DANISH METEOROLOGICAL INSTITUTE

––––– TECHNICAL REPORT –––––

99-23

Verification of Sea Level Forecasts 1998

Jacob Woge Nielsen
Mads Hvid Nielsen

DMI

Copenhagen 1999

ISSN-Nr. 0906-897X
ISBN-Nr. 1399-1388

Verification of Sea Level Forecasts 1998

Jacob Woge Nielsen
Mads Hvid Nielsen
e-mail: jw@dmi.dk , mhn@dmi.dk

Danish Meteorological Institute, Copenhagen, Denmark

Contents

1 Introduction

Sea level forecasts produced by the Danish Meteorological Institute's (DMI's) operational storm surge system *DKSS90* during 1998 are evaluated. Unbiased forecasts valid for the 12 hour time window analysis+06 hours to analysis+18 hours are compared with the observed sea level at 33 Danish coastal tide gauges, with special emphasis on high waters. Range-averaged error measures are presented for each station, and the general forecast quality is described in terms of station-averaged key numbers.

An upgrade setup *DKSS98* is also evaluated. The hydrodynamical model is basically unchanged, but bathymetry maps have higher spatial resolution, bed resistance and wind drag coefficients have been recalibrated, and a Smagorinsky (shear dependent) formulation of the eddy viscosity has been introduced. Also, the tidal boundary forcing has been revised. The upgrade setup has been run in hindcast mode for 1998. DKSS98 finally replaced DKSS90 September 1st 1999.

Outline: This report includes a short description of the operational storm surge warning system and of the upgrade setup (Ch. 2), an overview of the data (observations and forecasts) including an estimate of the return period of the most severe events (Ch. 3), and definitions of the statistical measures and key numbers used to describe the forecast quality (Ch. 4). Ch. 5 summarises the 1998 operational forecast quality, and Ch. 6 compares with previous years. Ch. 7 describes the quality of the upgrade setup. Ch. 8 concludes the work.

Lists of References, Figures and Tables are found at the end of the report.

This is one in a series of reports on verification of the operational storm surge system at the Danish Meteorological Institute. Two types of reports are issued on this subject, this one being a verification of a calendar year, while the second type describes a storm surge event. Previous and related reports may be found in the reference list at the end of the report.

2 The storm surge warning system

Operational setup DKSS90:

The storm surge warning system DKSS90 has been in operational use since October 1990, and has been verified since winter 1993-94 (report list in References). DKSS90 has three major components:

- a depth-integrated hydrodynamical model *System 21*
- an atmospheric forcing model
- open boundary tidal elevations

System 21 was developed by the Danish Hydraulic Institute (DHI, see [2] for a scientific documentation), and kindly made available to DMI for operational use. The model domain (Fig. 1, upper panel, and Table 1, left-side columns) consists of four computational domains, with high resolution of the narrows, belts and straits of the Wadden Sea and the Belt Sea. This nesting method, with full dynamical coupling of the model domains, allows the North Sea and the Baltic to be interconnected.

Operational (DKSS90)		Upgrade (DKSS98)	
Resolution	Region	Resolution	Region
18520 m	North Sea - Baltic Sea	16668 m	North Sea - Baltic Sea
6173 m	Trans. Area	5556 m	Trans. Area - Baltic Proper
2058 m	Belt Sea	1852 m	Belt Sea - Western Baltic
2058 m	Wadden Sea	1852 m	Wadden Sea

Table 1: Computational grids, DKSS90 and DKSS98. See also Fig 1.

The atmospheric forcing consists of mean sea level pressure (mslp) and 10m wind, obtained from a numerical weather prediction (NWP) model and interpolated onto the coarsest model grid. This constitutes the surface boundary condition for the hydrodynamical model, which converts surface wind to stress using a wind speed-dependent drag coefficient. Two setups of DKSS90 are run in parallel, differing only in the choice of forcing model. The *main* setup is forced by DMIs limited area NWP model D15-HIRLAM (D15, sometimes referred to as the E-model), with time/space resolution of 1 hour/0.15^o. New versions of HIRLAM are being used in DKSS90 when possible. This is the principal method of improving the system. The *backup* setup is forced by the British NWP model UK-LAM (UKM), which has a much coarser temporal and spatial resolution, 6 hours/1.25^o[1]. To make up for this coarseness, the UKM wind speed is augmented artificially by 10% over the North Sea and by 1-2 m/s over the Belt Sea. The backup setup provides the duty forecaster with extra general information, and serves as principal forecast in case main forecasts are for some reason unavailable. Both setups are fully operational, and both are verified and intercompared in this report.

The tidal open boundaries are positioned along 51^{o}N and 59^{o}N, respectively. Ten tidal constituents are used, $M_2, S_2, N_2, K_2, \mu_2, K_1, O_1, P_1, M_4, MS_4$. These have established by the DHI on the basis of detailed German tables compiled during Second World War (Oberkommando der Kriegsmarine, 1943) plus other, more recent information [12].

[1] The UKM is highly resolved, but DMI receives model fields in coarse resolution.

DKSS90 makes no use of the observed model state, in terms of tide gauge or current measurements.

Schedule:

A DKSS90 run produces a prediction of surface elevation and vertical mean current on each of the 4 model grids. The prediction consists of a *hindcast* (based on meteorological analyses meteorological and 1-5 hour forecasts) plus a short-range *forecast*. Medium-range forecasts to a maximum of 7 days ahead may be produced by supplementing D15/UKM with a coarser global forcing model obtained from *ECMWF*. These forecasts will, of course, in general be of lower quality. During 1998, DMI provided the Sound Link Construction Company with medium-range (5 days) current and sea level forecasts.

The schedule is shown in Fig. 2. DKSS90 is run twice daily, at 00 UTC and 12 UTC. Both runs start at 00 UTC the previous day, initialised by the same model state from a previous run. The 12 UTC run thus repeats the hindcast part of the preceding 00 UTC run. For the main setup only, the 24 hour hindcast part of the 00 UTC run is stored on tape. The hindcast is followed by a forecast of minimum length 36 hours. The model time step is 10 minutes, and the model state output time step is 30 minutes.

If the new NWP model output is not available within 5 hours from analysis time, DKSS90 uses substitute forcing (old forecasts and/or forecasts from the global forcing model). Such forecasts are not use for storm surge warning, and are also not verified.

For further information on DKSS90, see [12].

Upgrade setup DKSS98

An upgrade of DKSS90 termed *DKSS98* has been calibrated and run in hindcast mode for all of 1998, using 3-hourly D15 forcing (alternating hindcast and forecast)[1]. Work on DKSS98 was initiated late 1996, but due to errors in the tidal boundary forcing, discovered during 1998, sea level predictions were at first less good than the operational ones. For this reason, the system was not put in operational use during 1998.

DKSS98 uses a new hydro-dynamical model Mike21 (likewise developed by the DHI). A slightly refined grid configuration and bathymetry (see Fig. 1, lower panel, and Table 1, right-side columns) and revised tidal boundaries are introduced. All friction terms, including the wind drag coefficient, have been recalibrated. The model time step has been decreased from 10 min. to 5 min., and the output time step from 30 min. to 15 min.

The data-assimilation setup *DMIsurge* described in [3], is no longer in use. The system never managed to produce decent results. A data-assimilation scheme has been implemented in Mike21 by DHI but that version has not been released yet (DHI, pers. comm.).

The complete 1998 DMI storm surge setup inventory shown in Table 2.

[1] The retrieval of NWP model output from tape, plus postprocessing to generate the mslp, is very timecomsuming, one reason being that has full D15 model state has to be retrieved from tape for every time step. 1-hourly forcing is of course better than 3-hourly, but it was considered of minor importance since a direct comparison with operational forecasts is not done anyways.

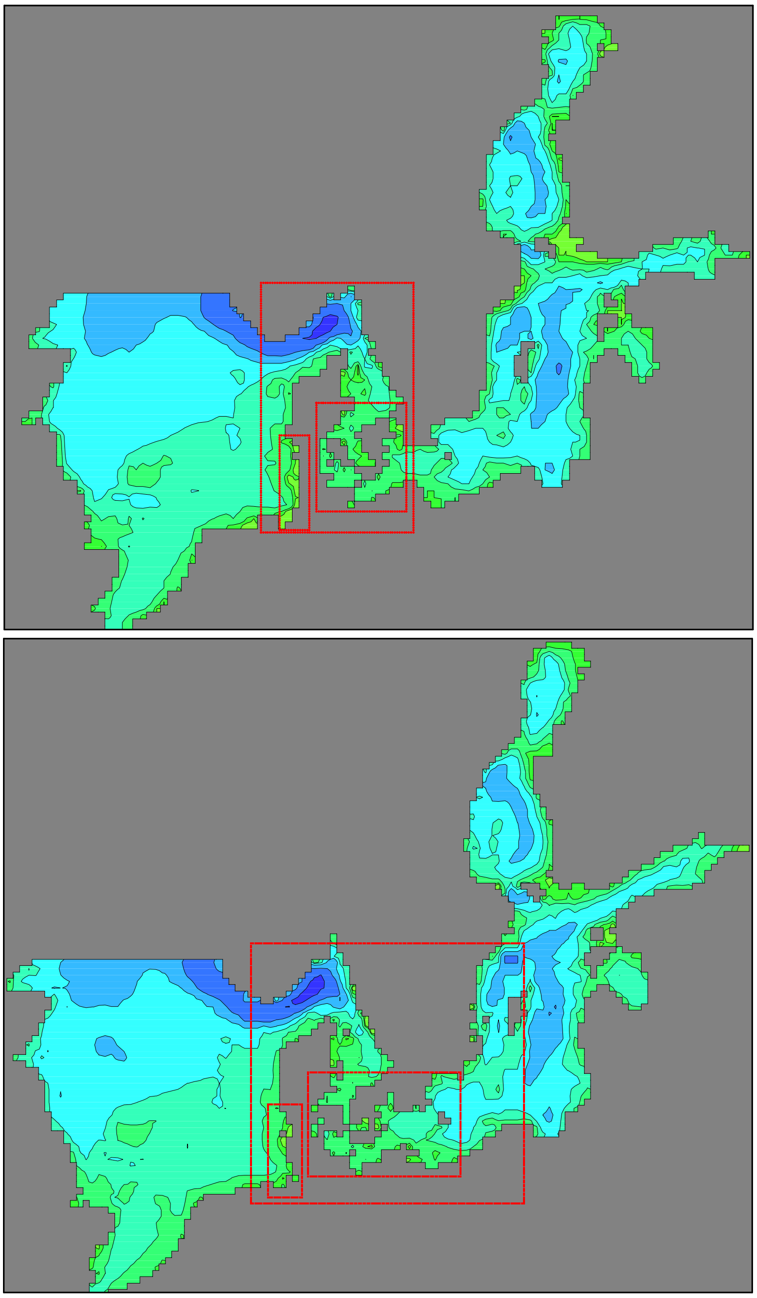


Figure 1: Upper: The DKSS90 model domain with nested grid arrangement. The model has four computational grids and two open tidal boundaries, at the Shetland-Bergen cross-section (59°N) and at the Dungeness-Wissant cross-section in the British Channel (51°N). Lower: In the upgrade setup (cf. Ch. 3), the Domestic model domain extends eastwards of Gotland, while the Belt Sea model domain is extended eastwards to a section just east of Bornholm.

Setup	Main/D15	Back-up/UKM	Upgrade/M21
Schedule name	DKSS90	DKSS90	DKSS98
Mode of run	Operational	Operational	Hindcast
Hyd. model	System 21	System 21	Mike 21
Atm. model	D15-HIRLAM	UK-LAM	D15-HIRLAM
Model time step	10 min.	10 min.	5 min.
Output time step	30 min.	30 min.	15 min.

Table 2: Storm surge setups at DMI, 1998.

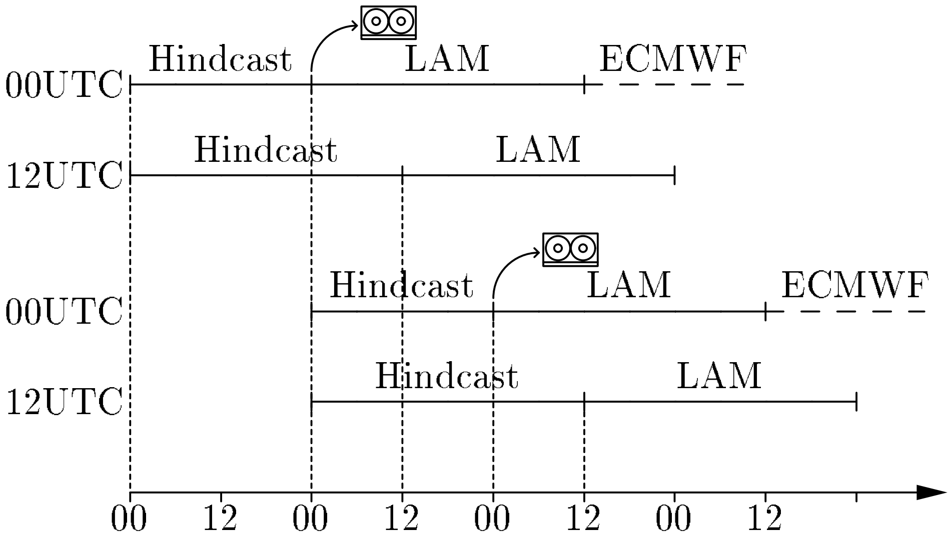


Figure 2: Schedule of the operational runs. 00 UTC runs store a 24 hour hindcast with 30 min. time resolution on tape. The runs may be extended to a maximum range of 7 days using ECMWF atmospheric forcing fields.

3 Verification data

The storm surge system produces predictions of sea level and vertical mean current. Only sea level predictions are verified in this report.

Observations:

Sea level from 33 automatic Danish tide gauge stations are used for verification. Station locations are shown in Fig 3 and Table 3.

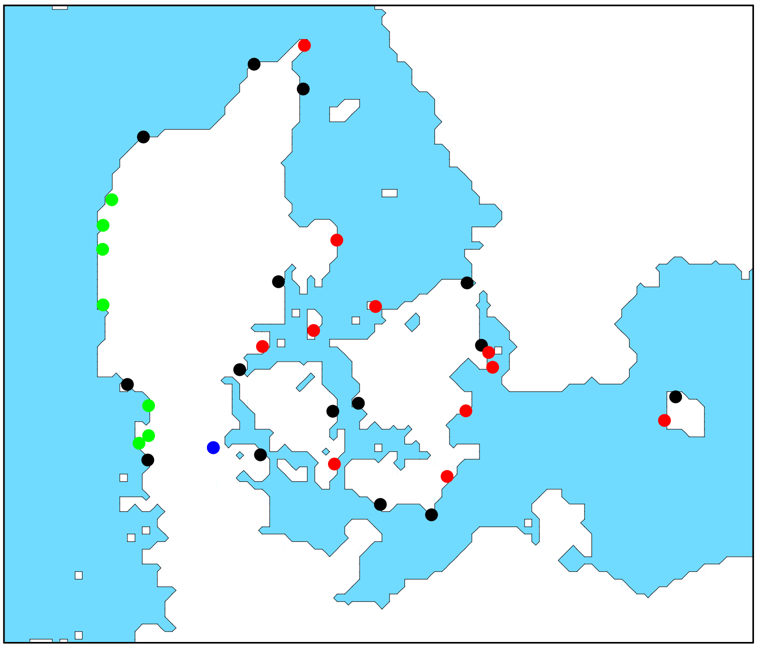


Figure 3: Location of verification stations. Black=DMI station, red=RDANH, green=KI, blue=AMT. The coastline represents the Transition Area model domain. On-line observations, tidal sea level and 12-hour forecasts from all stations may be found on the Internet Maritime Service address *http://www.dmi.dk/vejr/vandstand*. See also Table 3.

Of the stations used,

- 15 are operated by DMI
- 10 are operated by Royal Danish Academy of Navigation and Hydrography (RDANH)
- 7 are operated by the Coastal Authorities (KI)
- 1 is operated by a local authority (AMT)

The tide gauges measure the sea level relative to a local datum, with an accuracy of 1 cm and an observation frequency of 15 min. The RDANH stations do not record regularly at minutes 00, 15, 30 and 45, and the recording time interval may also be slowly drifting. The timestamp of those records is rounded to the nearest full quarter, resulting in a phase change of up to 7 minutes.

Station		Number	Position	Region	Owner
Skagen	n)	20002	$57^o43'N$ $10^o36'E$	Skagerrak	RDANH
Hirtshals		20047	$57^o36'N$ $09^o58'E$	Skagerrak	DMI
Frederikshavn		20101	$57^o26'N$ $10^o34'E$	Skagerrak	DMI
Hanstholm		21009	$57^o07'N$ $08^o36'E$	Skagerrak	DMI
Grenå	n)	22121	$56^o25'N$ $10^o56'E$	Kattegat	RDANH
Århus		22331	$56^o09'N$ $10^o13'E$	Kattegat	DMI
(Juelsminde)	n)	23132	$56^o43'N$ $10^o01'E$	Belt Sea	RDANH
Fredericia		23293	$55^o34'N$ $09^o45'E$	Belt Sea	DMI
Thyborøn		24006	$56^o42'N$ $08^o13'E$	West Coast	KI
Ferring	n)	24018	$56^o32'N$ $08^o07'E$	West Coast	KI
Torsminde		24122	$56^o22'N$ $08^o07'E$	West Coast	KI
Hvide Sande		24342	$56^o00'N$ $08^o08'E$	West Coast	KI
Esbjerg		25149	$55^o28'N$ $08^o26'E$	Wadden Sea	DMI
Ribe Sluse	*)	25343	$55^o20'N$ $08^o41'E$	Wadden Sea	KI
Havneby		26136	$55^o05'N$ $08^o34'E$	Wadden Sea	KI
Åbenrå		26239	$55^o03'N$ $09^o26'E$	Western Baltic	AMT
Ballum Sluse	*)	26346	$55^o08'N$ $08^o41'E$	Wadden Sea	KI
Vidå Sluse	*)	26359	$54^o58'N$ $08^o40'E$	Wadden Sea	DMI
Fynshav		26457	$55^o00'N$ $09^o59'E$	Western Baltic	DMI
Ballen	n)	27084	$55^o49'N$ $10^o38'E$	Kattegat	RDANH
Slipshavn		28233	$55^o17'N$ $10^o50'E$	Belt Sea	DMI
Spodsbjerg	n)	28582	$54^o56'N$ $10^o50'E$	Belt Sea	RDANH
Odden	n)	29002	$55^o58'N$ $11^o22'E$	Kattegat	RDANH
Korsør		29393	$55^o20'N$ $11^o08'E$	Belt Sea	DMI
Hornbæk		30017	$56^o06'N$ $12^o28'E$	Kattegat	DMI
København		30337	$55^o41'N$ $12^o30'E$	Belt Sea	DMI
Nordre Røse	n)	30346	$55^o38'N$ $12^o41'E$	Belt Sea	RDANH
Drogden Fyr	n)	30357	$55^o32'N$ $12^o43'E$	Belt Sea	RDANH
Rødvig	n)	31063	$55^o15'N$ $12^o23'E$	Western Baltic	RDANH
Hesnæs	n)	31493	$54^o49'N$ $12^o08'E$	Western Baltic	RDANH
Rødby		31573	$54^o39'N$ $11^o21'E$	Western Baltic	DMI
Gedser		31616	$54^o34'N$ $11^o56'E$	Western Baltic	DMI
Tejn		32048	$55^o15'N$ $14^o50'E$	Baltic	DMI
Rønne	n)	32096	$55^o06'N$ $14^o41'E$	Baltic	RDANH

Table 3: Verification stations. Station number, name, position and DKSS90 grid. Owner DMI=Danish Meteorological Institute; KI=Coastal Authorities, RDANH=Royal Danish Academy of Navigation and Hydrography, AMT = a local authority. *) = only used for peak statistics, n) = new since 1997. Juelsminde (in brackets) is not used for verification. A new tide gauge at Kolding will be included in the next report.

For verification, data subsets with half-hour time resolution are used. In 1998, all DMI stations were recalibrated to record sea level relative to the common datum DNN (Danish Normal Null). The subject of reference level and unbiasing is discussed further below.

Year	Coverage	Stations	<95%	Added
1994	96.9	17	4	(none)
1995	93.4	21	3	Havneby, Ribe, Vidå, Ballum
1996	91.1	21	7	(none)
1997	96.4	22	4	Aabenraa
1998	98.4	33	1	Ferring, 10 RDANH stations

Table 4: Observation data coverage (%) 1994-98. Also shown are stations added every year.

The 1998 data coverage (Table 4) is very high, with a bulk coverage of 98.4%. 26 stations have a data coverage above 98%, and DMI stations alone have an impressive average data coverage of 99.3%. Two RDANH stations have a data coverage below 95%:

- *Juelsminde:* 44%. Observations available since ultimo August 1998. Not used for verification.
- *Spodsbjerg:* 85%. Malfunctioning during April+May. Used for verification anyway since this is not during the stormy season.

Extreme sea level and surges:

The highest and the lowest sea level recorded at each station during 1998 are shown in Table 5. The Wadden Sea sluice stations (indicated by a *) in Table 3), Vidå Sluse, Ribe Kammersluse, and Ballum Sluse, have very non-linear behaviour at low waters, when the sea level is affected by locking/opening of sluice gates. This is not reproduced by the model. These stations are only used for verification of forecasts of very high waters (peaks).

Return periods for the highest 1998 sea level at some of the stations are estimated. For the estimation method, see [1],[11]. The results for DMI stations are shown in Table 5 and Fig. 4. A return period of two years or less indicates that the highest sea level was below median severity. Six stations have return periods of 3 years or more: Hirtshals, Hanstholm, Esbjerg, Slipshavn, Korsør and Gedser.

Region and date of the most severe 1998 surges are shown in Table 6. During 1998, DMI had warning responsibility for 7 stations (Gedser serves as backup station for Rødby). The warning thresholds, and the number of exceedances during 1998, are shown in Table 7.

Forecasts:

Each DKSS90 run produces and archives sea level predictions for all Table 3 stations[1]. Model grid points representing the stations have been carefully selected; usually (but not always) the grid point nearest the geographical position of the station is used. A prediction consists of a hindcast of variable length plus a 36 hour forecast, with a time resolution of 30 minutes. The 12 hour time window analysis+06 hours to analysis+18 hours (considered

[1] Sea level forecasts for a number of non-Danish stations are also archived, but not verified.

Station	Sea level (cm)		Return Period (yrs)	
	Lowest	Highest	DMI	KI
Skagen	-74	97	-	*3.2*
Hirtshals	-92	104	3.8	*1.6 *)*
Frederikshavn	-75	100	2.0	*1.4*
Hanstholm	-84	136	3.4	*3.7*
Grenå	-87	107	-	*<3.8*
Århus	-81	97	1.5	*<4.5*
Fredericia	-69	100	2.1	*<2.8*
Thyborøn	-125	202	-	-
Ferring	-115	212	-	-
Torsminde	-94	216	-	*2.1*
Hvide Sande	-131	258	-	-
Esbjerg	-202	297	4.4	-
Ribe K.sluse	-125	339	-	-
Havneby	-205	307	-	*<2.8*
Aabenraa	-86	126	-	*2.2*
Ballum Sluse	-15	329	-	-
Vidå Slusen	-85	326	1.9	*2.0*
Fynshav	-78	127	1.7	*1.0*
Ballen	-75	87	-	-
Slipshavn	-64	116	6.5	*5.0*
Spodsbjerg	-79	136	-	-
Odden	-83	86	-	-
Korsør	-40	102	3.0	*<2.5*
Hornbæk	-73	106	1.5	*<1.0*
København	-73	100	2.2	*1.4*
Nordre Røse	-28	97	-	-
Drogden	-66	93	-	-
Rødvig	-86	101	-	-
Hesnæs	-84	101	-	-
Rødby	-81	122	2.4	*1.2*
Gedser	-68	121	3.0	*1.9*
Tejn	-39	74	1.7	-
Rønne	-52	71	-	*1.1*

Table 5: Extreme sea level (cm) and return period (years), 1998. The DMI estimate is based on a Gumbel distribution and annual extremes, the KI estimate is based on a Weibull distribution and a peak over threshold (POT) method. "-" indicates no return period is calculated, "<" indicates that the return period could not be estimated but was below the indicated value. The KI estimate at Hirtshals *) was calculated using a different data set than the DMI estimate.

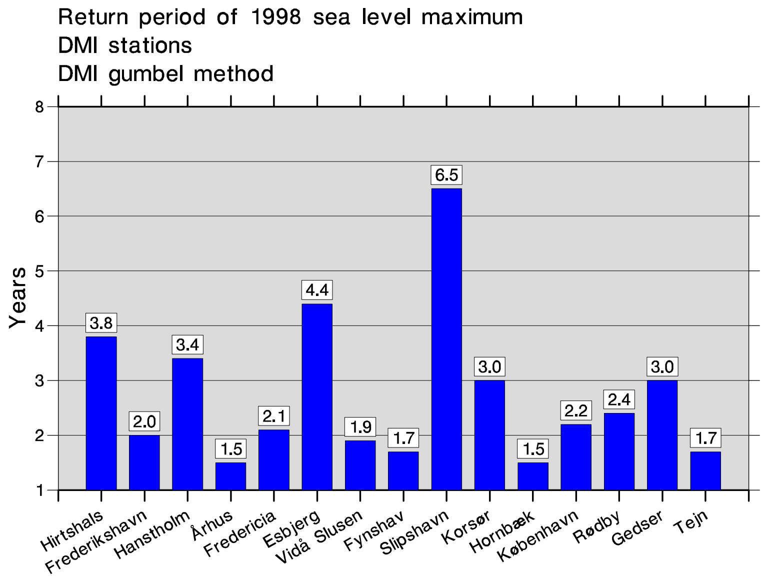


Figure 4: 1998 return periods at DMI stations.

the most important for storm surge warning purposes) of all forecasts are concatenated to make monthly pseudo time series. The remainder of the forecasts is ignored in this context.

For both setups, roughly 9% of the forecasts are missing. All October forecasts were lost due to a tape device error. This is particularly unfortunate since the two most severe storm surges were experienced during that month. For the main (D15) setup 00Z hindcasts have been substituted for October forecasts, leaving 2.5% (19) still missing. For the backup (UKM) setup hindcasts were not archived, so no data is available for that month.

Forecast postprocessing:

a) Unbiasing:

In general some mean error (bias) persists at individual stations. This is corrected for by shifting the forecast by the negative of this amount (unbiasing). The bias may vary from year to year; during 1998 three-year mean errors (1994-96) were used (Table 8, left-side column). This makes the forecasts more base level neutral relative to the station datum, leading to better forecasts on average (but not in every case). The improvement is substantial at stations with a large stable bias, but the forecasts may get worse at stations where the mean error has a large year-to-year variation [3]. For the upgrade setup, the 1998 hindcast mean error is used as bias (Table 8, right-side column).

b) Filtering:

In a practical warning situation, sea level forecasts are filtered using an autoregressive (AR) model for the forecast error. The discrepancy between the model prediction and the most recent observations is used to model the forecast error at future times, and the 'raw' forecast

Date	Region
January 4th	Wadden Sea, West Coast
January 20th	Bornholm
January 24th	Western Baltic
January 30th-31st	Belt Sea, Western Baltic
February 17th	Western Baltic
October 18th	Southern Kattegat
October 25th-26th	Wadden Sea, West Coast, Skagerrak, Kattegat
October 28th-29th	West Coast, Skagerrak, Kattegat
November 6th	The Sound
December 27th-28th	Wadden Sea, West Coast

Table 6: 1998 major storm surges.

Station	Thresh.	1994	1995	1996	1997	**1998**
Frederikshavn	90	0	2	2	2	**2**
Torsminde	200	0	0	0	1	**3**
Esbjerg	250	3	1	2	2	**3**
Aabenraa	100	-	-	-	2	**4**
Vidå Sluse	250	4	6	5	2	**4**
Slipshavn	100	0	2	1	1	**1**
Korsør	100	0	2	0	1	**1**
Rødby	130	0	2	0	0	**0**
Gedser	130	0	2	0	0	**0**

Table 7: Number of times the warning threshold was exceeded, 1994-98. The tide gauge at AAbenraa was established during 1996.

is corrected by this amount. The correction usually takes the shape of an exponential decay, possibly with a sine wave superimposed, but the exact nature of the filter varies from station to station. The filter constants have been calculated using hindcast errors from a previous year. The AR procedure is repeated every time a new sea level observation arrives, resulting in a forecast update every 15 min. For simplicity, only unfiltered forecasts are considered in this report.

Station	S21	M21
Skagen	no	-18
Hirtshals	-18	-18
Frederikshavn	-11	-11
Hanstholm	-10	-7
Grenå	no	2
Århus	-4	3
Fredericia	2	9
Thyborøn	-2	0
Ferring	no	1
Torsminde	2	3
Hvide Sande	7	9
Esbjerg	5	5
Ribe Sluse	7	-1
Havneby	-4	5
Aabenraa	no	17
Ballum Sluse	no	no
Vidå Sluse	5	18
Fynshav	8	19
Ballen	no	3
Slipshavn	-3	3
Spodsbjerg	no	0
Odden	no	-1
Korsør	1	13
Hornbæk	-7	4
København	0	-3
Nordre Røse	no	17
Drogden	no	8
Rødvig	no	12
Hesnæs	no	14
Rødby	15	16
Gedser	2	16
Tejn	-2	8
Rønne	no	3

Table 8: Biases used to postprocess sea level predictions (cm). S21 = operational bias (D15+UKM) based on 1994-96 average mean errors. M21 = upgrade bias based on 1998 hindcast mean errors. 'no' = no unbiasing done for this station. The tabulated value is added to the prediction.

4 Error measures

Previous studies [4] have shown that the forecast quality does not depend very much on forecast range. For instance, hindcasts are no better than short-range forecasts. The main reason is the lack of model initialisation using observed sea level. Therefore, only range-averaged error measures are considered below. Two types of error measures are used (Fig. 5):

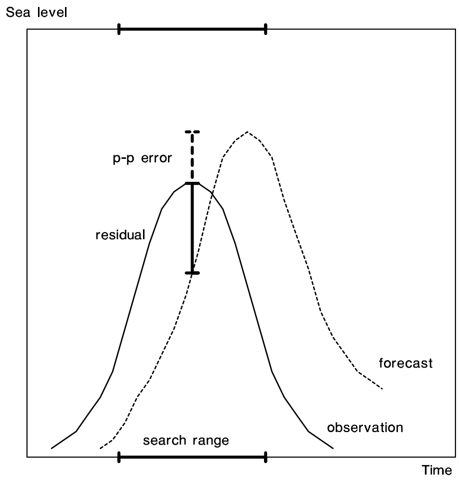


Figure 5: Definition of the residual and the peak error.

- *residual error measures*, based on forecast minus observation at a given time
- *peak-to-peak (p-p) error measures*, based on forecasted minus observed sea level high

The p-p error allows for a phase error less than the search range = 6 hours, but the phase error itself is not verified. If the nearest forecasted peak lies outside the search range, it is discarded. If more than one peak is forecasted within the search range, the highest peak is chosen.

Range-averaged monthly error measures are calculated for each station, using forecast pseudo time series as described in Ch. 4. Annual mean or extreme values of the monthly error measures are calculated for each station, with equal weight on each month. Finally, grand averages are computed with equal weight on every station.

The residual

Residual statistics are based on the unbiased forecast error time series. We define monthly error measures me = mean error, mae = mean absolute error, rms = root mean square error, max = maximum error and ev = explained variance by

$$\begin{aligned} res(i) &= frc(i) - obs(i) \\ me &= \overline{res} \\ mae &= \overline{|res|} \end{aligned}$$

$$
\begin{aligned}
rms &= \sqrt{\frac{N}{N-1}\overline{res^2}} \\
max &= max(|res(i)|), i = 1, ..., N \\
ev &= 1 - \frac{Var(res)}{Var(obs)}
\end{aligned}
$$

where overbar denote one-month average, N is the series length and $Var(x) = \overline{x^2 - \overline{x}^2}$. Each error measure is calculated for every station, and for every month. Annual error measures for each station are calculated as twelve-month averages. These are denoted by upper-case letters ME, MAE, RMS, MAX, EV. The MAX error has been redefined since 1997, when a grand maximum was used instead of a 12-month average. Finally, annual error measures are obtained by averaging over all stations, denoted Average ME, etc. One additional measure Average AME is defined as the station average of absolute annual mean errors.

The peak-to-peak error

Predictions of extreme (high) water levels are verified using the p-p error (Fig. 5)

$$
pe(t_p) = frc(t) - obs(t_p)
$$

where t_p is the time of the observed peak, t is the time of the predicted peak, and the phase error $|t - t_p| < 6$ hours, corresponding to half the diurnal tidal period. pe is only defined at times of observed peaks, and two observed peaks must be separated in time by at least 6 hours. Otherwise, the lowest peak is discarded.

For each station, the 10 highest sea level peaks are identified and the set $pe(t_i), i = 1, .., 10$ is calculated. Annual peak error measures $MPE =$ mean peak error (or peak bias), $MAPE =$ mean absolute peak error, and $MXPE =$ maximum absolute peak error are defined by (overbar indicate average over 10 highest peaks)

$$
\begin{aligned}
MPE &= \overline{pe} \\
MAPE &= \overline{|pe|} \\
MXPE &= MAX(|pe_i|), i = 1, ..., 10
\end{aligned}
$$

Grand peak error measures, termed Average MPE, etc., are obtained by averaging over all stations. One additional measure Average $AMPE$ is defined as the station average of absolute annual mean peak errors.

Extreme low waters or false alarms are not verified.

With a given success criterion, a simple count of the number of well predicted peaks is done. In 1998 the data set totals 330 observed peaks. The resulting *peak hit rate*, with equal weight on all stations, is

$$
PHIT = \frac{\#errors < criterion}{\#peaks} * 100\%
$$

Success criteria of 10cm and 20cm are employed.

5 Results

Annual averages of the error measures for each station are presented in Tables on the following pages. Operational unbiased D15 and UKM forecasts, and upgrade (M21) hindcasts are included. The operational results are discussed in relation to the Tables, while the upgrade results are discussed in Chapter 7. The D15 results before unbiasing are discussed where considered relevant, but are only shown as grand averages. D15 hindcast error measures have been calculated and compared with D15 forecasts. No major differences were found, so only forecasts are discussed in detail.

Tables 9-13 show residual statistics, Tables 14-16 peak statistics for each station. Figs. 6-14 show peak errors at the warning stations. Grand averages (key numbers) and peak hit rates are given in Table 18 at the end of this Chapter, including D15 hindcasts and unbiased upgrade forecasts. Key numbers have been calculated in previous years (1994-97), and comparison with this is made in Chapter 6.

The mean error (ME) (Table 9)

Station	D15	UKM	M21
Skagen	16.3	17.1	17.8
Hirtshals	-3.0	-2.8	16.8
Frederikshavn	-1.3	-0.5	11.1
Hanstholm	-3.5	-2.9	8.5
Grenå	-3.4	-2.4	-2.0
Århus	-8.5	-7.5	-3.6
Fredericia	-5.5	-4.0	-8.9
Thyborøn	-2.0	-0.9	0.4
Ferring	0.5	1.7	1.2
Torsminde	-1.2	0.1	-2.6
Hvide Sande	-3.4	-1.4	-9.4
Esbjerg	5.7	6.4	-1.2
Havneby	-6.8	-6.0	-4.9
Aabenraa	-11.5	-9.9	-16.6
Fynshav	-5.6	-3.7	-18.7
Ballen	-3.5	-2.3	-2.9
Slipshavn	-5.5	-3.9	-3.5
Spodsbjerg	2.8	4.5	0.7
Odden	0.7	1.5	0.8
Korsør	-9.5	-8.4	-12.6
Hornbæk	-11.3	-10.5	-4.1
København	-3.3	-2.5	-5.3
Nordre Røse	-14.4	-13.8	-17.1
Drogden	-0.8	0.6	-7.9
Rødvig	-3.2	-1.9	-11.5
Hesnæs	-5.8	-4.7	-13.9
Rødby	6.6	8.2	-15.6
Gedser	-5.2	-3.9	-15.5
Tejn	0.4	2.5	-8.1
Rønne	7.6	10.0	-2.6

Table 9: Average mean error, 1998. D15/UKM = main/backup setup, M21 = upgrade setup.

The ME (or bias) ranges roughly ±15 cm for both the D15 and the UKM setup. The UKM forecasted sea level is up to a few cm higher than the D15 forecasts on average. Unbiasing (using Table 8) has removed a large part of the ME formerly present at Hirtshals, Frederikshavn, Hanstholm, Hvide Sande, Fynshav and Rødby (see [3]), while increasing the ME by 7 cm at Hornbæk, less at other stations. The stations with ME in excess of 10 cm, Skagen, Aabenraa, Nordre Røse and Rønne, are new and not yet unbiased. The absolute average ME is 5 cm for both setups and unbiasing has not lowered this figure significantly. For the D15 setup, 14 stations have small ME (<5 cm), for the UKM setup 20 stations. The geographical distribution of the ME is more or less random.

The mean absolute error (MAE) (Table 10)

Station	D15	UKM	M21
Skagen	17.4	18.1	9.9
Hirtshals	9.8	10.2	9.9
Frederikshavn	10.4	10.6	10.4
Hanstholm	9.8	9.9	9.4
Grenå	11.0	11.8	11.1
Århus	12.8	13.0	11.1
Fredericia	10.0	10.0	9.6
Thyborøn	12.0	12.1	9.4
Ferring	12.2	12.4	10.8
Torsminde	13.8	14.2	11.8
Hvide Sande	13.1	12.7	12.8
Esbjerg	14.3	14.4	16.3
Havneby	14.7	14.5	16.7
Aabenraa	13.8	12.8	9.7
Fynshav	10.7	10.0	9.5
Ballen	10.1	10.6	10.2
Slipshavn	10.6	10.6	10.0
Spodsbjerg	9.5	10.0	9.6
Odden	9.0	9.8	9.7
Korsør	12.4	12.1	9.6
Hornbæk	13.1	13.6	9.2
København	11.7	11.8	11.0
Nordre Røse	15.5	15.3	8.6
Drogden	7.7	7.6	7.7
Rødvig	7.8	7.5	7.6
Hesnæs	9.0	8.8	7.9
Rødby	10.0	10.4	8.5
Gedser	8.9	8.3	8.0
Tejn	6.4	6.4	7.8
Rønne	9.1	10.4	8.1

Table 10: Average mean absolute error, 1998. D15/UKM = main/backup setup, M21 = upgrade setup.

The MAE ranges from 6 cm (Tejn) to 18 cm (Skagen) for both the D15 and the UKM setup. The D15-UKM difference does not exceed 1 cm at any station. Low MAE (<10 cm) is found at stations in the Western Baltic, at Odden, Spodsbjerg and the Skagerrak stations Hirtshals and Hanstholm. High MAE (>15 cm) is found at Skagen and Nordre Røse. These stations are not yet unbiased. Unbiasing has removed a large part of the MAE at Hirsthals, Frederikshavn and Fynshav, but increased the MAE at Hornbæk. The MAE of D15 hindcasts (not shown) is equal to the forecasts within a few mm.

The root mean square error (RMS) (Table 11)

Station	D15	UKM	M21
Skagen	20.3	21.0	12.9
Hirtshals	12.4	12.9	12.9
Frederikshavn	13.1	13.3	13.3
Hanstholm	12.4	12.6	12.1
Grenå	13.9	14.9	14.2
Århus	15.5	16.0	14.2
Fredericia	12.6	12.0	12.3
Thyborøn	15.0	15.2	11.9
Ferring	15.3	15.5	13.6
Torsminde	17.0	17.6	14.8
Hvide Sande	16.2	15.9	15.8
Esbjerg	18.3	18.6	19.8
Havneby	18.1	18.1	20.6
Aabenraa	16.7	15.6	12.2
Fynshav	13.5	12.7	12.0
Ballen	12.6	13.5	13.0
Slipshavn	13.1	13.4	12.7
Spodsbjerg	11.9	12.5	12.2
Odden	11.4	12.4	12.4
Korsør	14.6	14.5	12.0
Hornbæk	15.2	15.8	11.7
København	14.8	14.8	13.8
Nordre Røse	17.7	17.4	11.0
Drogden	9.6	9.2	9.4
Rødvig	9.8	9.2	9.4
Hesnæs	11.1	10.6	9.6
Rødby	11.9	12.1	10.5
Gedser	11.1	10.3	9.8
Tejn	7.6	7.3	9.1
Rønne	10.3	11.4	9.6

Table 11: Average RMS error, 1998. D15/UKM = main/backup setup, M21 = upgrade setup.

The RMS error ranges from 7 cm (Tejn) to 20 cm (Skagen) 14 cm for both setups. The RMS error is 15-25% higher than the MAE in general. Low RMS ($<12\frac{1}{2}$ cm) is found at stations in the Western Baltic, at Odden, Spodsbjerg, and the Skagerrak stations Hirtshals and Hanstholm. High RMS ($>17\frac{1}{2}$) is found at Skagen, Esbjerg, Havneby and Nordre Røse. Maximum D15-UKM differences are about 1 cm.

The maximum error (MAX) (Table 12)

Station	D15	UKM	M21
Skagen	54.8	55.5	48.4
Hirtshals	42.6	44.0	48.4
Frederikshavn	42.1	42.2	47.8
Hanstholm	44.8	47.9	47.2
Grenå	44.2	46.5	50.4
Århus	45.5	46.4	51.5
Fredericia	43.6	40.2	41.4
Thyborøn	51.8	53.3	40.9
Ferring	53.4	52.9	46.9
Torsminde	56.1	60.1	48.1
Hvide Sande	57.2	56.8	51.9
Esbjerg	69.9	69.2	59.6
Havneby	60.9	65.0	63.4
Aabenraa	53.1	50.1	39.5
Fynshav	45.2	44.5	38.1
Ballen	38.5	42.7	47.2
Slipshavn	42.6	44.5	44.7
Spodsbjerg	45.4	46.6	48.3
Odden	36.1	40.5	43.2
Korsør	36.6	38.5	37.2
Hornbæk	39.0	39.6	43.2
København	48.5	50.6	46.8
Nordre Røse	44.9	45.8	40.0
Drogden	33.0	29.4	30.0
Rødvig	35.1	31.3	28.7
Hesnæs	37.7	34.0	29.2
Rødby	34.6	35.3	33.2
Gedser	37.8	35.5	31.0
Tejn	21.0	18.2	23.0
Rønne	28.2	26.8	28.2

Table 12: Average maximum error, 1998. D15/UKM = main/backup setup, M21 = upgrade setup.

The MAX error has been redefined compared with 1997 (cf. Ch. 4) and now represents the the average of the worst prediction for each month. The MAX error ranges roughly from 20 cm (Tejn) to 70 cm (Esbjerg) for both setups. High MAX error (>50 cm) is found all along the North Sea coast, and at Aabenraa and Skagen. Low MAX (<30 cm) is found at Bornholm, where the sea level variation is rather weak. The D15-UKM setups differ up to 5 cm. No setup is evidently better in this respect.

The explained variance (EV) (Table 13)

Station	D15	UKM	M21
Skagen	64.7	61.3	62.7
Hirtshals	69.2	65.4	65.1
Frederikshavn	59.5	55.2	57.2
Hanstholm	75.1	71.5	74.6
Grenå	68.8	60.3	63.4
Århus	63.2	55.4	56.1
Fredericia	61.7	57.9	58.7
Thyborøn	78.0	75.6	87.1
Ferring	81.4	79.5	86.2
Torsminde	79.5	77.2	85.6
Hvide Sande	87.2	86.1	86.8
Esbjerg	93.2	92.8	90.8
Havneby	94.8	94.5	92.0
Aabenraa	70.5	70.9	76.7
Fynshav	67.1	67.6	73.7
Ballen	67.0	58.4	60.4
Slipshavn	71.8	66.0	68.1
Spodsbjerg	64.7	61.2	60.4
Odden	61.9	53.6	52.4
Korsør	44.2	38.4	38.0
Hornbæk	70.9	62.5	64.7
København	48.0	42.6	52.5
Nordre Røse	41.5	38.0	41.6
Drogden	69.4	73.4	76.2
Rødvig	77.8	80.6	82.8
Hesnæs	80.5	82.2	85.4
Rødby	79.7	81.6	82.8
Gedser	78.0	80.9	83.1
Tejn	84.3	88.8	82.4
Rønne	83.9	86.6	81.8

Table 13: Explained variance, 1998. D15/UKM = main/backup setup, M21 = upgrade setup.

The D15 EV ranges from 41% (Nordre Røse) to 95 % (Havneby). Very low EV, less than 50%, is found the Belt Sea stations Korsør, København and Nordre Røse. During summer months, even negative EV is found at Korsør, indicating that the sea level prediction varies out of phase with the observed sea level. Very high EV, more than 75%, is found in the Wadden Sea, at the West Coast, and in the Western Baltic. The D15 setup is better in the Wadden Sea, at the West Coast, in the Skagerrak, Kattegat, the Great Belt and the Sound, while the UKM setup is better at stations in the Western Baltic. The EV is not affected by unbiasing.

The mean peak error (MPE) (Table 14)

Station	D15	UKM	M21
Skagen	2.8	15.9	-7.4
Hirtshals	-18.1	-3.5	-10.1
Frederikshavn	-18.9	-2.8	-11.0
Hanstholm	-13.3	-0.5	-1.4
Grenå	-29.4	-26.1	-20.7
Århus	-31.9	-31.6	-20.2
Fredericia	-21.0	-28.0	-18.5
Thyborøn	-8.7	8.1	-15.1
Ferring	-13.7	7.3	-17.7
Torsminde	-8.9	3.9	-15.3
Hvide Sande	-16.5	6.9	-7.5
Esbjerg	5.1	22.8	4.0
Ribe Kammersluse	14.0	34.2	15.0
Havneby	-5.1	11.8	7.9
Aabenraa	-21.5	-31.9	-13.0
Ballum Sluse	7.0	25.6	11.6
Vidå Slusen	6.7	23.4	4.2
Fynshav	-13.7	-19.7	-11.1
Ballen	-21.2	-25.0	-11.9
Slipshavn	-23.2	-31.1	-20.7
Spodsbjerg	-35.3	-37.2	-39.9
Odden	-13.4	-16.9	-6.1
Korsør	-20.9	-22.5	-10.8
Hornbæk	-24.3	-24.6	-6.7
København	-22.7	-26.8	-26.6
Nordre Røse	-28.1	-29.9	-16.2
Drogden	-14.4	-16.7	-14.1
Rødvig	-13.8	-17.8	-13.3
Hesnæs	-17.4	-18.7	-12.7
Rødby	-5.2	-6.3	-12.2
Gedser	-15.1	-14.0	-11.2
Tejn	-5.3	-4.7	-10.1
Rønne	-2.2	-2.4	-15.3

Table 14: Mean peak error, 1998. D15/UKM = main/backup setup, M21 = upgrade setup.

Both operational setups underpredict sea level peaks on average at most stations. For the D15 setup, 28 out of 33 stations have negative MPE (peak bias), ranging down to -35 cm (Spodsbjerg). Positive peak bias is found only at Wadden Sea stations and at Skagen, and large negative bias is found at stations in the Kattegat, Belt Sea and the Skagerrak. For the UKM setup, positive bias is also found at West Coast stations. For most of the North Sea coast, UKM peak predictions are 15-25 cm higher than D15 peak predictions on average. While removing a mean error at many stations, unbiasing is found to have a negative impact on the mean peak error at the Skagerak stations. At Skagen, Aabenraa and Nordre Røse, where the ME exceeds 10 cm (Table 9), unbiasing would diminish the MPE at Aabenraa and Nordre Røse but introduce a large negative bias at Skagen.

The mean absolute peak error (MAPE) (Table 15)

Station	D15	UKM	M21
Skagen	13.4	21.3	17.4
Hirtshals	18.3	17.1	13.1
Frederikshavn	19.3	14.8	13.8
Hanstholm	17.1	18.9	11.8
Grenå	29.4	28.5	22.1
Århus	31.9	31.6	21.6
Fredericia	21.0	28.0	18.5
Thyborøn	16.9	14.3	15.3
Ferring	15.9	14.1	19.3
Torsminde	18.7	21.7	17.5
Hvide Sande	20.3	14.7	10.7
Esbjerg	16.3	26.2	10.2
Ribe Kammersluse	23.0	34.4	16.6
Havneby	15.9	25.0	11.7
Aabenraa	21.5	31.9	13.6
Ballum Sluse	18.4	27.6	12.4
Vidå Slusen	18.9	25.0	9.6
Fynshav	15.1	19.7	12.7
Ballen	21.2	25.0	14.1
Slipshavn	23.2	31.1	20.7
Spodsbjerg	35.3	37.2	39.9
Odden	13.4	18.3	10.5
Korsør	20.9	22.5	11.6
Hornbæk	24.3	25.2	12.3
København	22.7	26.8	26.6
Nordre Røse	28.1	29.9	18.2
Drogden	16.4	17.1	15.3
Rødvig	15.0	18.0	14.1
Hesnæs	18.0	18.7	15.3
Rødby	10.6	11.3	14.4
Gedser	16.1	16.6	15.4
Tejn	8.7	7.9	12.1
Rønne	6.8	5.4	16.5

Table 15: Average absolute peak error, 1998. D15/UKM = main/backup, M21 = upgrade setup.

The D15 MAPE ranges from 7 cm (Rønne) to 35 cm (Spodsbjerg). Large MAPE (>20 cm) is found at most stations in the Kattegat and the Belt Sea, and at Hvide Sande and Ribe. Small MAPE (<10 cm) is found only at Bornholm. The UKM MAPE follows very much the same pattern, but the MAPE is larger at 24 stations, smaller only at 9 stations, indicating that the D15 setup is better at predicting sea level peaks. In the Wadden Sea, D15 MAPE is in general 10 cm smaller than UKM MAPE. Unbiasing enlargens the MAPE significantly at Skagerrak stations, while improving the MAPE at Fynshav and Rødby.

The maximum peak error (MXPE) (Table 16)

Station	D15	UKM	M21
Skagen	41	45	42
Hirtshals	38	42	39
Frederikshavn	46	42	40
Hanstholm	34	43	25
Grenå	50	58	51
Århus	56	65	48
Fredericia	39	54	39
Thyborøn	47	33	44
Ferring	44	35	31
Torsminde	36	48	42
Hvide Sande	43	46	38
Esbjerg	40	82	18
Ribe Kammersluse	52	94	58
Havneby	30	78	18
Aabenraa	42	49	32
Ballum Sluse	45	85	37
Vidå Slusen	44	81	36
Fynshav	42	32	31
Ballen	37	36	22
Slipshavn	48	54	41
Spodsbjerg	74	77	78
Odden	26	28	20
Korsør	38	52	27
Hornbæk	34	51	23
København	37	50	52
Nordre Røse	49	51	33
Drogden	27	35	31
Rødvig	30	29	30
Hesnæs	31	30	27
Rødby	30	32	30
Gedser	40	41	29
Tejn	19	19	25
Rønne	14	12	28

Table 16: Maximum peak error, 1998. D15/UKM = main/backup, M21 = upgrade setup.

The MXPE is the error of the worst peak prediction of the 10 highest events at each station. The D15 MXPE ranges from 14 cm (Rønne) to 74 cm (Spodsbjerg) with 56 cm (Århus) as second worst. The typical range is 40-45 cm. For the UKM setup, a single event in the Wadden Sea is very badly predicted, with peak errors up to almost 100 cm. The D15 setup is better than the UKM setup at 24 stations, the UKM setup being much better only at Fynshav.

Peak errors at the warning stations (Table 17, Figs. 6-14)

For the 8 warning stations, the ten highest peaks and peak errors are shown in order of decreasing peak magnitude. The peaks are separated in time by at least one tidal period. A positive error indicates that the forecast was too high, a negative that it was too low. The D15 results for each station is summarised below, using a 10/20 cm success criteriae for the peak hit rate. Unfortunately, UKM peaks from October were not available. The UKM results are shown on the Figures (lower panel) for comparison.

Station	<10cm	10-20cm	>20cm	Under	Over
Torsminde	3	3	4	7	3
Esbjerg	3	5	2	3	6
Vidå	3	2	5	3	7
Frederikshavn	3	3	4	9	1
Aabenraa	2	3	5	10	0
Slipshavn	1	3	6	10	0
Korsør	1	4	5	10	0
Rødby	5	4	1	6	4
Gedser	4	3	3	9	1
Total	25	30	35	67	22

Table 17: D15 peaks at warning stations, 1998.

Of the 90 peaks considered here, less than one-third is predicted with 10cm error or less, one-third is predicted with an error of 10-20 cm, and little more than one-third are not well predicted. The largest peak error is in the order of 45-50 cm. At Frederikshavn, Aabenraa, Slipshavn, Korsør and Gedser, sea level peaks are nearly always underpredicted. At Torsminde, Esbjerg, Vidå and Rødby, peaks may be over- or underpredicted.

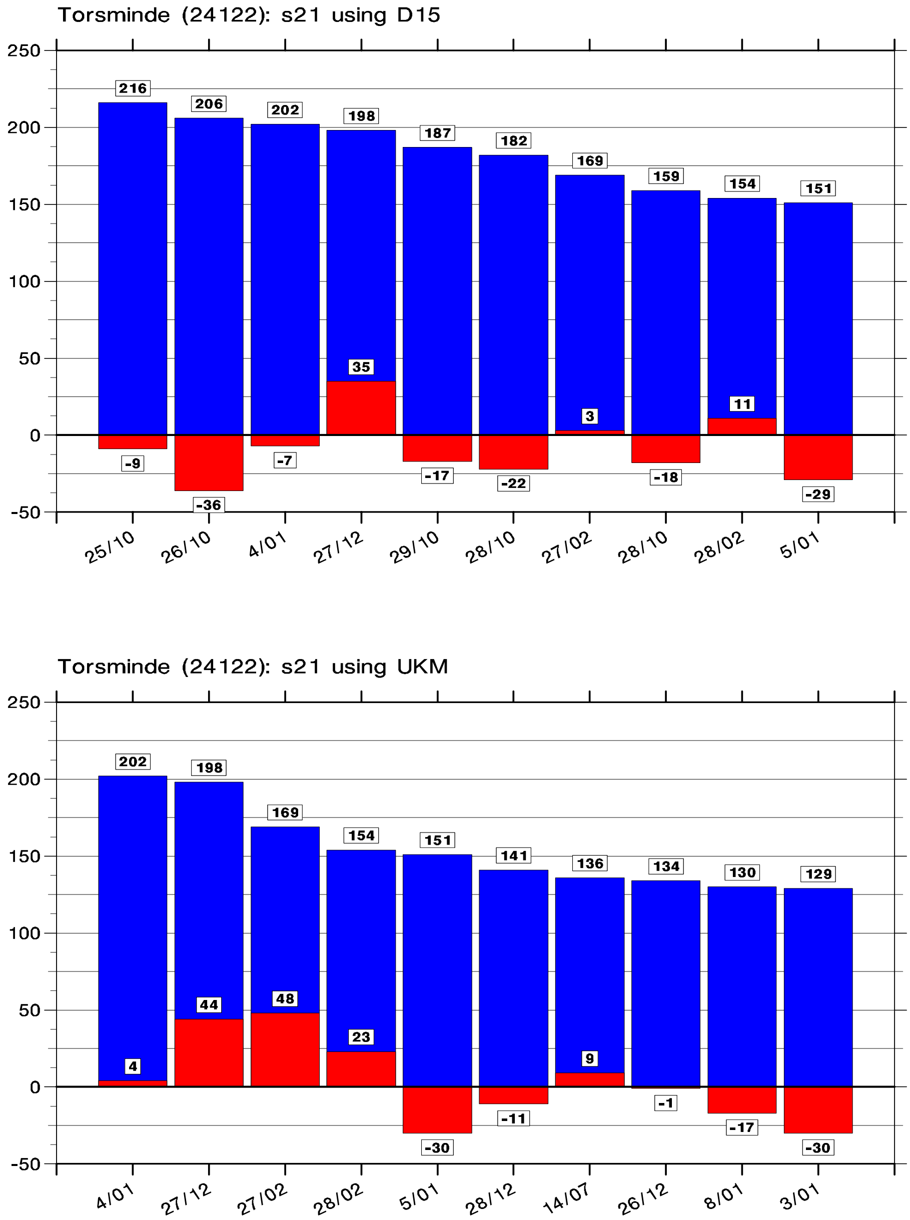


Figure 6: Torsminde 10 highest peaks and errors, 1998. Upper panel: Main (D15) setup. Lower panel: Backup (UKM) setup.

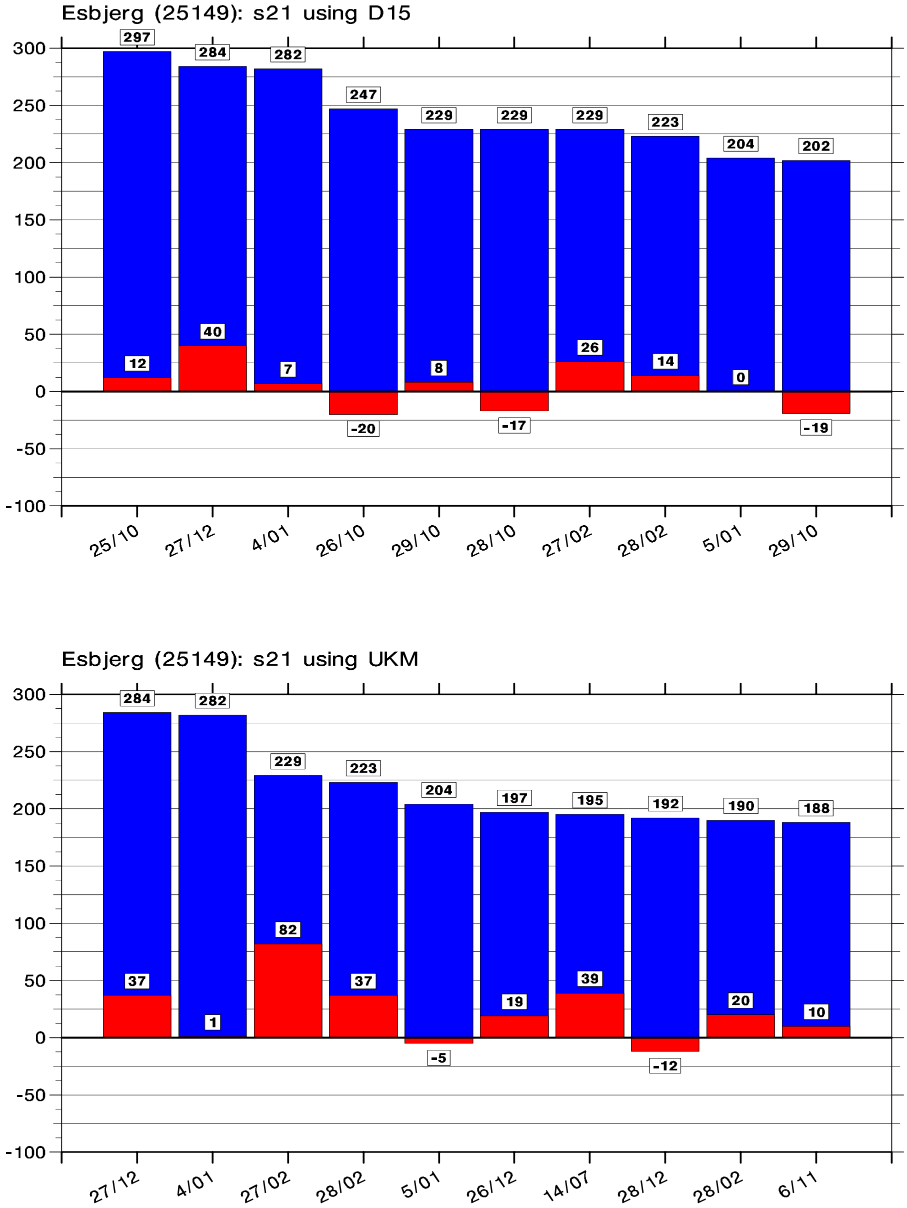
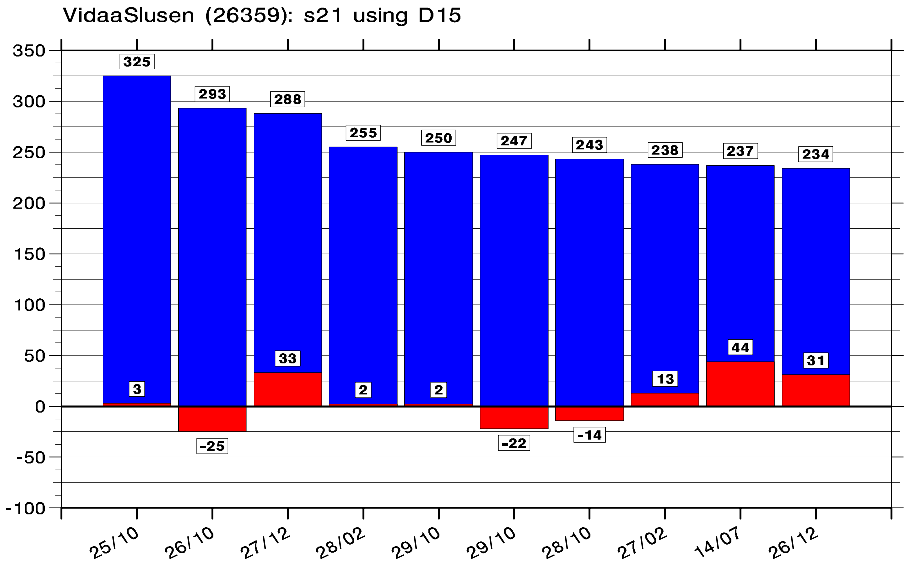


Figure 7: Esbjerg 10 highest peaks and errors, 1998. Upper panel: Main (D15) setup. Lower panel: Backup (UKM) setup.



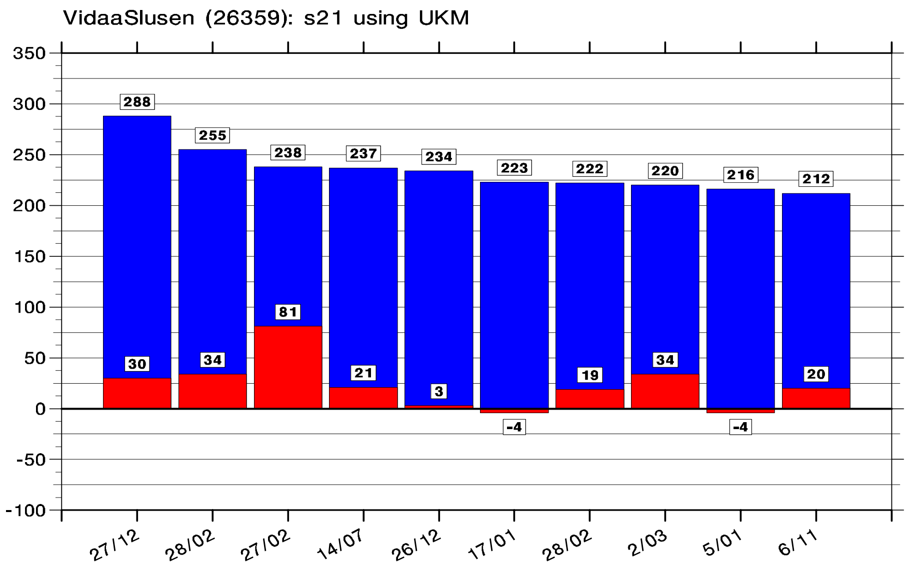


Figure 8: Vidå Sluse 10 highest peaks and errors, 1998. Upper panel: Main (D15) setup. Lower panel: Backup (UKM) setup.

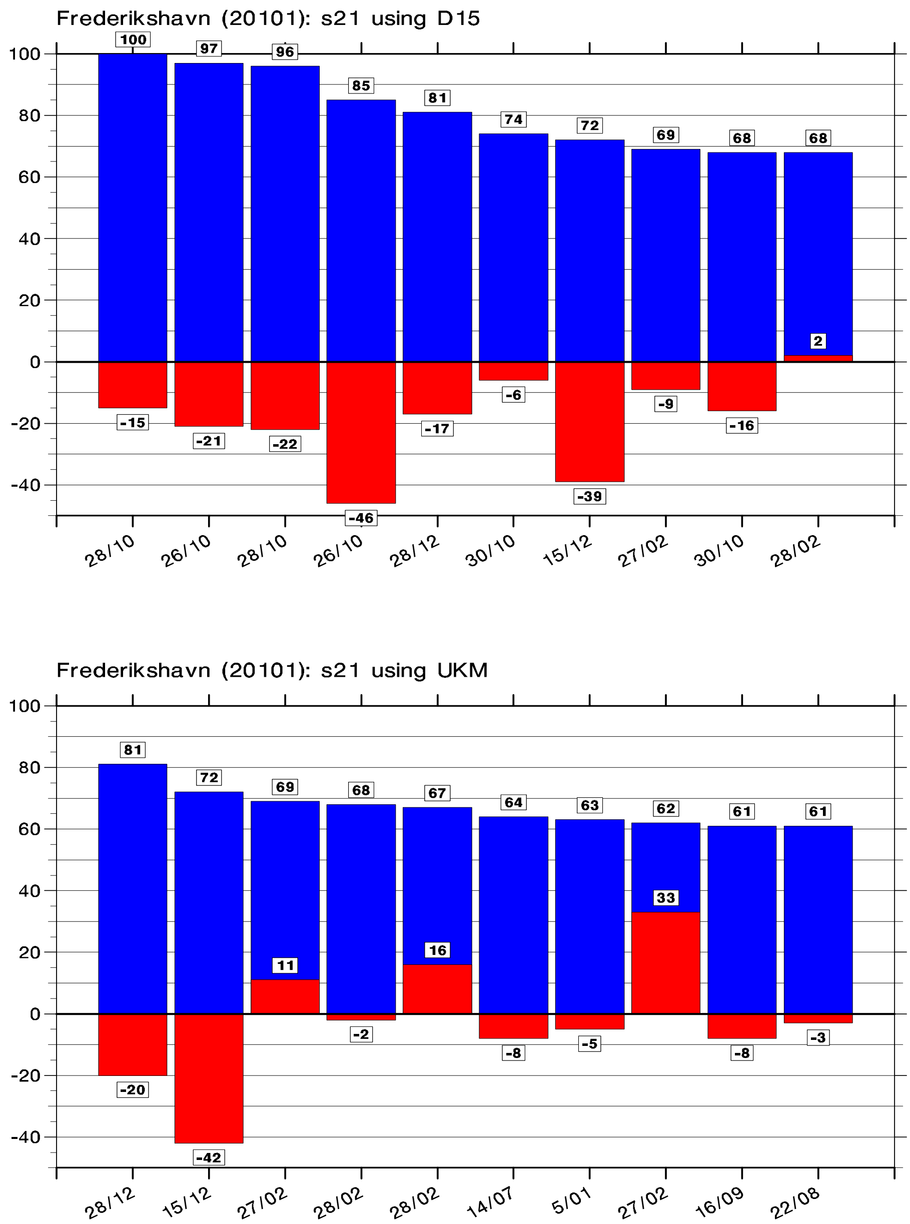


Figure 9: Frederikshavn 10 highest peaks and errors, 1998. Upper panel: Main (D15) setup. Lower panel: Backup (UKM) setup.

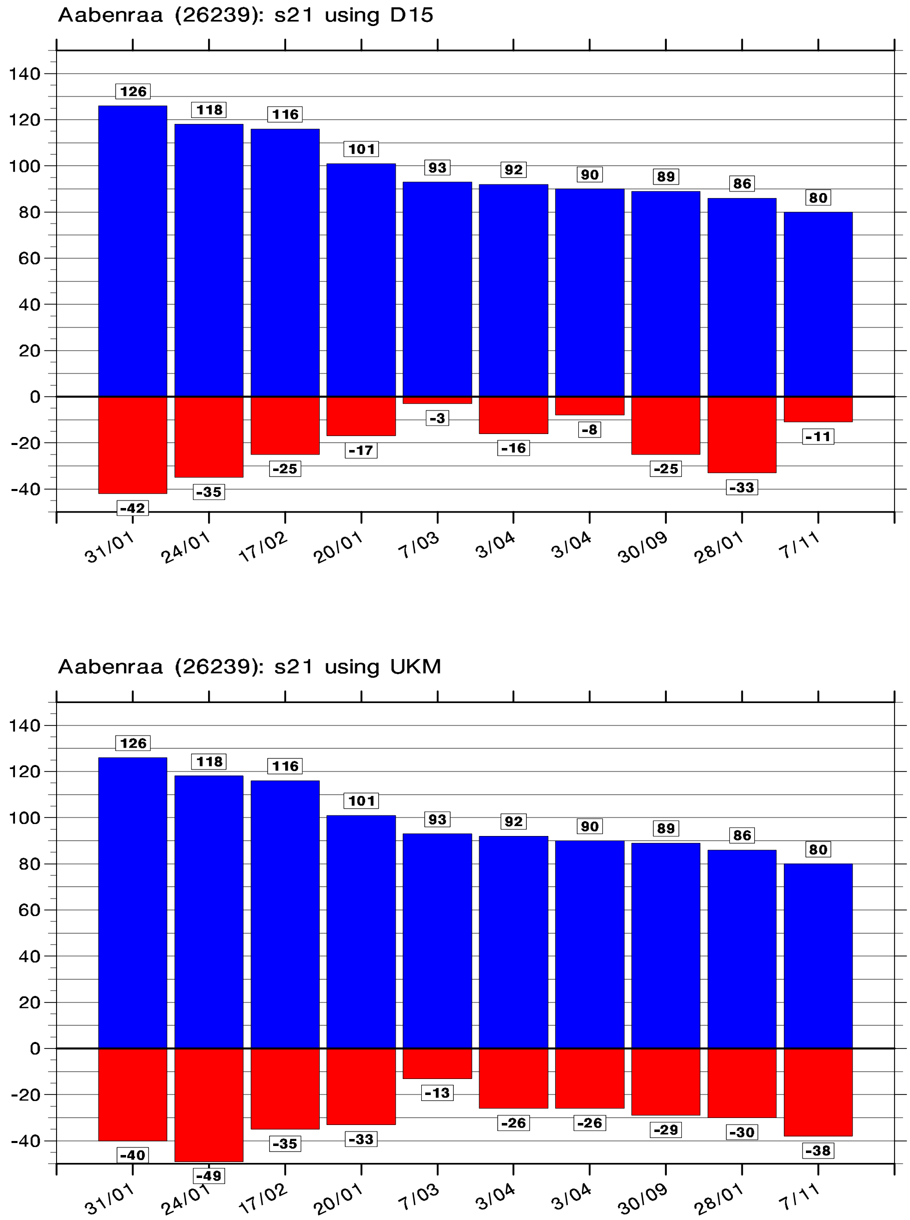


Figure 10: Aabenraa 10 highest peaks and errors, 1998. Upper panel: Main (D15) setup. Lower panel: Backup (UKM) setup.

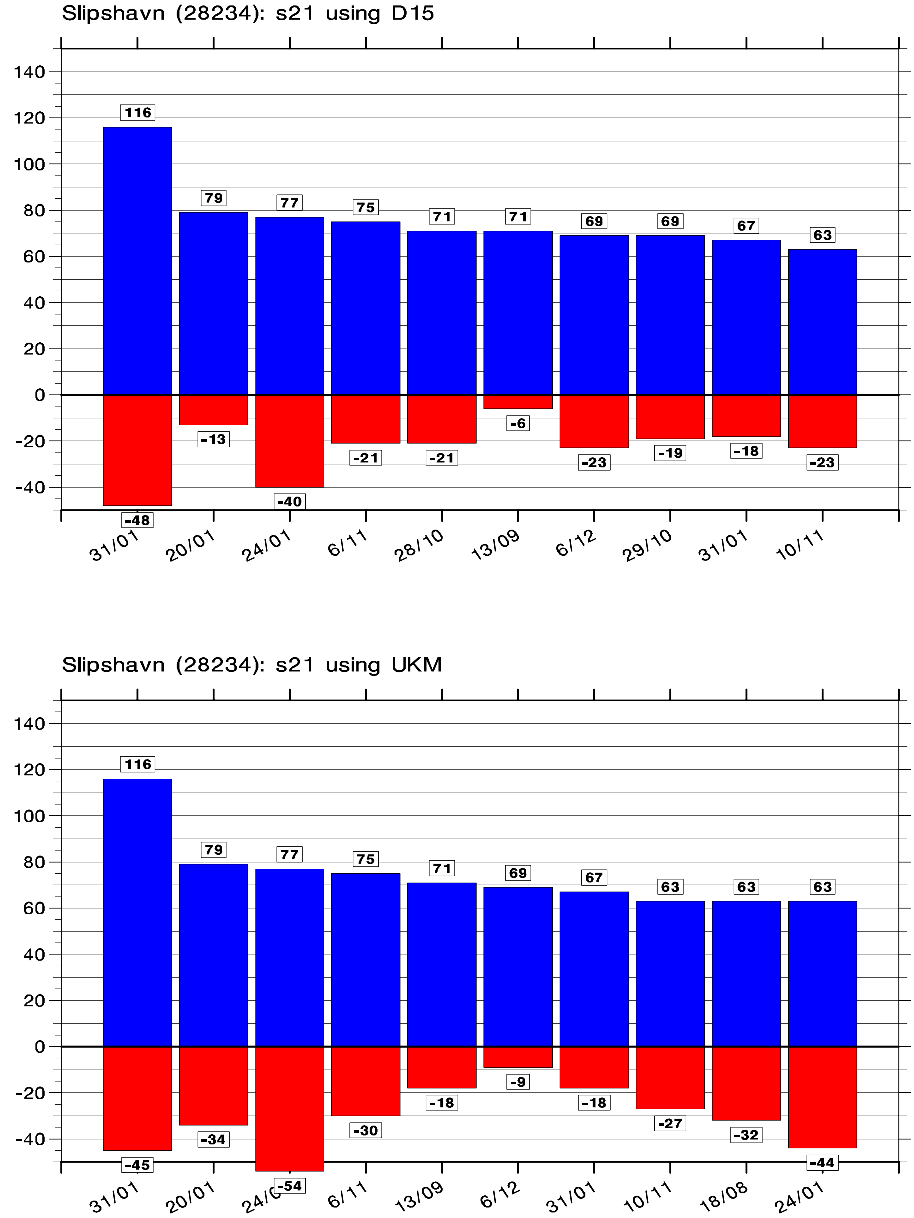


Figure 11: Slipshavn 10 highest peaks and errors, 1998. Upper panel: Main (D15) setup. Lower panel: Backup (UKM) setup.

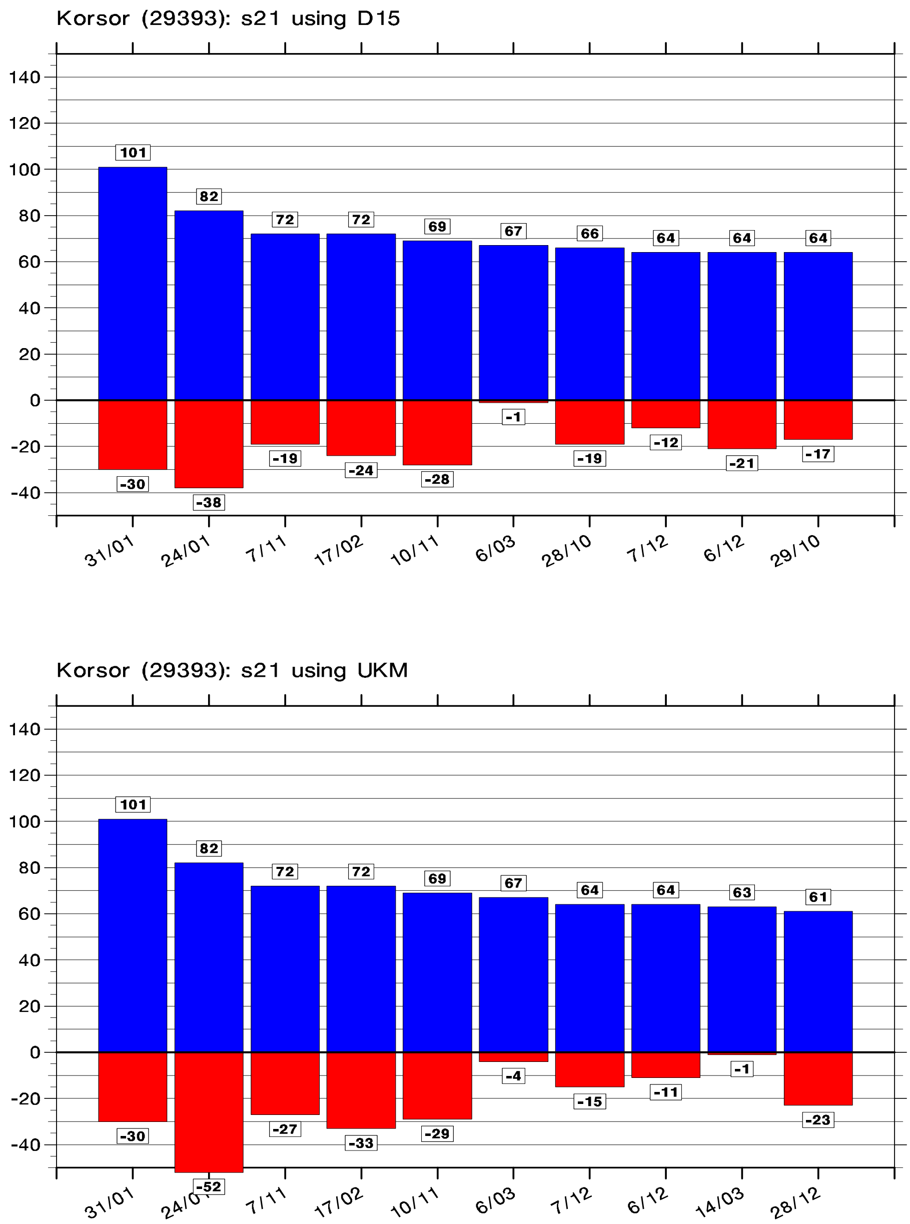


Figure 12: Korsør 10 highest peaks and errors, 1998. Upper panel: Main (D15) setup. Lower panel: Backup (UKM) setup.

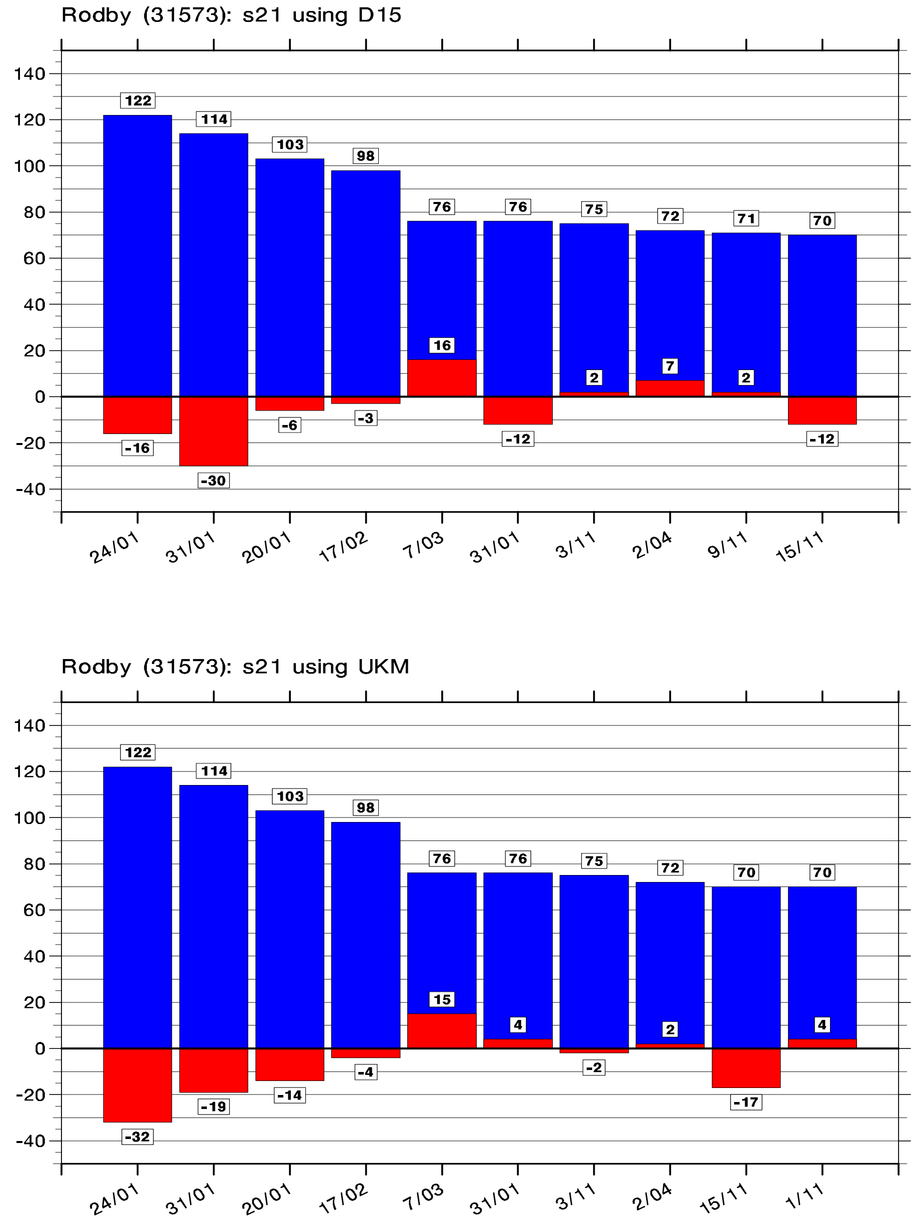


Figure 13: Rødby 10 highest peaks and errors, 1998. Upper panel: Main (D15) setup. Lower panel: Backup (UKM) setup.

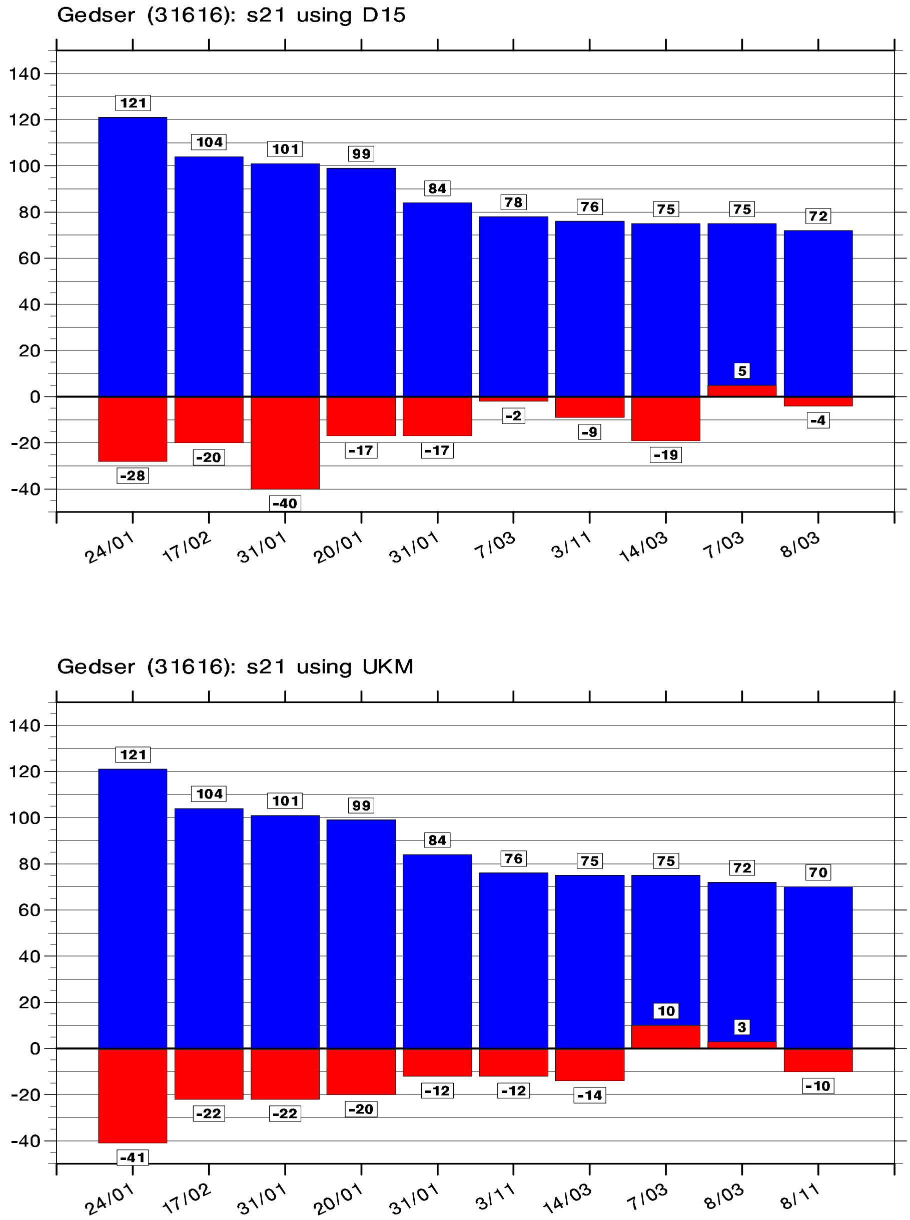


Figure 14: Gedser 10 highest peaks and errors, 1998. Upper panel: Main (D15) setup. Lower panel: Backup (UKM) setup.

Key numbers (Table 18)

Station-averaged error measures corresponding to Tables 9-16 are presented in Table 18 below. The Table includes operational (D15+UKM) error measures, upgrade (M21) error measures, and M21 results before unbiasing. M21 results are discussed in Ch. 7. D15 hindcast results were practically identical to the forecast results and are not presented here.

Residual error measure	D15	UKM	M21	M21B
Avg mean error	-3 cm	-1 cm	(0 cm)	-4 cm
Avg absolute mean error	5 cm	5 cm	(0 cm)	8 cm
Avg mean absolute error	11 cm	11 cm	10 cm	13 cm
Avg rms error	14 cm	14 cm	13 cm	16 cm
Avg maximum error	44 cm	45 cm	43 cm	47 cm
Avg explained variance	71 %	69 %	71 %	71 %
Peak error measure	D15	UKM	M21	M21B
Avg mean peak error	-15 cm	-9 cm	-12 cm	-16 cm
Avg absolute mean peak error	17 cm	19 cm	14 cm	19 cm
Avg maximum peak error	41 cm	50 cm	38 cm	42 cm
<10 cm peak hit rate	48 %	47 %	57 %	41 %
<20 cm peak hit rate	78 %	77 %	85 %	72 %

Table 18: 1998 station-averaged annual error measures. D15/UKM = main/backup setup, M21 = backup setup, D15H = operational hindcasts, M21B = upgrade hindcasts before unbiasing.

The D15 average ME is -3 cm, indicating that sea sevel is underpredicted on average. The UKM setup is more base level neutral, with an average ME of -1 cm. The average value of absolute ME is 5 cm for both setups. The average MAE is 11 cm for both setups, and the average RMS error 14 cm. The average MAX error is 44 cm for the D15 setup, 45 cm for the UKM setup. Finally, the D15 average EV is 71%, which is a little better than the UKM average EV of 69%. D15 hindcasts (not shown) have a slightly higher average EV of 72%.

To sum up, residual error measures are of almost the same quality for both setups, perhaps with a slight edge to the D15 (Main) setup.

The D15 average peak error MPE (peak bias) is -15 cm. High sea level peaks are in genereal underpredicted. For the UKM setup the peak bias is only -9 cm but it should be kept in mind that UKM October forecasts are not included. The D15 average MAPE is 17 cm, the UKM average MAPE 19 cm, indicating that D15 peak prediction are on average superior to the UKM predictions. Also the average maximum peak error is lower for the D15 setup (41 cm) than for the UKM setup (50 cm). Finally, the D15 peak hit rates of 48% (10 cm criterion) and 78% (20 cm criterion) are a little better than the UKM hit rates.

To sum up, D15 peak error measures are better than UKM results, except that the peak D15 bias is larger.

6 Comparison with previous years

The 1994-98 time development of nine key numbers is shown in Fig. 15. It should be noted that the data sets are not homogeneous. New stations are added every year, and 11 new stations are included in the calculation of 1998 key numbers. From Fig 15 the following conclusions are made:

ME: The D15 average ME has dropped below -2 cm. The bias at individual stations (not shown) displays rather large year-to-year variations.

MAE and RMS: The average MAE stays constant in the 11-12 cm range for both setups. The average RMS error stays constant in the 14-15 cm for both setups, 15-25% higher than the average MAE.

MAX: The calculation of the average MAX error has been changed. Instead of the single largest error, we now average the largest error from each month. In 1998, the average MAX error is roughly 45cm for both setups.

EV: The average EV fluctuates a bit but stays around 70% $\pm$ a few percent, for both setups. The D15 average EV is a little higher than the UKM average EV.

MPE: The average peak bias is negative at all times, in the range -10 to -20 cm. The UKM peak bias is not representative because October predictions are missing. There is no simple time development.

MAPE: The decrease of the D15 average MAPE, first evident in 1997, persists. Average MAPE is now down to 17 cm. In 1998, UKM average MAPE has also fallen below 20 cm. 1994 results are not representative because 4 Wadden Sea stations were not included yet.

MXPE: The D15 average maximum peak error stays at the low 1997 level, with a small tendency to increase. As for the MAPE, 1994 results are not representative.

PHIT: The peak hit rate (10 cm criterion) is just below 50% for both setups, with the D15 score in general a little higher than the UKM score.

The average MAE and the average RMS show a weak decrease with time. The other key numbers show no steady time development.

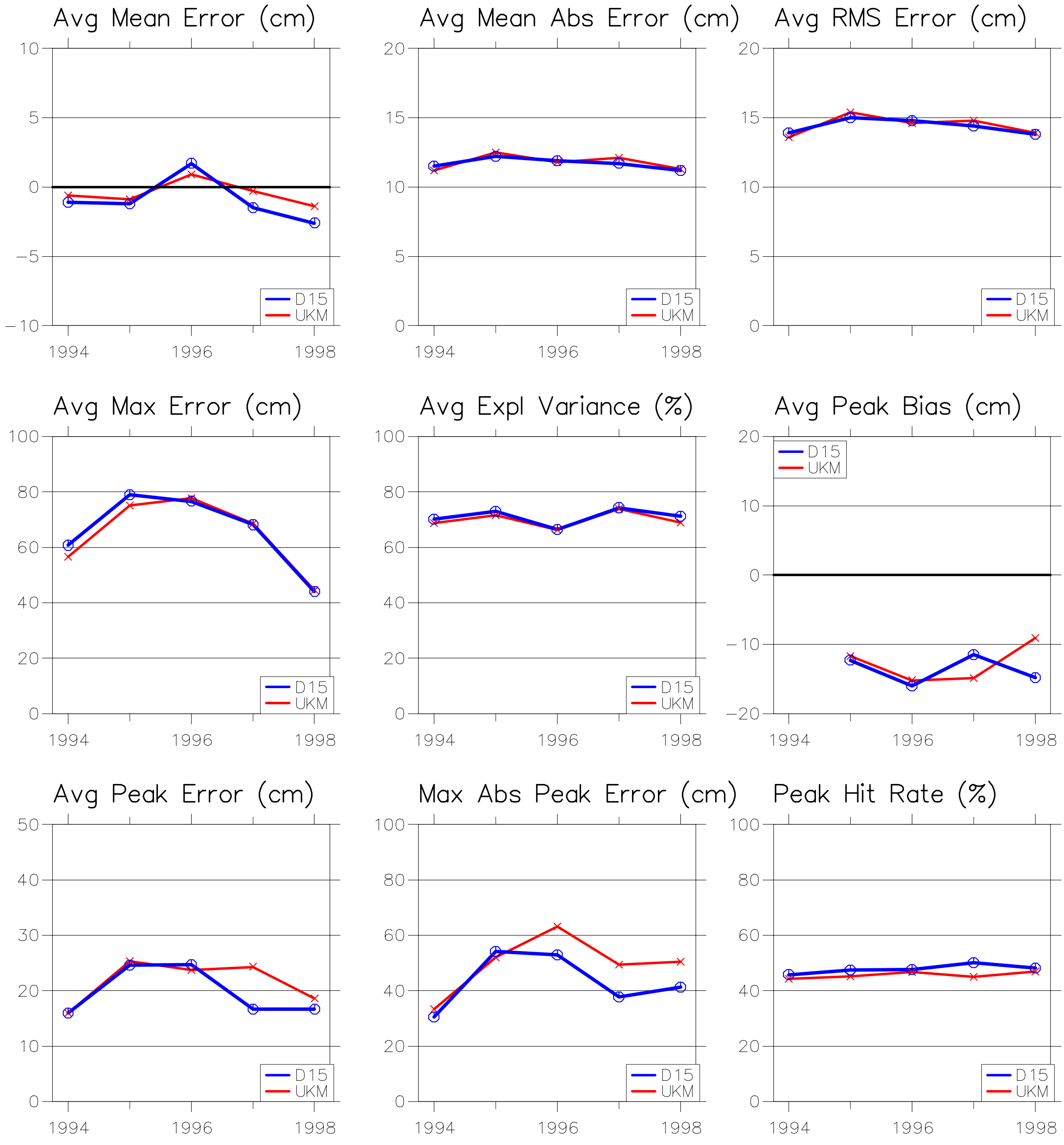


Figure 15: Station-averaged error estimates, 1994-98.

7 Comparison with upgrade setup

The upgrade setup *DKSS98* has been run in hindcast mode for 1998. Below, this is compared with main (D15) forecasts. The effect of unbiasing upgrade (M21) forecasts is examined on average. The results are found in the Tables presented in Ch. 5. Statistics on the peak errors at the warning stations are presented in Table 19 below.

The unbiased M21 hindcasts are better than the D15 forecasts. The average MAE, RMS, MAX and EV all drop significantly, as do the peak error measures and the peak hit rate. If M21 are not unbiased, the results are worse than D15 forecasts.

When unbiased, the M21 ME drops to zero at all stations. Before unbiasing, ME in excess of 10 cm was found at almost half the stations (Table 9).

The M21 MAE (Table 10) is better than the D15 MAE at most stations, and significantly worse only at Esbjerg, Havneby and Tejn. The same result applies to the RMS error (Table 11).

The M21 MAX error (Table 12) may be better or worse than the D15 MAX error at individual stations. The average MAX errors differ only by 1 cm.

The M21 EV (Table 13) shows no improvement over the D15. At some stations the predictions get worse, at others they get better. Average EV is the same (71%) for both setups.

The M21 peak bias (Table 14) is lower than the D15 bias at most stations, but at some station (Torsminde, Thhyborøn, Rødby, Tejn, Rønne) the bias is larger. At Spodsbjerg a very large peak bias of -40 cm is found. On average, the bias has dropped to -12 cm.

The M21 MAPE (Table 15) may also be better or worse than D15 MAPE at individual stations. On average, the MAPE has dropped to 14 cm.

The M21 maximum peak error (Table 16) is also slightly smaller than the D15 MXPE.

The M21 peak hit rate (Table 18) shows a 7-9% increase over the D15 peak hit rate, depending on success criterion.

Table 19 below show the distribution of M21 peak errors at the warning stations. More peaks are well predicted ($<$10 cm error) and less badly predicted ($>$20 cm error) than with the D15 setup (comp. with Table 17).

It should be noted that the results described above apply to hindcasts with no bias at all. In the future, the unbiasing will be imperfect due to the natural variations in mean sea level at the individual stations.

Station	<10cm	10-20cm	>20cm	Under	Over
Torsminde	3	3	4	8	2
Esbjerg	5	5	0	3	7
Vidå	8	1	1	3	5
Frederikshavn	5	2	3	7	2
Aabenraa	5	3	2	9	1
Slipshavn	2	3	5	10	0
Korsør	4	4	2	8	1
Rødby	3	4	3	9	1
Gedser	2	5	3	7	3
Total	37	30	23	64	24

Table 19: M21 peaks at warning stations, 1998. Compare with Table 17.

8 Conclusion

Operational storm surge forecasts from 1998 are verified. Sea level predictions, produced by the hydrodynamical model System 21 forced by the D15-HIRLAM (D15) and the UK-LAM (UKM) atmospheric models are evaluated on a monthly basis, using observations from 33 Danish tide gauge stations. Three Wadden Sea sluice stations are only used for peak statistics. Only the time window analysis+06 hours to analysis+18 hours is considered, and the dependency on forecast range is not taken into account. The predictions are unbiased using 1994-96 mean errors, but new (in 1998) stations are not unbiased. The AR filtering process that the predictions undergo in an actual warning situation is not considered here.

The two setups are intercompared, and further comparison is made with unbiased results, with results from previous years and with results from an upgrade setup.

The D15 setup is used for storm surge warning purposes. The UKM setup serves mainly as a backup. The two setups are of very similar quality, with the D15 setup scoring marginally better in most respects. Mean absolute error (MAE) ranges from 6-18 cm with a station average of 11 cm. Explained variance ranges from 41-95% with a station average of 71%. For predicting sea level peaks, the D15 setup scores better. D15 mean peak error is 17 cm, the UKM mean peak error 19 cm. Maximum peak errors are 41 cm and 50 cm, respectively. There is a general tendency to underpredict sea level peaks, mostly so in the Belt Sea, the Kattegat, and the Skagerrak. The D15 peak hit rate (10 cm criterion) is 48%, again a few percent better than the UKM peak hit rate. It should be noted that UKM predictions from October are missing, while D15 hindcasts are used instead of forecasts for that month.

Comparing with 1994-97, we find almost no changes but the predictions are definitely not getting any worse.

The M21 upgrade setup has been run in hindcast mode only. M21 predictions are unbiased using mean errors from the same year, making the comparison a little unfair to the D15 setup.

From the key numbers, the M21 predictions are definitely better than D15 predictions when unbiasing is applied - and definitely worse when the 'raw' predictions are used. In forthcoming years, still using 1998 mean errors as biases, the unbiasing will only partly remove the forecast mean error.

Finally on the concept of unbiasing we note that although unbiasing may remove a large part of the mean error, maen absolute error etc., it may at the same time lead to larger peak errors. This aspect of unbiasing needs to be examined further in the future.

The data-assimilation setup termed *DMIsurge* [3] is no longer in use.

References

[1] Højvandsstatistikker 1997. Technical report, Kystinpektoratet, 1997.

[2] DHI. System 21 HD Scientific Documentation. Technical report, Danish Hydraulic Institute, 1988.

[3] Jacob Woge Nielsen. Verification of Sea Level Forecast: 1997. Technical Report 98-5, Danish Meteorological Institute, 1998.

[4] Jacob Woge Nielsen and Karsten Bolding. Verification of Sea Level Forecasts: October 1993 – March 1994. Technical Report 94-11, Danish Meteorological Institute, 1994.

[5] Jacob Woge Nielsen and Karsten Bolding. Verification of Sea Level Forecast: 1994. Technical Report 95-4, Danish Meteorological Institute, 1995.

[6] Jacob Woge Nielsen and Karsten Bolding. Verification of Sea Level Forecasts: October 1994 – March 1995. Technical Report 95-9, Danish Meteorological Institute, 1995.

[7] Jacob Woge Nielsen and Karsten Bolding. Verification of Sea Level Forecasts: 1995. Technical Report 96-2, Danish Meteorological Institute, 1996.

[8] Jacob Woge Nielsen and Karsten Bolding. Stormflodsrapport: Højvandet den 29. og 30. oktober 1995. Technical Report 97-2, Danish Meteorological Institute, 1997.

[9] Jacob Woge Nielsen and Karsten Bolding. Stormflodsrapport: Højvandet den 6. og 7. november 1995. Technical Report 97-6, Danish Meteorological Institute, 1997.

[10] Jacob Woge Nielsen and Karsten Bolding. Verification of Sea Level Forecasts: 1996. Technical Report 97-1, Danish Meteorological Institute, 1997.

[11] Jacob Woge Nielsen, Karsten Bolding, and Lonny Hansen. Extreme sea level highs: A statistical tide gauge data study. Technical Report 96-3, Danish Meteorological Institute, 1996.

[12] Hans Jacob Vested, Henrik Rene Jensen, Helmer Petersen, Anne Mette Jørgensen, and Bennert Machenhauer. An operational hydrographic warning system for the North Sea and the Danish Belts. *Continental Shelf Research*, 12(1):65–81, July 1992.

List of Tables

List of Figures