centimeters above mean sea level is given along the y-axis. The bottom left panel contains a menu that gives the user options to change between mean sea level and chart datum as reference levels, the possibility to view the forecast with added ensemble predictions, and also to view forecasts for the last seven days. The top right panel contains the same information as the graph, but in tabular text form. The columns, left to right are time (NNT), astronomical tides, storm surge, total water level and observation. All numbers are in centimeter. The bottom right panel show in tabular text form a summary of the critical numbers for each forecast day. The columns are named "Day0" to "Day4", and the rows are, top to bottom, highest total water level, lowest total water level, deviation from yellow criterion, deviation from orange criterion and deviation from red criterion. Positive numbers for deviation indicate total water level above criterion, whereas a negative number indicate total water level below criterion. The bottom two lines of text in the lower right panel indicate how much the forecast was adjusted based on observations, in this case -3 centimeters, and how large the biggest difference between the forecasted and observed total water level was, in this case +6 centimeters

for further investigation into the event. All text in the dashboard is written in Norwegian, since this is the working language at MET Norway. We emphasize that if the forecasters need more detailed information, for instance from the EPS model, this view mode may be selected by the user. During a time of hectic activities in the forecasting room, such as is bound to happen during extreme weather events like "Elsa", this type of graphical overview providing detailed information at a glance, indeed helps to lessen the burden on the forecaster on duty.

7 Summary and some final remarks

Considered is the Norwegian water level forecasting system from observations to dissemination as schematically shown by Fig. 3. The system has evolved over the past 50 years or so. Its main function is to work as an advance warning system so as to mitigate any dangerous events that may threaten life and property along the Norwegian coast. As such it is one



Fig. 14 The modeled water level due to storm surge during the extreme weather event "Elsa" on February 10th 2020 at 22UTC. The color scale gives storm surge in meters. Contour lines are drawn every 0.2 meters

of many parts of the Norwegian system for forecasting dangerous weather events, such as for example heavy precipitation, strong winds and snow storms.

At the core of the system is a well calibrated and validated numerical storm surge model which at present is the Regional Ocean Modeling System (ROMS) run in barotropic mode. Of equal importance in any such system though is the surrounding infrastructure, such as the flow of real-time observations, protocols for dissemination of forecasts to relevant key personnel and authorities responsible for safeguarding life and property, as well as informing the general public. To achieve the latter it is paramount to have an operational forecast-ing center where the water level forecasts produced by the models are monitored 24/7 and necessary warnings are issued.

Above we have therefore focused on the performance of the storm surge model, both what we call the deterministic model and the parallel Ensemble Prediction System (EPS). In addition we have given a detailed account of the dissemination system including a recently developed web based dashboard system that helps to lessen the burden on the fore-casting center, in particular when a dangerous event is about to happen.

To assess the performance of the deterministic model and the EPS we have used observations from the 23 tide gauge stations along the Norwegian coast from the Swedish border in the Skagerrak to the Russian border in the north. By focusing on the three years 2018, 2019 and 2020, we find that the performance of both systems is satisfactory. The mean error (or bias) is approximately 0 cm at +0 h forecast lead time, decreasing to -1 cm at +120 h lead time for the average over all stations. Regarding the mean absolute error and the root mean square error they are, respectively, about 3 and 5 cm at +0 h forecast lead time. In addition we observed that the forecasts produced by both the deterministic model and the EPS showed oscillations at a period which appeared to be close to the M2 tidal frequency. A spectral analysis of the deterministic model results revealed spikes in the spectra not only at the M2 period, but also at the M4 and M6 periods. This is hardly surprising since both of our systems forecasts the storm surge only. To forecast the total water level we simply add the tidal contribution from harmonic analysis. Since the storm surge and the

tides are inseparable such oscillations are bound to crop up when the storm surge contribution is predicted separately. Ideally the model should therefore forecast the total water level, that is, both the tides and the storm surge together, as argued by [10].

More surprisingly the spectral analysis also revealed a spike at the 12 h period, which at first glance may be attributed to the fact that we restart the model every 12 h. However, by examining the spectra from the individual stations separately we found that this was only seen in the spectra from stations located inside long, narrow fjords and typically where the land mask of the model surrounding the station grid point consist of a one or two grid point bay. Since the 12 h period did not show up at the more open ocean stations we thus attributed this to be caused by the complex bathymetry and special grid configurations inside fjords and not to the model restart configuration.

Since the EPS forecasts are trustworthy we also find that it is a useful tool to the forecasters responsible for monitoring possible dangerous events. This makes the forecasters able to assess the probability of a possible damaging event early on. It should be emphasized though that we do see a clear tendency in the reliability diagrams (see Fig. 11) for all lead times that the low probabilities occur more often in the observations than in the forecast, and the high probabilities tend to occur more often in the forecast than the observations. As is also shown in the rank histograms in Fig. 9 and 10, this means that the EPS overall has too little spread. We also note that the current EPS has zero uncertainty in the initial state which is unrealistic. Adding uncertainties also in the initial state for the storm surge EPS could improve the system and add value to the forecast for short term forecasting.

Perhaps the most useful recent advancement regarding the storm surge forecasting and monitoring of possible dangerous storm surge events at MET Norway is the development of the dashboard as illustrated by Fig. 12. This helps the forecaster on duty to quickly and at a glance to assess whether an event might occur, to assess its severity, and most importantly to evaluate whether the event needs more investigations and whether a warning should be issued to key users.

Appendix A The governing equations of the storm surge model

The governing equations of the deterministic storm surge model and the ensemble prediction system are exactly the same, and make use of the option of running ROMS in its barotropic mode only. In this mode these equations conform to the the shallow water equations which may be found in any textbook on the subject ([e.g.,][22],[page 7]). In flux form they are

$$\partial_t \mathbf{U} + \nabla \cdot (\frac{\mathbf{U}\mathbf{U}}{h}) + f\mathbf{k} \times \mathbf{U} = -gh\nabla\zeta + \rho_0^{-1}(\boldsymbol{\tau}_s - \boldsymbol{\tau}_b) + \mathbf{X},$$
(2)

$$\partial_t h + \nabla \cdot \mathbf{U} = 0. \tag{3}$$

Here U is the volume flux in a water column of height $h = H + \zeta$ where H is the equilibrium depth and ζ is the sea surface deviation as illustrated by Fig. 15. Furthermore, τ_s and τ_b denote the wind and bottom stress, respectively, while X represents the internal mixing often referred to as the eddy viscosity. Finally, f is the Coriolis parameter and ρ_0 is the density of sea water. In addition appropriate boundary conditions must be imposed. These, and



further details on the parameterization of \mathbf{X} may be found in, e.g., [8] or on the ROMS web page⁶.

Acknowledgements This research was funded by the Norwegian Meteorological Institute. We would also like to give a special thanks, and give credit, to Anne-Mette Olsen and Merete Øiestad at the MET Norway regional office in Bergen who are the main architects and developers behind the water level dashboard.

Funding Open access funding provided by the Norwegian Meteorological Institute.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

- Baxter PJ (2005) The east coast Big Flood, 31 January 1 February 1953: a summary of the human disaster. Philos Trans R Soc Lond 363A:1293–1312. https://doi.org/10.1098/rsta.2005.1569
- Gerritsen H (2005) What happened in 1953? The Big Flood in the Netherlands in retrospect. Philos Trans R Soc Lond 363A:1271–1291. https://doi.org/10.1098/rsta.2005.1568
- 3. Gjevik B, Røed LP (1976) Storm surges along the western coast of Norway. Tellus 28:166-182
- 4. Johansen S (1959) On the effect of meteorological conditions upon the height of the sea level at the coast of southern Norway. Met Ann, 14
- Martinsen EA, Gjevik B, Røed LP (1979) A numerical model for long barotropic waves and storm surges along the western coast of Norway. J Phys Oceanogr 9:1126–1138
- Blumberg AF, Mellor GL (1987) Three-dimensional coastal ocean models. Coastal and estuarine sciences. In: Heaps N (ed) A description of a three-dimensional coastal ocean circulation model. American Geophysical Union, Washington, D.C., pp 1–16
- Engedahl H (1995) Implementation of the Princeton Ocean Model (POM/ECOM-3D) at the Norwegian Meteorological Institute (DNMI). Research Report 5, Norwegian Meteorological Institute
- Haidvogel DB, Arango H, Budgell PW, Cornuelle BD, Curchitser E, Lorenzo ED, Fennel K, Geyer WR, Hermann AJ, Lanerolle L, Levin J, McWilliams JC, Miller AJ, Moore AM, Powell TM, Shchepetkin AF, Sherwood CR, Signell RP, Warner JC, Wilkin J (2008) Ocean forecasting in terrain-following coordinates: formulation and skill assessment of the regional ocean modeling system. J Comput Phys 227(7):3595–3624. https://doi.org/10.1016/j.jcp.2007.06.016
- Shchepetkin AF, McWilliams JC (2005) The regional ocean modeling system (ROMS): a split-explicit, free-surface, topography-following coordinate ocean model. Ocean Model 9:347–404

⁶ https://www.myroms.org/wiki/Documentation_Portal

- Zijl F, Verlaan M, Gerritsen H (2013) Improved water-level forecasting for the Northwest European Shelf and North Sea through direct modelling of tide, surge and non-linear interaction. Ocean Dyn 63:823–847. https://doi.org/10.1007/s10236-013-0624-2
- Chapman DC (1985) Numerical treatment of cross-shelf open boundaries in a barotropic coastal ocean model. J Phys Oceanogr 15:1060–1075
- 12. Flather RA (1976) A tidal model of the northwest European continental shelf. Mem Soc R Sci Liege 6:141–164
- Ponte RM (2006) Low-frequency sea level variability and the inverted barometer effect. J Atmos Ocean Technol 23(4):619–629. https://doi.org/10.1175/JTECH1864.1
- Charnock H (1955) Wind stress on a water surface. Q J R Meteorol Soc 81(350):639–640. https://doi. org/10.1002/qj.49708135027
- Molteni F, Buizza R, Palmer TN, Petroliagis T (1996) The ECMWF ensemble prediction system: methodology and validation. Q J R Meteorol Soc 122(529):73–119. https://doi.org/10.1002/qj.49712 252905
- Leutbecher M, Palmer TN (2007) Ensemble forecasting (514):31. https://doi.org/10.21957/c0hq4yg78. https://www.ecmwf.int/node/10729
- Hamill TM (2001) Interpretation of rank histograms for verifying ensemble forecasts. Mon Weather Rev 129:550–560
- Sætra Ø, Hersbach H, Bidlot J-R, Richardson DS (2004) Effects of observation errors on the statistics for ensemble spread and reliability. Mon Weather Rev 132:1487–1501. https://doi.org/10.1175/1520-0493(2004)132<1487:EOOEOT>2.0.CO;2
- Wilks DS (2011) Statistical methods in the atmospheric sciences: an introduction, 3rd edn. Academic Press, Oxford, p 704
- Sætra Ø, Bidlot J-R (2004) Potential benefits of using probabilistic forecasts for waves and marine winds based on the ECMWF ensemble prediction system. Weather Forecast 19:673–689. https://doi. org/10.1175/1520-0434(2004)019<0673:PBOUPF>2.0.CO;2
- Selberg LA, Skjerdal MS, Julsrud IR, Øiestad M, Andersen ASH, Kristensen NM (2020) MET Info Hendelsesrapport, Ekstremværet Elsa mandag 10. og tirsdag 11. februar 2020. MET Info 17, Norwegian Meteorological Institute
- Røed LP (2019) Atmospheres and oceans on computers, Fundamental numerical methods for geophysical fluid dynamics, 1st edn. Springer, Cham, p 275. https://doi.org/10.1007/978-3-319-93864-6

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.