DANISH METEOROLOGICAL INSTITUTE

––––– TECHNICAL REPORT –––––

98-5

Verification of Sea Level Forecasts 1997

Jacob Woge Nielsen

DMI

Copenhagen 1999

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

Verification of Sea Level Forecasts 1997

Jacob Woge Nielsen
e-mail: jw@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 1997 are evaluated. Forecasts valid for the 12 hour time window analysis+06 hours to analysis+18 hours are compared with the observed sea level from 22 Danish coastal tide gauges, with special emphasis on the most severe surges. The forecast quality is compared with previous years, and the importance of unbiasing the forecasts is examined.

Two test set-ups are also evaluated. The first one is a DMI model of the North Sea-Kattegat using data assimilation, and the second one is an upgrade of the operational system, using revised bathymetry and resistance matrices, and a slightly higher spatial resolution. Both have been running in semi-operational mode during 1997.

The report includes a brief description of the operational storm surge system (Ch. 2) and of the test set-ups (Ch. 3), an outline of the data including an estimate of the return period of the most severe events (Ch. 4), and definitions of the statistical measures used to describe the forecast quality (Ch. 5). Ch. 6 summarises the 1997 operational forecast quality, and in Ch. 7 the effect of unbiasing the forecasts is examined. Chs. 8-9 compares with previous years and with test set-ups, respectively. Ch. 10 concludes the work.

The report includes 25 Figures, inserted after the relevant text sections. A reference list and a Figure inventory is given 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 Operational set-up

The warning system DKSS90 has been used for operational storm surge prediction since October 1990, and has been verified since October 1993[1]. It has three components:

- a depth-integrated hydro-dynamical model *System 21*
- a limited area atmospheric forcing model *LAM*
- open boundary tidal elevations

System 21 is developed by the Danish Hydraulic Institute (DHI, cf. [1]), and kindly made available to DMI for operational use. The model domain (Fig. 1) consists of

- A coarse grid model (10 nautical mile resolution) covering the North Sea, the Transition Area, and the Baltic.
- An embedded sub-model covering the Danish Domestic Waters (resolution 3.3 n.m)
- Two further embedded sub-models covering the Wadden Sea and the Belt Sea, respectively (resolution 1.1 n.m).

The fine grid spacing allows flow to pass through narrows, belts and straits of the Wadden Sea and the Belt Sea. Each sub-model is dynamically coupled to the coarser grid model.

Two set-ups of DKSS90 are run in parallel, differing only in the atmospheric forcing. Both set-ups are fully operational, and both are verified and compared in the present report.

Main *(DKV)* set-up:

> System 21 forced by the Danish *DMI-HIRLAM*[1]
> Spatial resolution: 0.21^o
> Since September 1997: 0.15^o
> Temporal resolution: 1 hour

Back-up *(UKM)* set-up:

> System 21 forced by the British *UK-LAM*
> Spatial resolution: 1.25^o
> Temporal resolution: 6 hours[2]

The back-up set-up is used for storm surge warning when DMI-HIRLAM (in short: DKV) forecasts are not available. The atmospheric fields (msl pressure, 10m wind) constituting the surface boundary condition for the hydro-dynamical model are interpolated to the coarsest hydro-dynamical model grid. The UKM wind speed is then augmented artificially by 10% over the North Sea and by 1-2 m/s over the Belt Sea (the DKV wind speed is not modified). Finally, the surface stress is calculated using a speed-dependent drag coefficient.

A DKSS90 run produces a hindcast (based on analysed meteorological fields) plus a short-range 36 hour forecast of surface elevation and vertical mean current on the model grids. The forecast range can be extended to 5 days by supplementing DKV/UKM with a coarser global forcing model *ECMWF*.

[1] Cf. [3],[4],[5],[6],[9],[7],[8]
[1] A technical description of HIRLAM may be found in [11]
[2] The UKM is higher resolved, but DMI receives model fields in this resolution.

The DKSS90 schedule is shown in Fig. 2. The system is run twice daily, the 00 UTC and the 12 UTC run respectively, when a new LAM forecast is available. If LAM is unavailable at run-time, DKSS90 instead uses yesterday's forecast plus the global forcing model.

The *00 UTC* run:

a 24 hour hindcast
a 36 hour forecast

The *12 UTC* run:

a 36 hour hindcast
a 36 hour forecast

Each model run is initiated by a model state output from a previous run, the 12 UTC run using the same initial model state as the previous 00 UTC run. No use is made in the model of the observed model state, in terms of tide gauge or current data.

The 24 hour hindcast produced at 00 UTC is stored on tape. Since October 1996, DMI provides the Sound Link Construction Company with medium-range (5 days) current and sea level forecasts.

For further information on DKSS90, see [13].

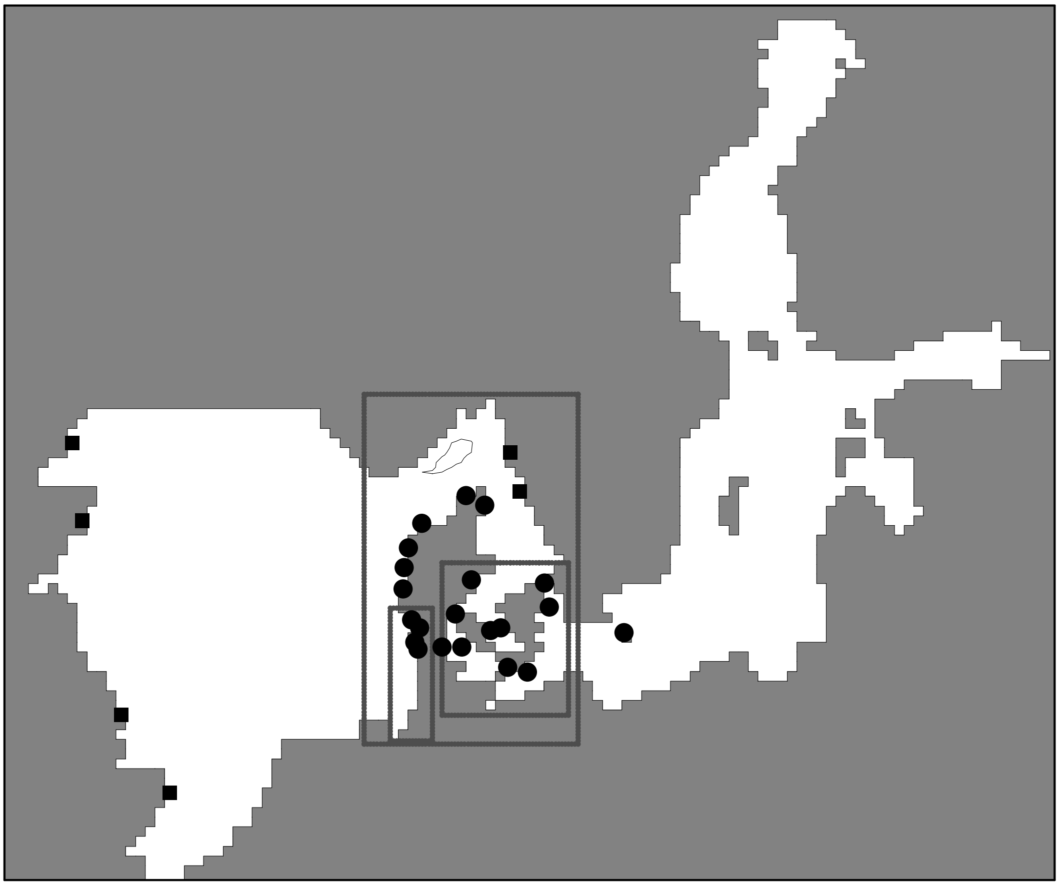


Figure 1: The DKSS90 model domain with nested grid arrangement, totalling four computational grids. Open tidal boundaries are present at the Shetland-Bergen section and in the British Channel. In the upgrade set-up (cf. Ch. 3), the Domestic model domain extends eastwards to a position just west of Gotland, while the Belt Sea model domain is extended somewhat less far east, to a position just east of Bornholm. The coarse grid assimilation set-up model domain is truncated by a further open boundary running from Djursland to Sweden. Locations of the tide gauge stations used for verification are indicated by solid circles. The assimilation set-up makes use of sea level from 8 western and northern Jutland stations, plus 6 British/Swedish stations (indicated by squares).

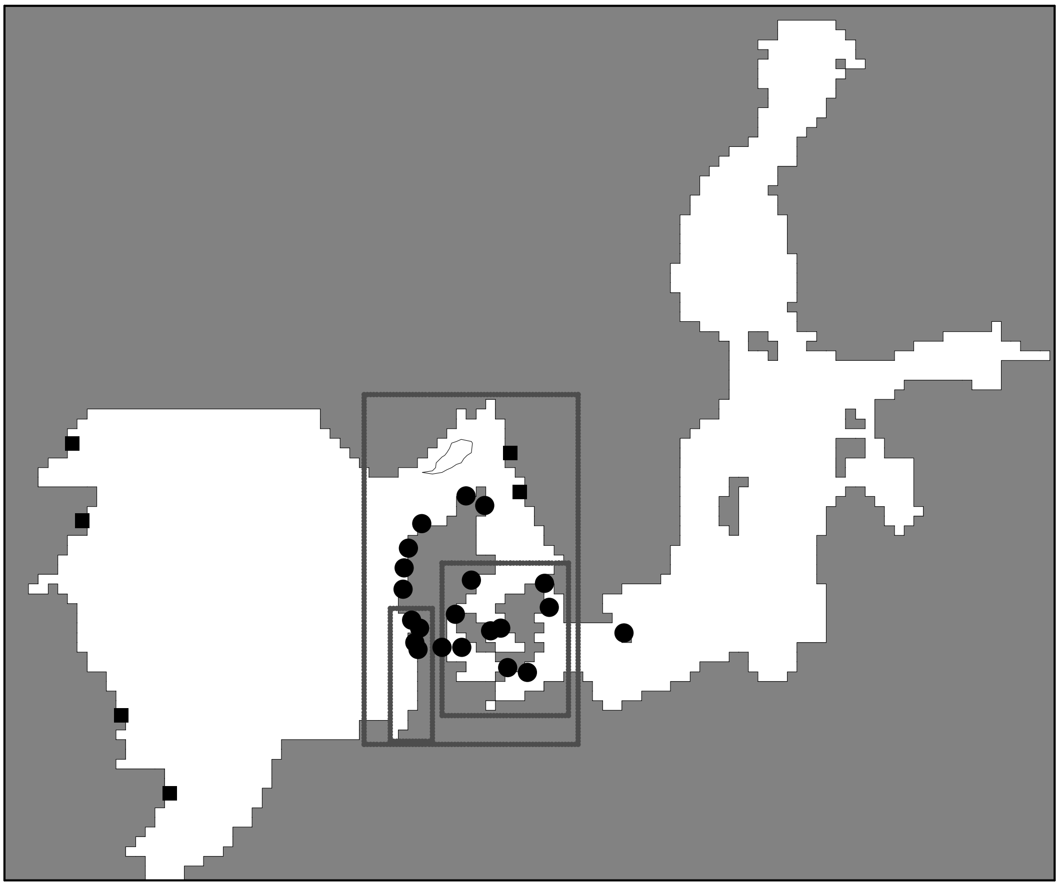


Figure 2: Schedule of the operational runs. From every 00 UTC run, a 24 hour hindcast with a time resolution of 30 min. is stored on tape. The runs may be extended to a range of 5 days using ECMWF atmospheric forcing fields.

3 Test set-ups

During 1997, two test set-ups were run in semi-operational mode (operational schedule but no data dissemination). Both were forced by the Danish DKV-HIRLAM.

DKSS96:

Since August 1996, an upgrade of DKSS90 termed DKSS96 is running in semi-operational mode. DKSS96 uses a new hydro-dynamical model (Mike21, likewise developed by the DHI) a slightly refined grid configuration (9 n.m. resolution in the coarsest grid, revised bathymetries), and revised tidal boundaries. Also, all friction terms have been reevaluated. Sea level forecasts have been archived since October 1996. Results from 1997 are discussed below.

DMIsurge:

Since October 1995, a coarse-grid (9 n.m.) quasi-linear hydro-dynamical model of the North Sea - Kattegat termed DMIsurge is running in semi-operational mode. The bed friction depends on velocity squared, other terms are linear. A Riemann open boundary condition is implemented in the southern Kattegat. The model assimilates tide gauge data available at run-time from 14 coastal stations (cf. Fig. 1) using a stationary Kalman filter technique (see, e.g., [2]), and steps forward in time using the explicit Sielecki scheme [12]. Preliminary results have been archived since March 1996. Results from 1997 are discussed below.

A complete DMI storm surge set-up inventory is given below:

Set-up	Main	Back-up	Upgrade	Data-ass.
Schedule name	DKSS90	DKSS90	DKSS96	DMIsurge
Short name	DKV	UKM	M21	DMI
Mode of run	Operational	Operational	Semi-oper.	Semi-oper.
Hyd. model	System 21	System 21	Mike 21	DMIsurge
Atm. model	DKV-HIRLAM	UK-LAM	DKV-HIRLAM	DKV-HIRLAM
Model domain	N.Sea - Baltic	N.Sea - Baltic	N.Sea - Baltic	N.Sea - Kattegat
Assimilation	No	No	No	Yes
Grid nesting	Yes	Yes	Yes	No

Table 1: Storm surge set-ups at DMI

4 Data

Observations:

Sea level from 22 automatic tide gauge stations are used for verification (cf. Table 2, Fig 1).

Station		Number	Position	Region (grid)	Owner
Hirtshals	d)	20047	$57^o36'N$ $09^o58'E$	Domestic	DMI
Frederikshavn	d)	20101	$57^o26'N$ $10^o34'E$	Domestic	DMI
Hanstholm	d)	21009	$57^o07'N$ $08^o36'E$	Domestic	DMI
Aarhus		22331	$56^o09'N$ $10^o13'E$	Belt Sea	DMI
Fredericia		23293	$55^o34'N$ $09^o45'E$	Belt Sea	DMI
Thyborøn	d)	24006	$56^o42'N$ $08^o13'E$	Domestic	KI
Torsminde	d)	24122	$56^o22'N$ $08^o07'E$	Domestic	KI
Hvide Sande	d)	24342	$56^o00'N$ $08^o08'E$	Domestic	KI
Esbjerg	d)	25149	$55^o28'N$ $08^o26'E$	Wadden Sea	DMI
Ribe Sluse	*)	25343	$55^o20'N$ $08^o41'E$	Wadden Sea	KI
Havneby	d)	26136	$55^o05'N$ $08^o34'E$	Wadden Sea	KI
Åbenrå		26239	$55^o03'N$ $09^o26'E$	Belt Sea	AMT
Ballum Sluse	*)	26346	$55^o08'N$ $08^o41'E$	Wadden Sea	KI
Vidaa Sluse	*)	26359	$54^o58'N$ $08^o40'E$	Wadden Sea	DMI
Fynshav		26457	$55^o00'N$ $09^o59'E$	Belt Sea	DMI
Slipshavn		28233	$55^o17'N$ $10^o50'E$	Belt Sea	DMI
Korsør		29393	$55^o20'N$ $11^o08'E$	Belt Sea	DMI
Hornbæk		30017	$56^o06'N$ $12^o28'E$	Belt Sea	DMI
København		30337	$55^o41'N$ $12^o30'E$	Belt Sea	DMI
Rødby		31573	$54^o39'N$ $11^o21'E$	Belt Sea	DMI
Gedser		31616	$54^o34'N$ $11^o56'E$	Belt Sea	DMI
Tejn		32048	$55^o15'N$ $14^o50'E$	Baltic	DMI

Table 2: Verification stations. Station number, name, position and DKSS90 grid. Owner DMI=Danish Meteorological Institute; KI=Coastal Authorities, AMT = a local authority. *) = only used for peak statistics. d) = also used for assimilation.

Of the stations used,

- 15 are operated by DMI
- 6 are operated by the Coastal Authorities (KI)
- 1 is operated by a local authority (AMT)

Furthermore, of these stations

- 3 (marked *)) are sluice stations, unable to record low water correctly
- 8 (marked d)) are used for data-assimilation in the DMIsurge set-up

Since ultimo 1997, tide gauge data from 13 stations and current data from 5 moorings operated by the Royal Danish Academy of Navigation and Hydrography (RDANH) have kindly been made available to the DMI. These are, however, not verified in the present report.

The 1997 data coverage (Fig. 3 and Table 3) is rather high, with a bulk coverage of 96.4%. DMI stations alone have a data coverage of 97.7%. This is a substantial improvement from 1995 and 1996. Four stations have a data coverage below 95%:

Year	1997	1996	1995	1994
Data Coverage %	96.4	91.1	93.4	96.9
# of stations	22	21	21	17
# below 95%	4	7	3	4

Table 3: Observation data coverage (%) 1994-97.

- *Hvide Sande:* drifted towards higher sea level from February 11th through April 7th; that period has been deleted
- *Ribe Kammersluse:* malfunctioning from July 27th through August 20th, plus several periods of shorter duration when the tide gauge got stuck at a constant sea level
- *Åbenrå:* twice drifted towards higher sea level during August 1st through September 16th; that period has been deleted
- *København:* malfunctioning and repaired from July 25th through September 8th.

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. For verification, a subset of the data with a half-hour time resolution is used. In 1998, all DMI stations will be re-calibrated to record sea level relative to the common datum DNN (Danish Normal Null). The subject of reference level is discussed further below.

Surges:

The highest and the lowest sea level recorded at each station during 1997 are shown in Fig. 4. The Wadden Sea sluice stations (indicated by a *) in Table 2), Vidå Sluse, Ribe Kammersluse, and Ballum Sluse, do not record sea level below datum accurately and are only used for verification of forecasts of very high waters (peaks).

The dates of the 7 most severe surges during 1997, plus the region affected, are tabulated below:

Date	Region
February 21st	Skagerrak
February 25th	West Coast
April 11th	Belt Sea/Western Baltic
October 2nd	The Sound
October 10th	Wadden Sea
November 3rd	Belt Sea/Western Baltic
December 25th	Wadden Sea

Table 4: 1997 major storm surges

Return periods for the highest sea level at each station are estimated using a Gumbel distribution [10]. This is done for the DMI stations only, since the record lengths at other stations are yet too short for the analyses to be reliable. The results are shown in Fig. 5. A return period of two years or less corresponds to a year where the highest sea level was below median severity. High return periods (> 3 years) is noted at stations located in the Little Belt, the Great Belt, the Sound, and the Western Baltic, while the North Sea coast experienced a rather calm year.

During 1997, DMI had warning responsibility for 7 stations: The warning thresholds, and the number of times this has been exceeded during 1997, are shown in Table 5 (Gedser serves as a back-up station for Rødby).

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

Table 5: Warning thresholds (cm), and number of times this has been exceeded during 1994-97. The tide gauge at Åbenrå was established during 1996.

Forecasts:

As part of the scheduled model runs, time series of forecasted sea level at model grid points representing each station in Table 2 are archived[1]. Due to the coarseness of the grid, model grid points may be situated a few km away from the nearest tide gauge station. The archived series consist of a hindcast and a 36 hour forecast, with time resolution 30 min.

The verification period spans one year totalling 730 forecasts. The 12 hour time window analysis+06 hours to analysis+18 hours (considered the most important for storm surge warning purposes) of all forecasts are concatenated to make a pseudo time series. The remaining part of the forecasts is ignored in this context.

The 1994-97 forecast coverage for both operational set-ups is shown in Table 6. Also shown are the 1997 test set-up coverages. Forecasts may be missing for one of these reasons:

- the atmospheric model was not run in time
- the hydro-dynamical model was not run
- time series archiving failed

In case of delayed or missing atmospheric forcing fields, DKSS90 tries a substitute forcing model, or failing that, runs just the hindcast part of the prediction. In either case, that run is discarded for verification purposes. If atmospheric forcing fields were available, the hydro-dynamical model will generally also be run.

During 1997, operational forecasts from July were not archived due to tape station problems. In total, 68 forecasts are missing. For the back-up (UKM) set-up, forecasts from July 1st through September 8th have been deleted as the results were erroneous. The reason for this is unresolved. In total, 147 forecasts are missing. From 1994-97, the DKV coverage is generally 5-10% higher than the UKM coverage.

[1]Sea level at grid points representing to 4 British and 6 Swedish stations are also archived, but not verified. Since ultimo 1997, time series representing RDANH stations are also archived.

Set-up	1997	1996	1995	1994
Main (DKV)	90.6	80.3	97.7	65.4
Back-up (UKM)	80.2	76.0	91.8	56.6
Upgrade (M21)	96.5	-	-	-
Data-ass. (DMI)	93.2	-	-	-

Table 6: 1997 forecast data coverage. In 1997, there is a total of 730 36-hour forecasts. In total, 17520 data points per station are used for verification.

Both test set-ups were more stable, with less than 10% data loss. The Upgrade set-up failed 26 times, mainly during March and January. The Data-Ass. set-up failed 50 times, mainly during January.

Unbiasing and filtering:

Previous investigations have shown that a mean error persists at individual stations. This may be corrected for by shifting the forecast by this amount (unbiasing). Since October 7th 1997, operational sea level forecasts are unbiased using 1994-96 mean errors (Table 7, obtained from [4],[6],[9]) in an attempt to make the forecasts base level neutral relative to the station datum.

To verify a homogeneous forecast series, no unbiasing is applied (forecasts from October 7th onwards are returned to unbiased state). Instead, the full effect of unbiasing all 1997 forecasts in retrospect, using 1994-96 mean errors, is discussed in a special section.

Station	Number	Bias	Station	Number	Bias
Hirtshals	20047	-15	Åbenrå	26239	no=0
Frederikshavn	20101	-9	Ballum Sluse	26346	no=0
Hanstholm	21009	-8	Vidaa Sluse	26359	5
Aarhus	22331	-1	Fynshav	26457	16
Fredericia	23293	6	Slipshavn	28233	0
Thyborøn	24006	-2	Korsør	29393	3
Torsminde	24122	3	Hornbæk	30017	-5
Hvide Sande	24342	8	København	30337	0
Esbjerg	25149	-2	Rødby	31573	5
Ribe Sluse	25343	7	Gedser	31616	4
Havneby	26136	2	Tejn	32048	-3

Table 7: 1994-96 average mean error used as bias (cm)

Since ultimo 1997, operational sea level forecasts are filtered using an autoregressive model for the forecast error, resulting in an updated forecast whenever new observations are available. The filtered forecasts are not archived. This report only considers unfiltered forecasts.

The test set-ups are neither unbiased or filtered.

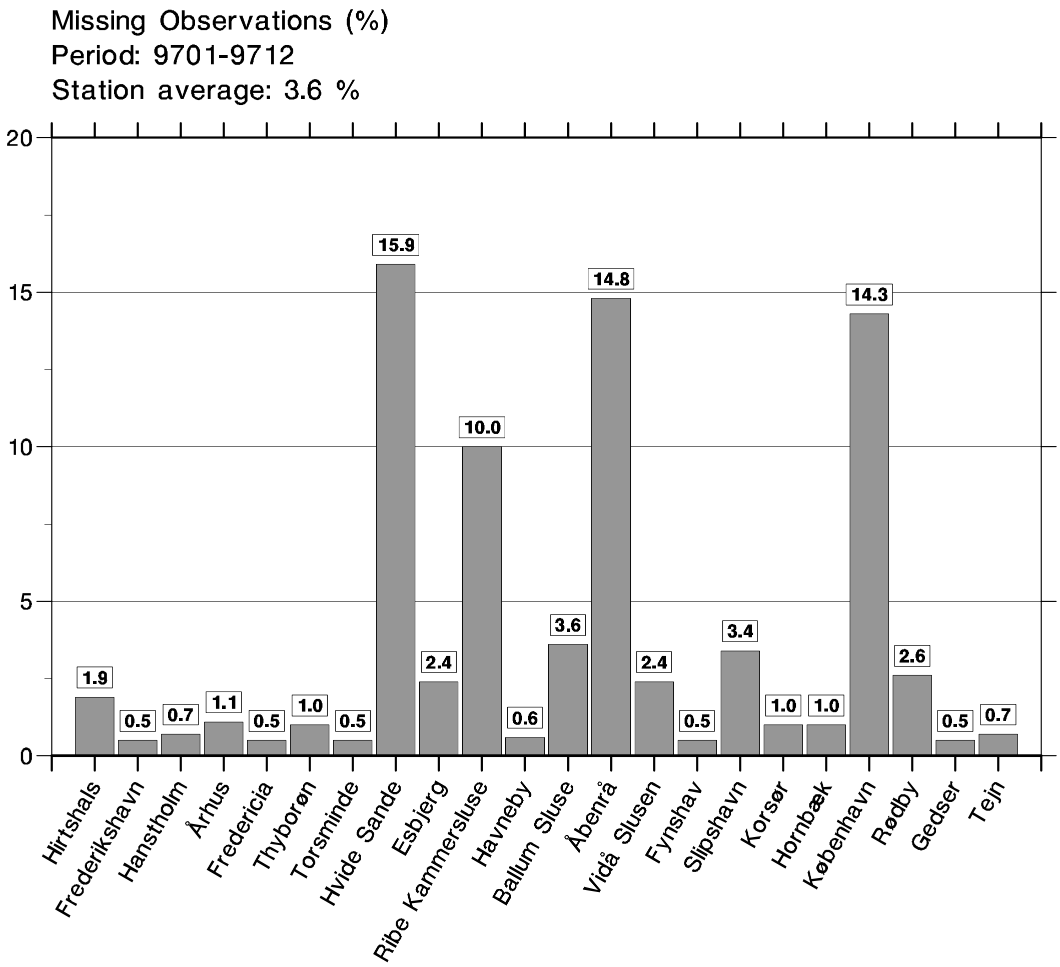


Figure 3: Missing observations, 22 stations, 1997.

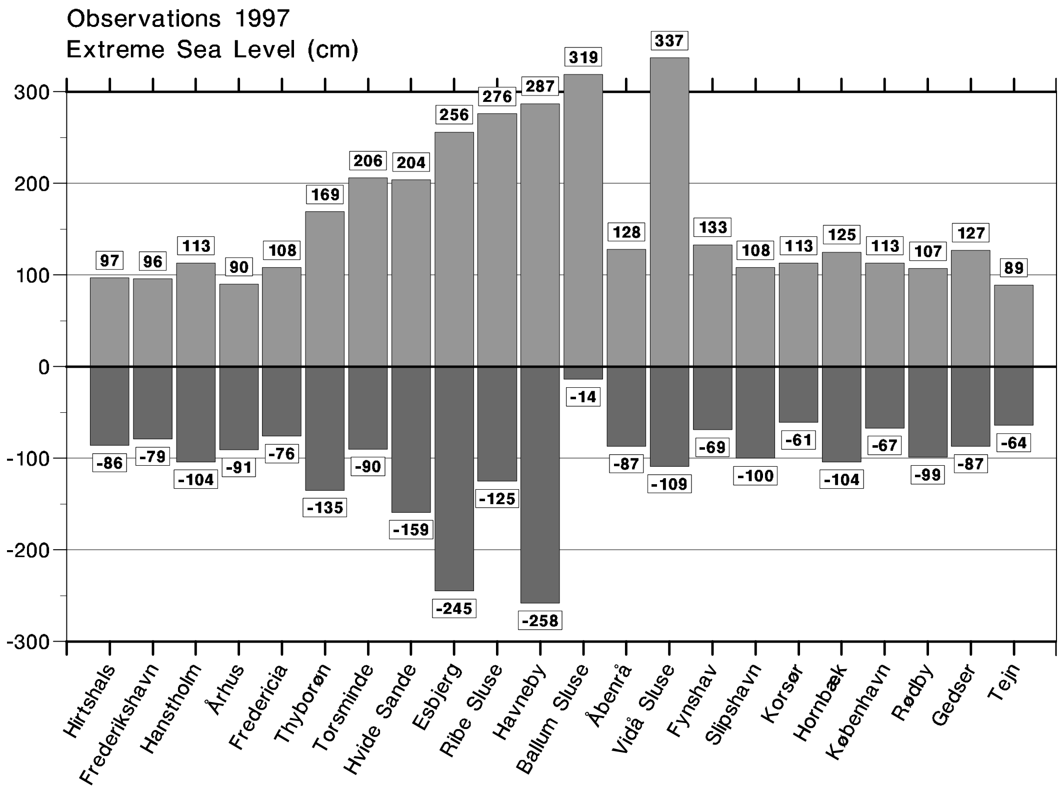


Figure 4: Extreme sea level, 22 stations, 1997.

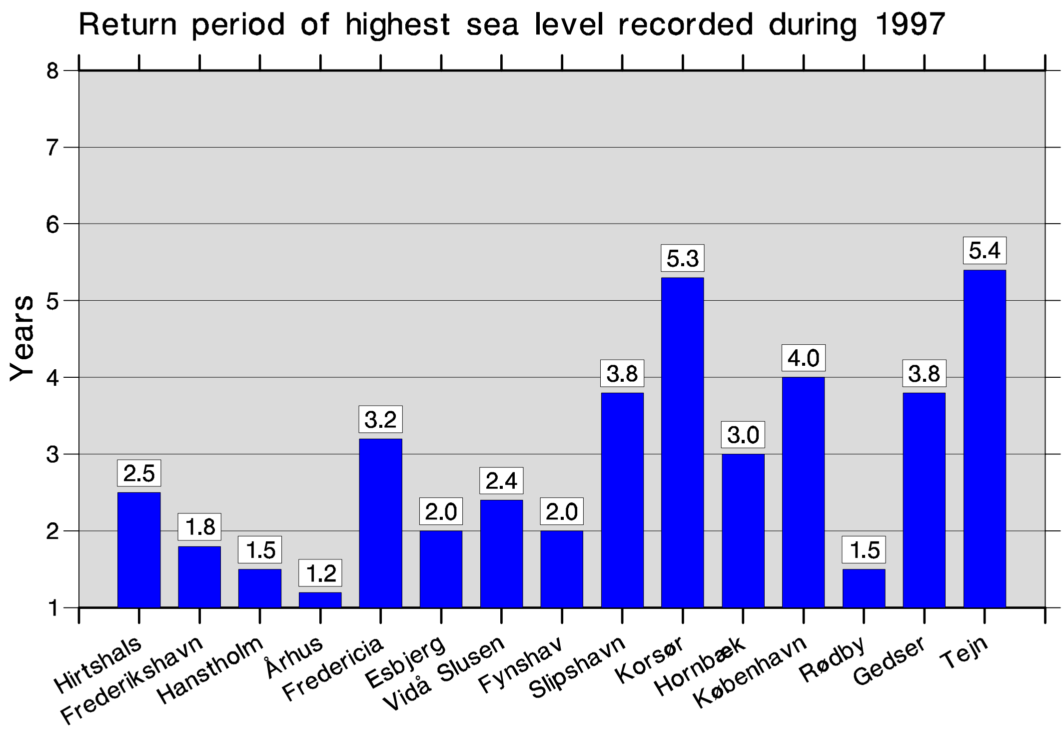


Figure 5: Return periods, 15 DMI stations, 1997.

5 Error measures

Two types of error measures are considered (Fig. 6):

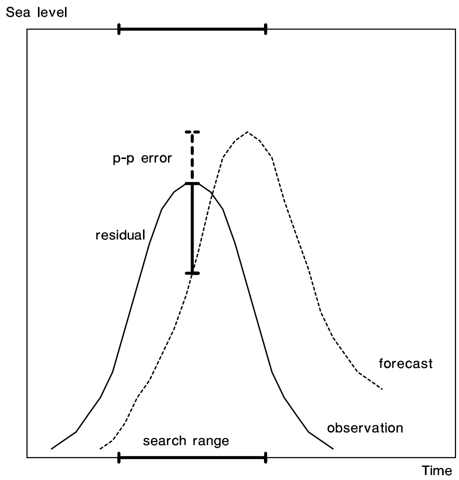


Figure 6: Definition of the residual and the peak error.

- *residual error measures*, based on forecast minus observation at a given time
- *peak-to-peak error measures*, based on the forecasted sea level high minus observed sea level high, allowing for a finite phase error. This may be larger or smaller than the residual at peak time.

Since no simple dependence on the forecast range has been found [3], range-averaged monthly error measures are calculated for each station, using forecast pseudo time series as described above. Subsequently, 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 forecast error time series. The model datum and the local datum of each tide gauge may differ considerably, and the residuals may be unbiased by subtracting a mean error for every station in question. The (possibly bias corrected) residual is used to calculate error measures. We define

$$\begin{aligned} RES(i) &= FRC(i) - OBS(i) \\ (RES_{unbiased}(i) &= RES(i) - BIAS) \end{aligned}$$

where $BIAS$ is the 1994-96 mean error for each station. Monthly error measures me = mean error, mae = mean absolute error, rms = root mean square error, max = maximum error and $e.v.$ = explained variance are defined as (overbar denote one-month average)

$$me = \overline{RES}$$

$$
\begin{aligned}
mae &= \overline{|RES|} \\
rms &= \sqrt{\frac{N}{N-1}\overline{RES^2}} \\
max &= MAX(|RES(1)|, ..., |RES(N)|) \\
e.v. &= 1 - \frac{Var(RES)}{Var(OBS)}
\end{aligned}
$$

where $Var(x) = \overline{x^2 - \overline{x}^2}$. Each error measure is calculated for every station, and for every month. Annual error measures for each station are calculated as twelve-month averages or maxima. These are denoted by upper-case letters $ME, MAE, RMS, MAX, E.V.$. For each item, a single annual error measure is then 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

To verify predictions of extreme (high) water levels, residual statistics are not adequate. A small phase error may lead to a rather large residual (Fig. 6). To evaluate the model performance in this respect we use the peak error, defined as

$$
PE(t_0) = FRC(t) - OBS(t_0)
$$

where t_0 is the time of the observed peak, t is the time of the predicted peak, and $|t - t_0| <$ 6 hours, corresponding to half the diurnal tidal period. If no peak is forecasted within this time window, that sea level high is disregarded.

For each station, the 10 highest observed peaks during 1997 are identified and corresponding peak errors are calculated. Annual peak error measures $MPE =$ mean peak error, $MAPE =$ mean absolute peak error, and $MXPE =$ maximum absolute peak error are defined by (overbar indicate average over 10 highest peaks)

$$
\begin{aligned}
MPE &= \overline{PE} \\
MAPE &= \overline{|PE|} \\
MXPE &= MAX(|PE_1|, ..., |PE_{10}|)
\end{aligned}
$$

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

Given some success criterion, we count the number of well predicted peaks. This is accumulated for the 10 highest peaks at all stations, totalling 220 peaks. The resulting *peak hit rate*, with equal weight on all stations, is

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

Success criteria of 10cm, 15cm and 20cm are employed.

6 Results

Station-averaged results are presented in Table 8 below, while annual averages for each station are shown in the Figures at the end of this section. Unbiased forecasts and update set-up (M21) results are both included below but discussed in separate sections. Figs. 7-11 show the residual statistics for each station, Figs. 12-14 show the peak statistics for each station, and Figs. 15-22 show peak errors at the warning stations[1].

Annual station averages (1997)	DKV	UKM	Unbias	M21
Average mean error	-1.5 cm	-0.3 cm	-1.4 cm	-5.1 cm
Average absolute mean error	5.3 cm	5.3 cm	3.0 cm	7.7 cm
Average mean absolute error	11.7 cm	12.1 cm	10.8 cm	13.1 cm
Average rms error	14.4 cm	14.8 cm	13.3 cm	15.7 cm
Average maximum error	68.2 cm	68.5 cm	64.3 cm	73.4 cm
Average explained variance	74.3 %	73.8 %	74.3 %	72.7 %
Average mean peak error	-11.5 cm	-14.9 cm	-10.6 cm	-29.0 cm
Average absolute mean peak error	13.2 cm	15.3 cm	10.6 cm	29.8 cm
Average mean absolute peak error	16.7 cm	24.3 cm	15.3 cm	29.7 cm
Average maximum peak error	37.8 cm	49.4 cm	36.5 cm	52.2 cm
Peak hit rate (10cm criterion)	50.1 %	45.0 %	54.6 %	40.3 %
Peak hit rate (15cm criterion)	68.0 %	61.4 %	72.1 %	57.5 %
Peak hit rate (20cm criterion)	81.7 %	76.0 %	83.5 %	71.0 %

Table 8: 1997 station-averaged annual error measures

The mean error (ME) (Fig. 7)

The average ME is very small (less than 2 cm) and negative, indicating that the predicted sea level is on average slightly too low. The UKM predicted mean sea level is a little higher than the DKV predicted mean sea level at all stations. At individual stations the ME ranges from -16 cm (Fynshav) to +14 cm (Hirtshals). The set-ups differ at most a few cm, mostly so at the three Baltic stations, Rødby, Gedser and Tejn. Positive ME (indicating that the predicted sea level on average is too high) is found at 9 stations, negative ME at 10 stations with no simple geographical distribution. Very small ME (less than 1 cm) is found at Thyborøn, Esbjerg, Havneby, Hornbæk and Tejn.

The absolute mean error (AME)

The average AME is around 5 cm for both set-ups.

The mean absolute error (MAE) (Fig. 8)

The average MAE is around 12 cm, the DKV error being a little smaller than the UKM error. At individual stations the MAE ranges from 6 cm (Tejn) to 17 cm (Fynshav). The set-ups differ 2 cm or less, with the DKV set-up better at 13 stations, the UKM set-up at 6 stations. The MAE decreases from the Jutland West Coast stations towards the Baltic stations, Fynshav being an exception.

[1] Please note that in the Figures, the UKM set-up is referred to as 'LAM'.

The root mean square error (RMS) (Fig. 9)

The average RMS error is 14-15 cm, the DKV error being a little smaller than the UKM error. At individual stations the RMS ranges from 7 cm (Tejn) to 20 cm (Fynshav). The set-ups differ 2 cm or less, with a spatial distribution similar to the MAE.

The maximum error (MAX) (Fig. 10)

The MAX is based on one event per station, and depends strongly on the data coverage (more data gives higher MAX error). The average MAX is 68-69 cm for both set-ups. At individual stations, MAX ranges from 26 cm (Tejn) to 97 cm (Esbjerg, Havneby), with set-up differences up to 25 cm. The DKV set-up is better at 8 stations (in the Kattegat, the Belt Sea and the Baltic), the UKM set-up at 10 stations (along the West Coast and in the Skagerrak). The MAX error is generally in the range 60-80 cm, with high values in the Wadden Sea and low values in the Baltic and the Belt Sea.

The explained variance (E.V.) (Fig. 11)

The E.V. is free of any bias induced error. The average E.V. is around 74 %, the DKV set-up being a little better than the UKM set-up. At individual stations, the E.V. ranges from 48 % (København) to 95 % (Havneby), with the set-ups differing up to 6 %. The DKV set-up is better at 11 stations (mainly in the Belts and the Sound), the UKM set-up at 8 stations (mainly in the Baltic). At Skagerrak, West Coast and Wadden Sea stations, the set-ups are almost equally good. The E.V. is high in the Wadden Sea and in the Baltic, low in the narrows of the Sound and the Little Belt.

The mean peak error (MPE) (Fig. 12)

The average MPE is negative, indicating that prediction of very high sea level is on average too low. The DKV set-up MPE is -11.5 cm, the UKM set-up is 2.5 cm worse. At individual stations, the MPE ranges from -25 cm (Hvide Sande, Fynshav, København) to +12 cm (Hirtshals), with noticeable larger negative MPE at most Domestic stations for the UKM set-up. The DKV set-up has numerically smaller MPE at 12 stations, the UKM set-up at 8 stations. Very high sea level peaks are underpredicted at nearly all stations, but this varies widely from station to station. Positive MPE is found only at Hirtshals, negligible MPE at Frederikshavn, Hanstholm and Tejn, and negative MPE at the rest of the stations.

The absolute mean peak error (AMPE)

The average AMPE is almost identical to the negative of the average MPE, because peaks are underpredicted at nearly all stations. Average AMPE is 13 cm for the DKV set-up, a few cm worse for the UKM set-up.

The mean absolute peak error (MAPE) (Fig. 13)

The average MAPE is around 17 cm for the DKV set-up, but markedly higher (24 cm) for the UKM set-up. At individual stations, the MAPE ranges from 5 cm (Tejn) to 24 cm (Torsminde, Hvide Sande, Fynshav and København), with noticeable larger MAPE at most stations for the UKM set-up. The set-ups may differ by up to 15 cm. The DKV set-up is better at 20 stations, the UKM set-up only at 2 stations.

The maximum peak error (MXPE) (Fig. 14)

Like the MAX error, the MXPE is based on a single event per station and depends on data coverage. The average MPPE is around 38 cm for the DKV set-up, but markedly higher (49 cm) for the UKM set-up. At individual stations, the MXPE ranges from 18 cm (Tejn) to 59 cm (Torsminde), with UKM set-up errors up to 40 cm larger. The DKV set-up is better at 18 stations, the UKM set-up at 2 stations.

The peak hit rate (PGOOD) (Table 8)

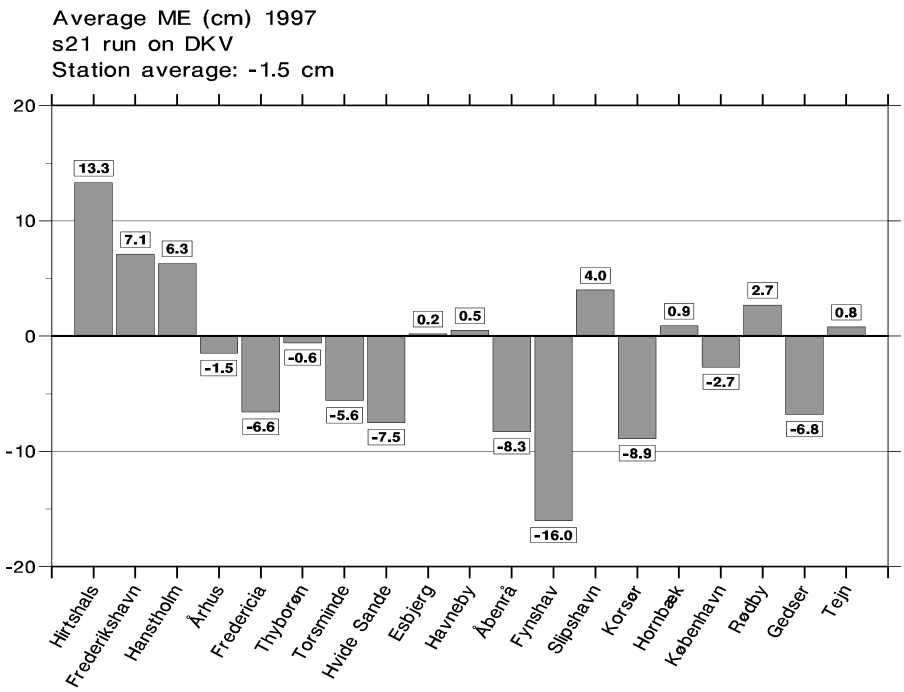
The hit rate PGOOD is not evaluated at single stations. The PGOOD is 4-6 % higher for the DKV set-up than for the UKM set-up, regardless of success criterion. The DKV peak hit rate is roughly 1 out of 2 (10 cm criterion), 2 out of 3 (15 cm criterion), and 5 out of 6 (20 cm criterion).

Peak errors at the warning stations (Figs. 15-22)

For the 8 warning stations (Table 5), the ten highest peaks and peak errors are shown in order of decreasing peak magnitude. A positive error indicates that the forecast was too high, a negative that it was too low. The DKV results for each station is summarised below, using a 10 cm success criterion for the peak hit rate.

Torsminde: 6 peaks are underpredicted, by up to 60 cm. 4 peaks are correctly predicted.
Esbjerg: 3 peaks are underpredicted by up to 50 cm. 7 peaks are correctly predicted.
Vidå: 4 peaks are underpredicted by up to 50 cm. 6 peaks are correctly predicted.
Frederikshavn: Peaks are over- or underpredicted, by up to 22 cm. 6 peaks are correctly predicted.
Åbenrå: All peaks are underpredicted, by up to 44 cm. 3 peaks are correctly predicted.
Korsør: Peaks are usually underpredicted, by up to 35 cm. 1 peak is correctly predicted.
Rødby: Peaks are over- or underpredicted, by up to 23 cm. 6 peaks are correctly predicted.
Gedser: Peaks are usually underpredicted, by up to 30 cm. 6 peaks are correctly predicted.

With an average peak hit rate of roughly 50 %, we may conclude that the peak hit rate at Torsminde, Åbenrå and Korsør is below average, while at Esbjerg, Vidå, Frederikshavn, Rødby and Gedser the peak hit rate is above average.



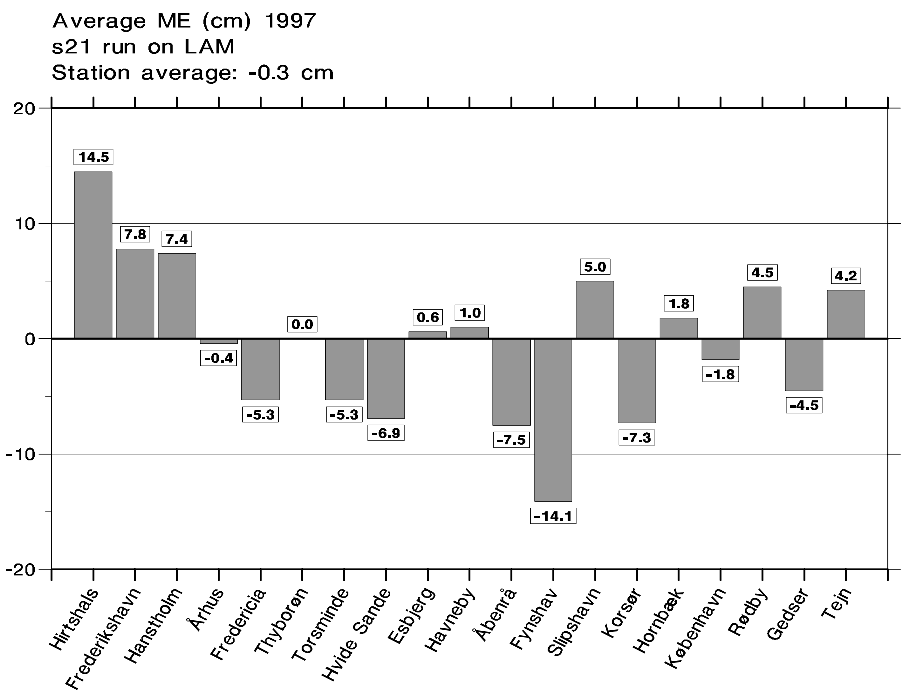
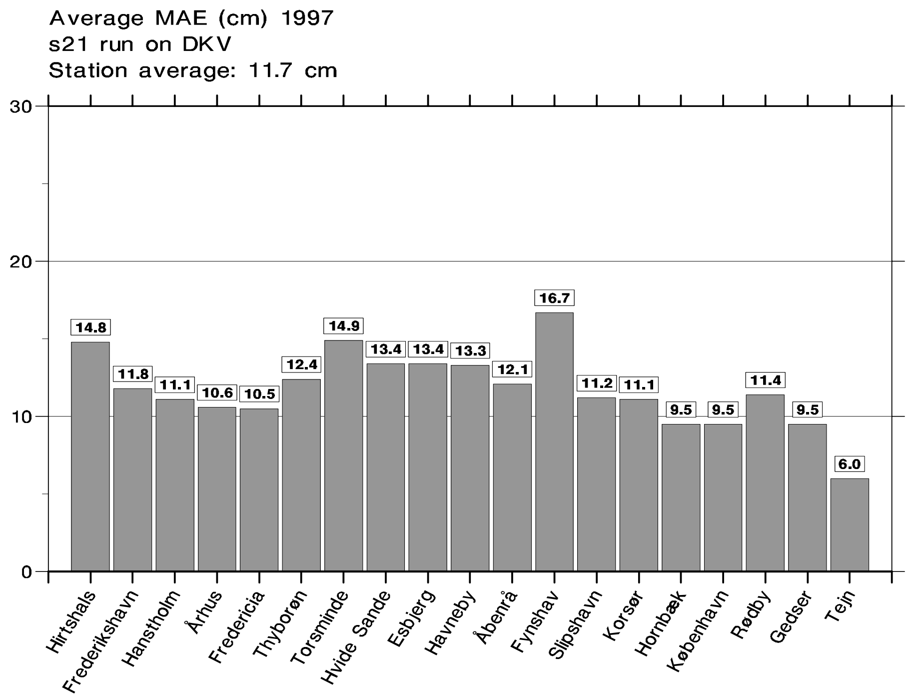


Figure 7: Average mean error at 19 stations, 1997.



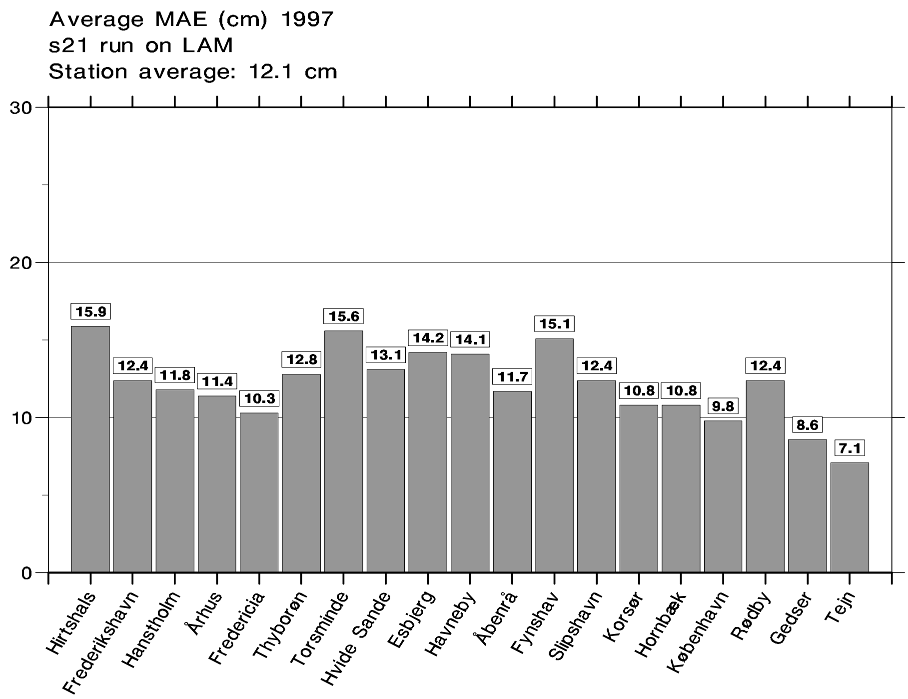
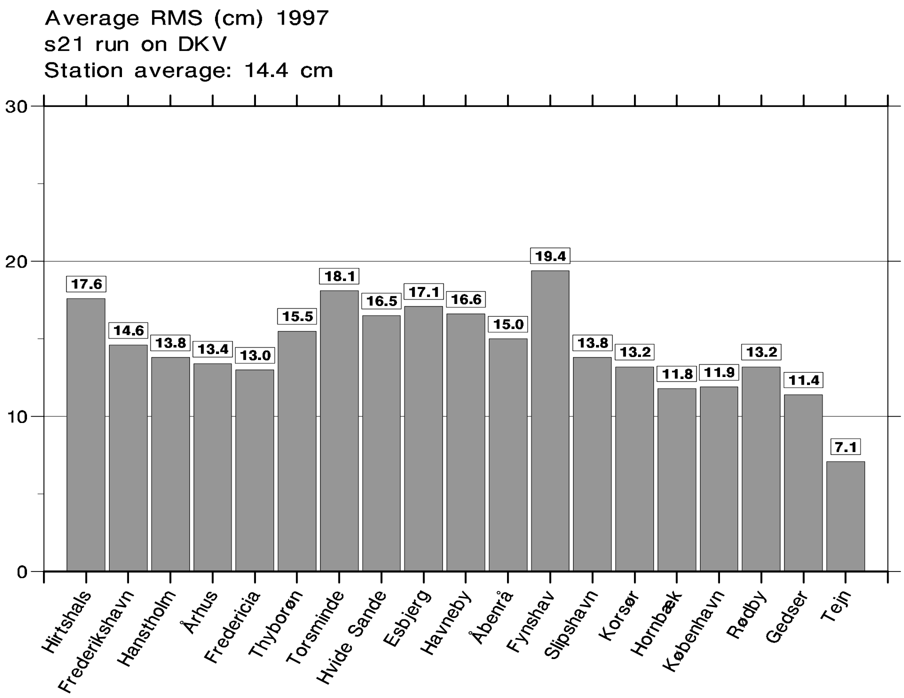


Figure 8: Average mean absolute error at 19 stations, 1997.



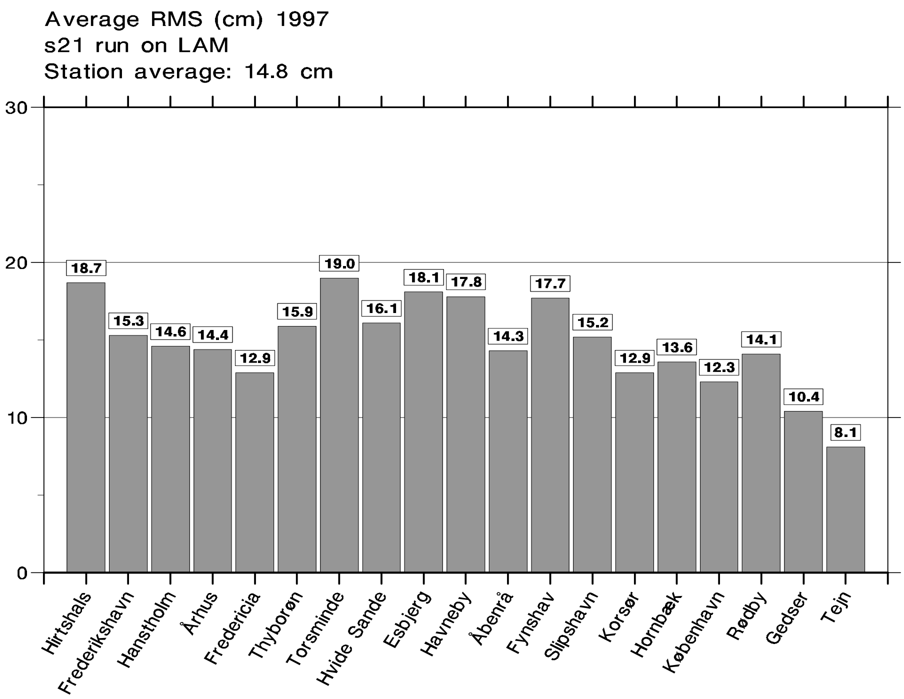
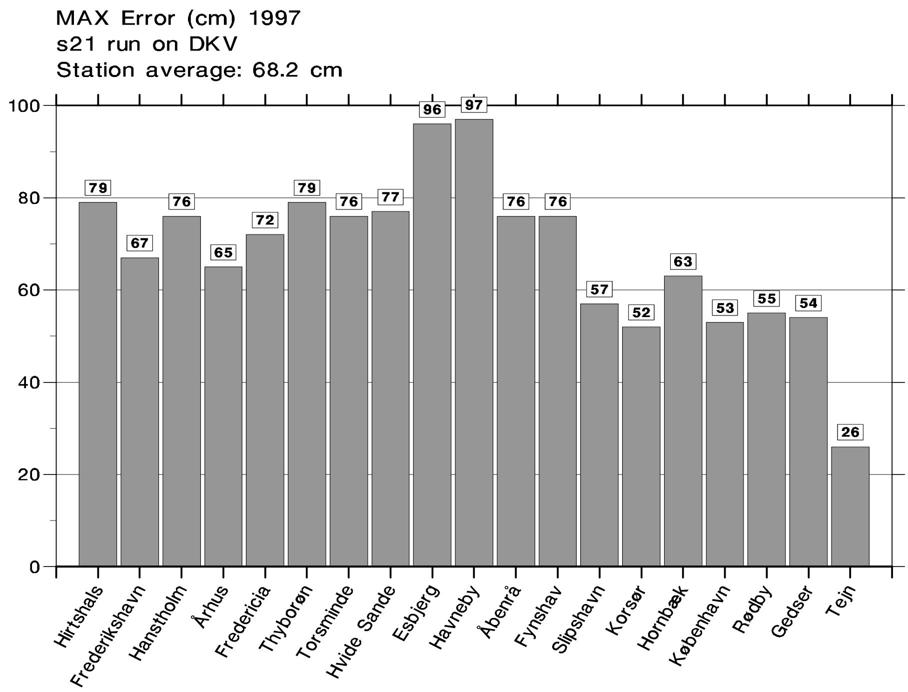


Figure 9: Average RMS error at 19 stations, 1997.



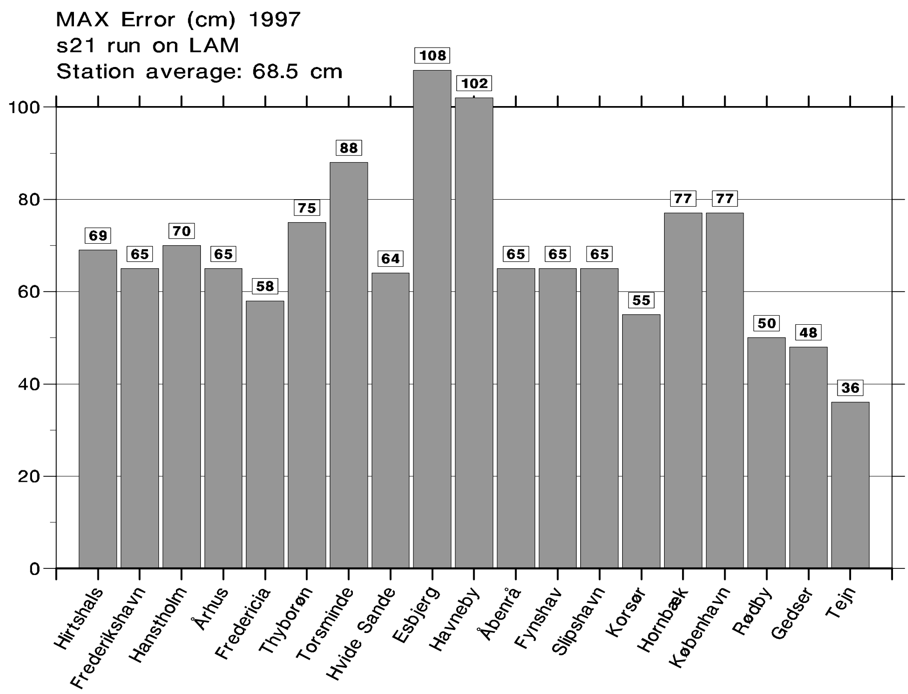
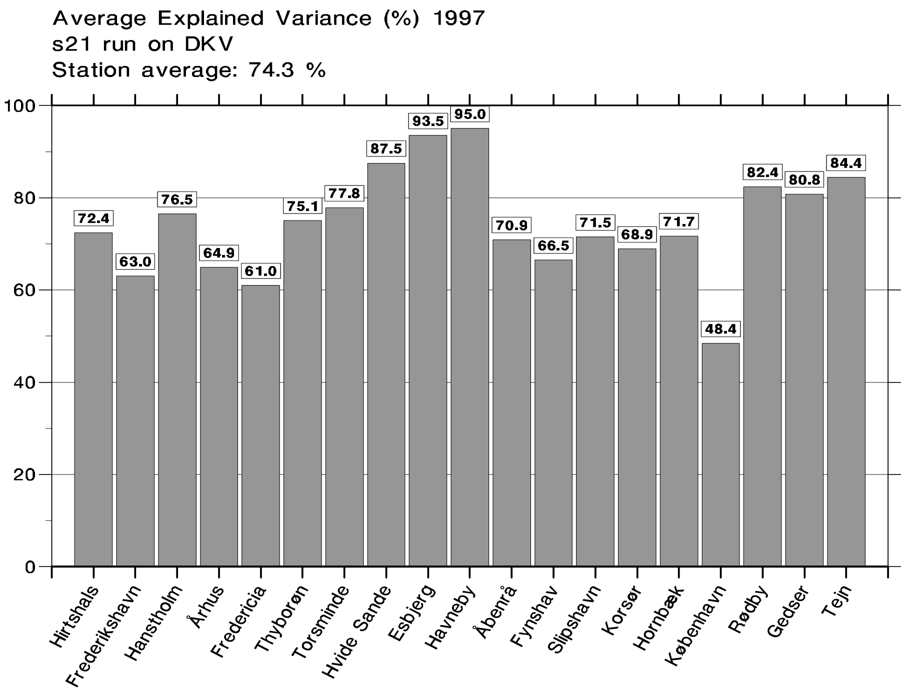


Figure 10: Maximum error at 19 stations, 1997.



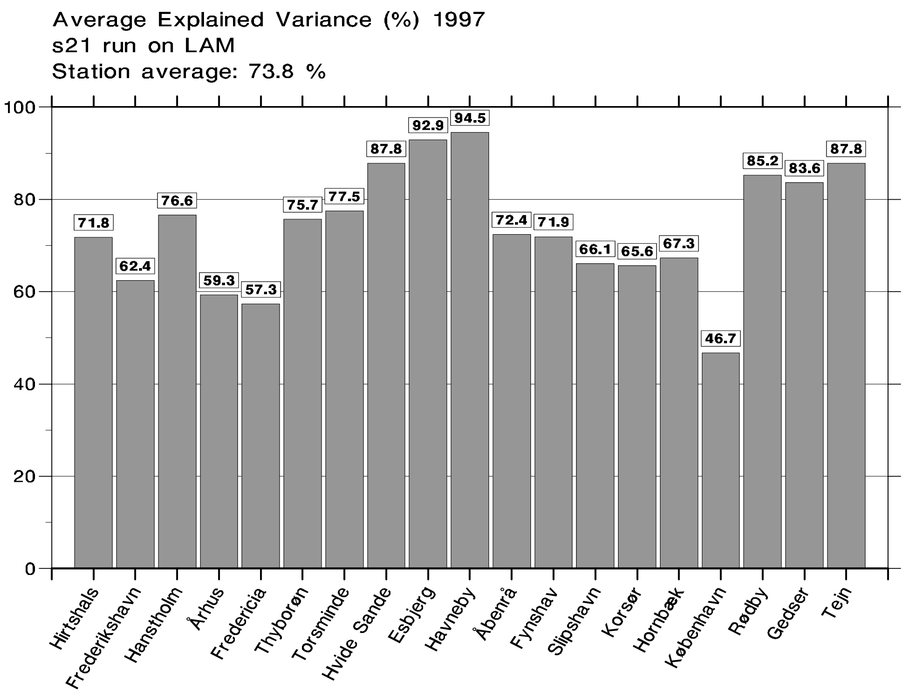
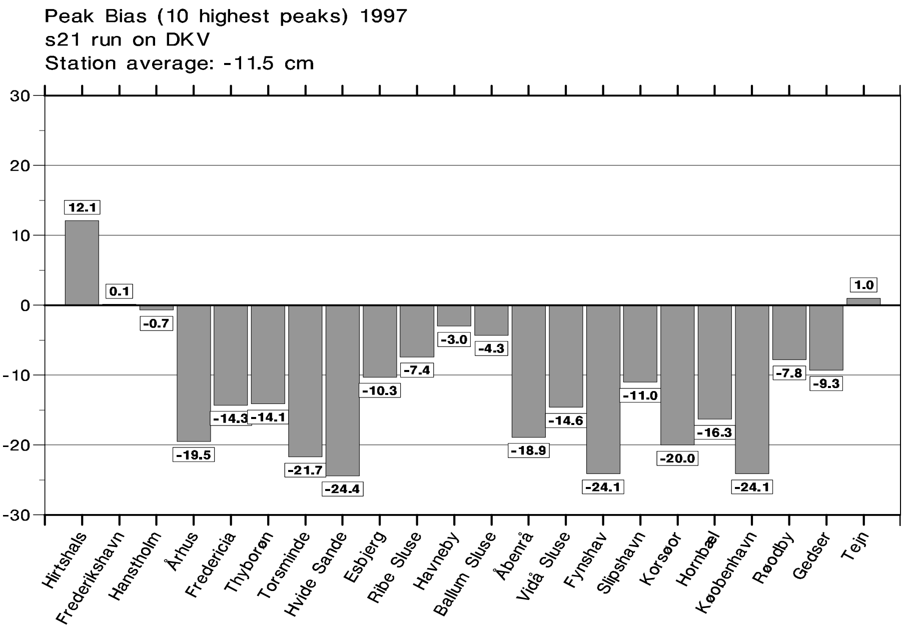


Figure 11: Explained variance at 19 stations, 1997.



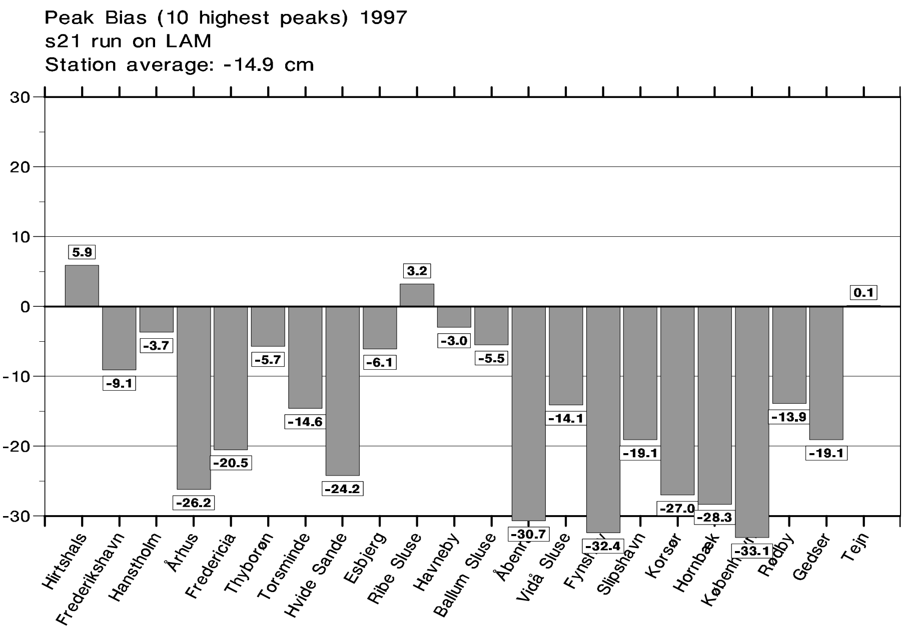
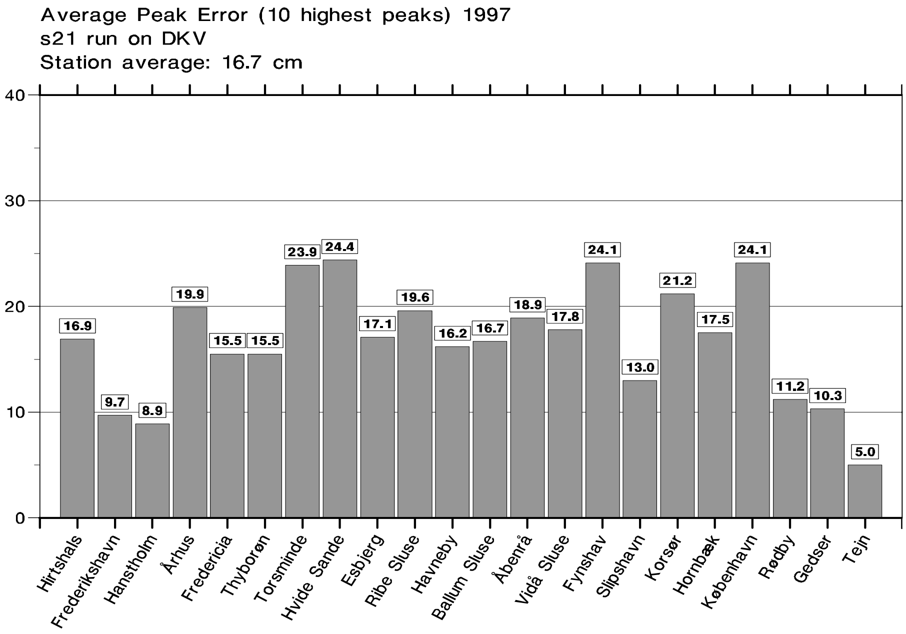


Figure 12: Mean peak error at 22 stations, 1997.



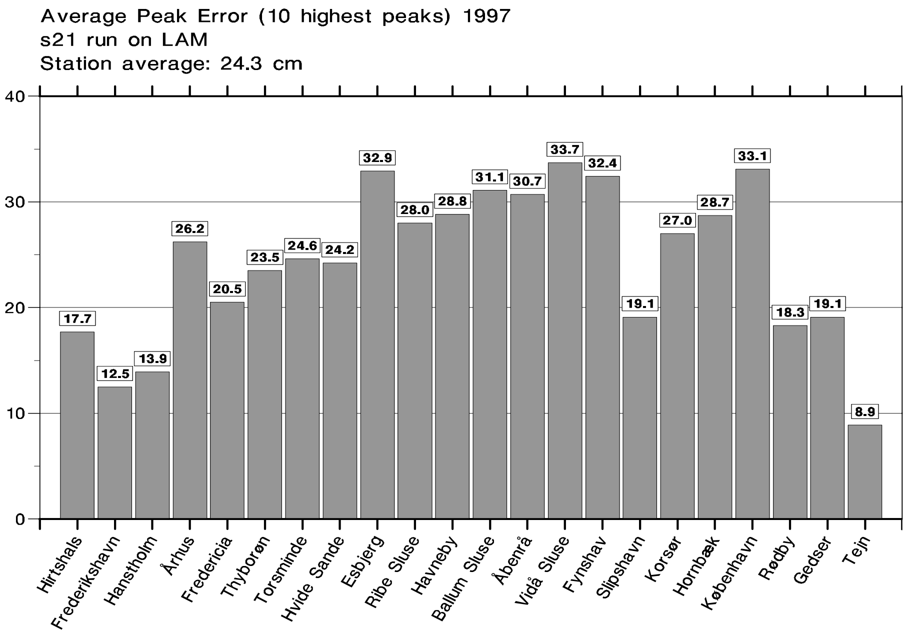


Figure 13: Average absolute peak error at 22 stations, 1997.

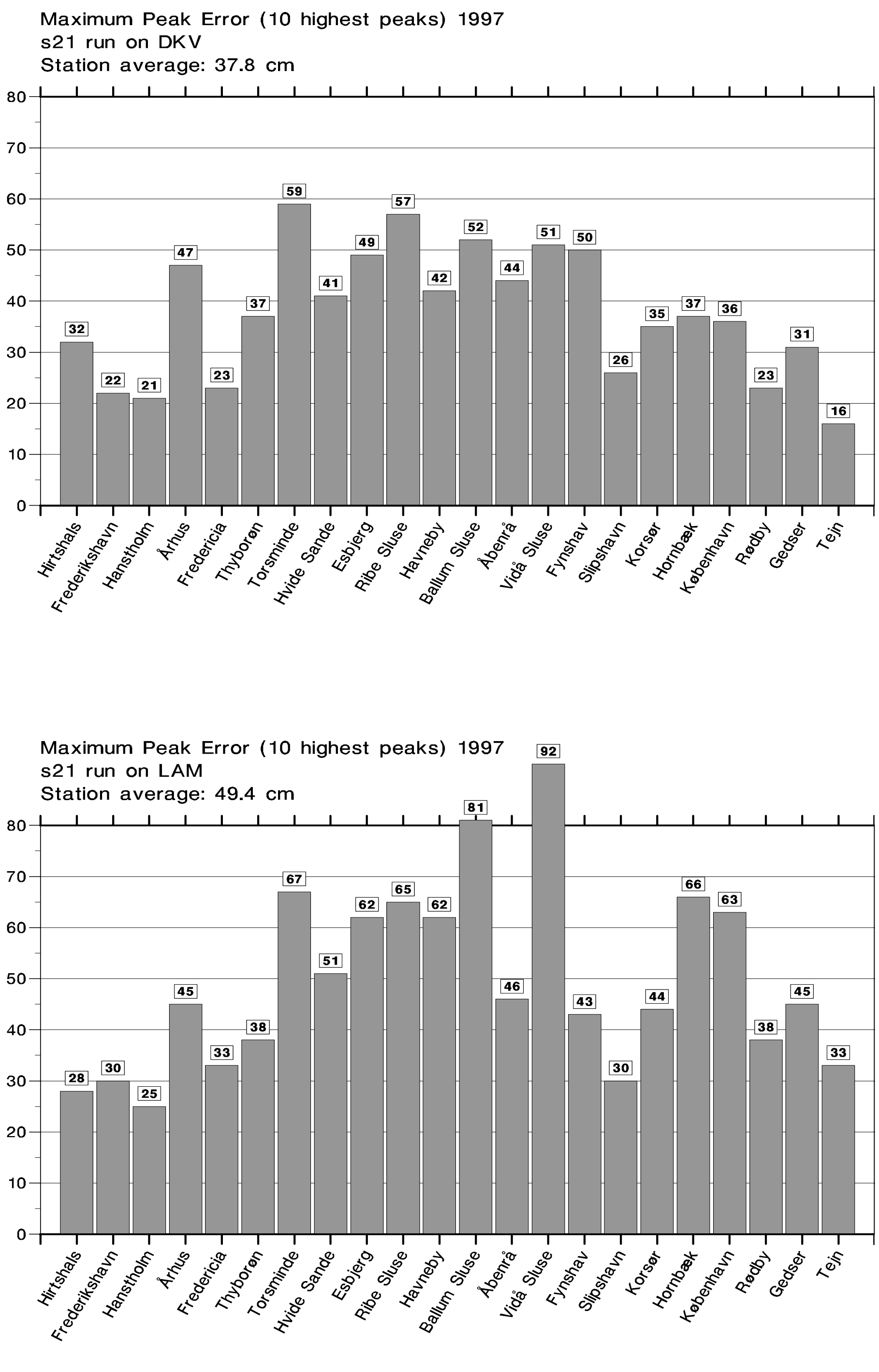
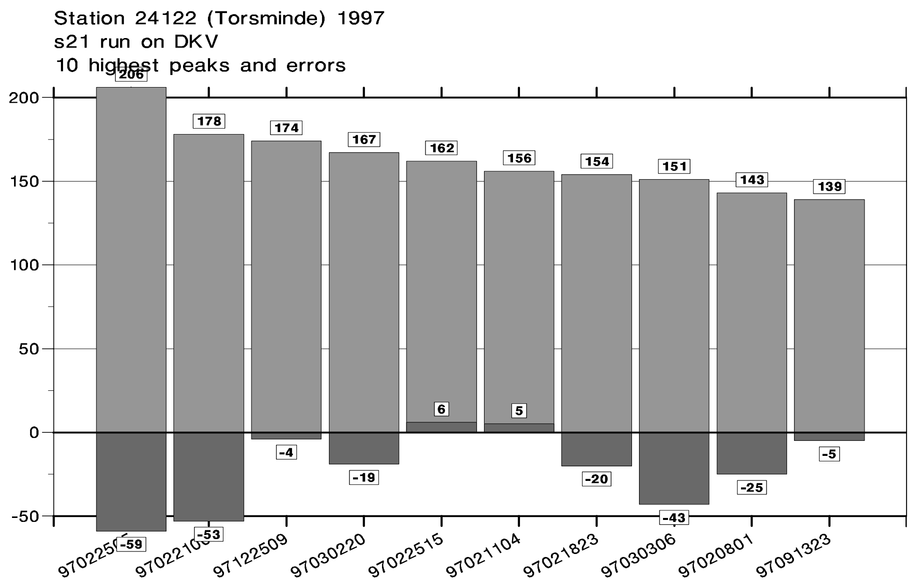


Figure 14: Maximum absolute peak error at 22 stations, 1997.



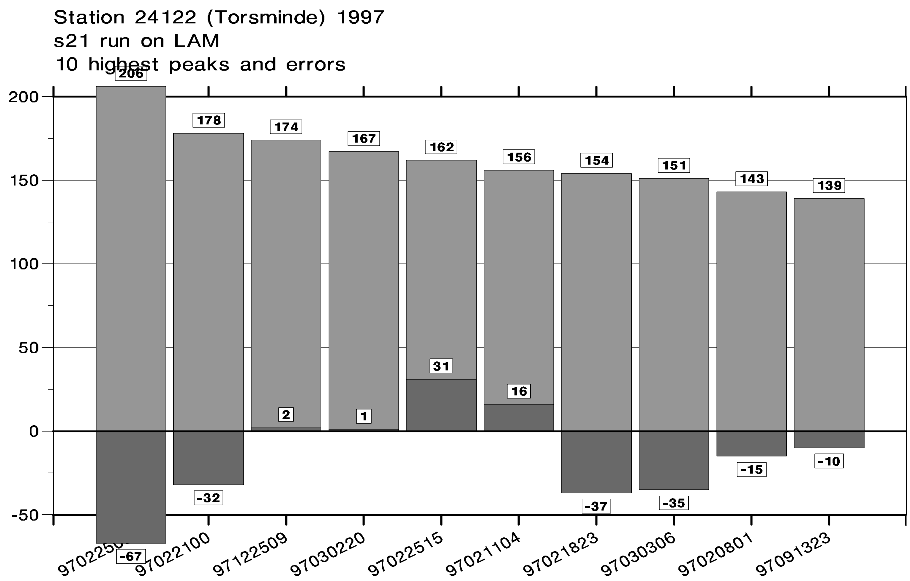
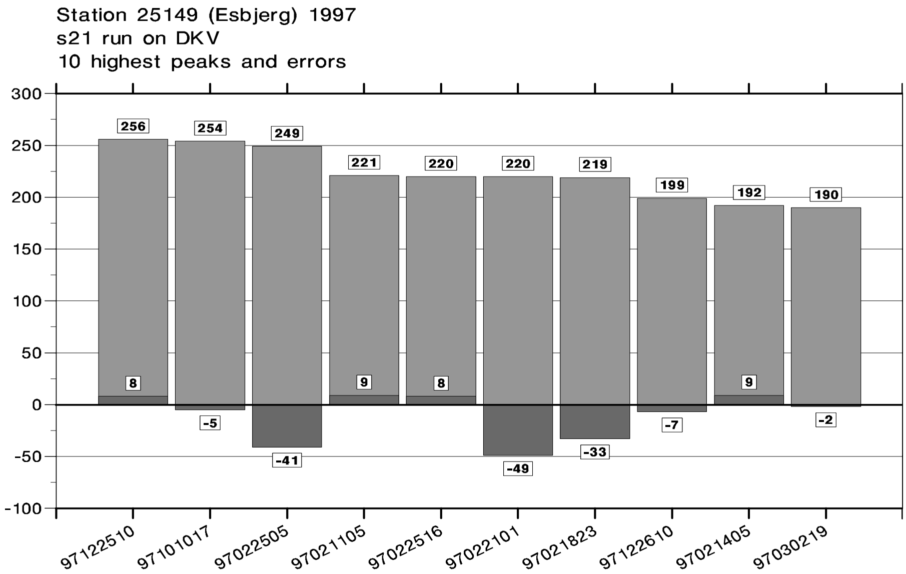


Figure 15: Torsminde 10 highest peaks and errors, 1997.



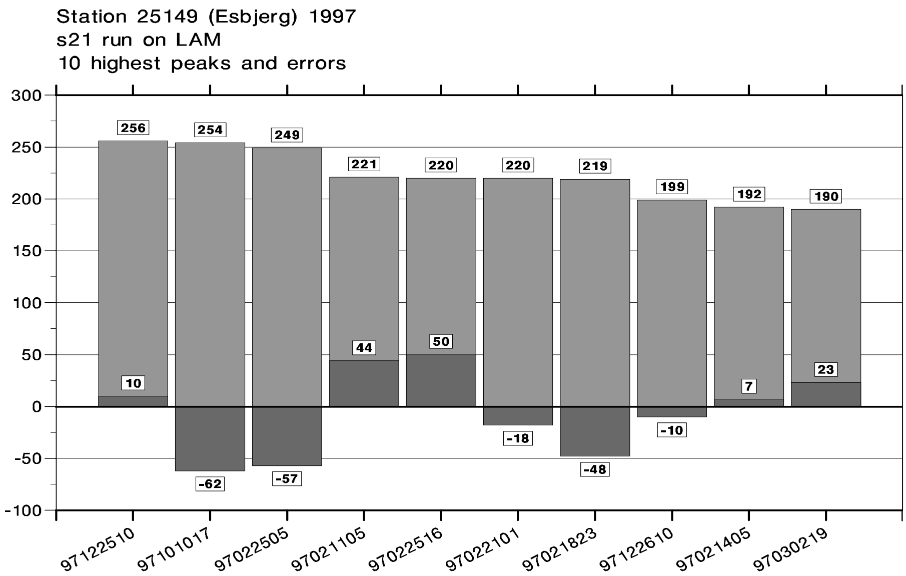
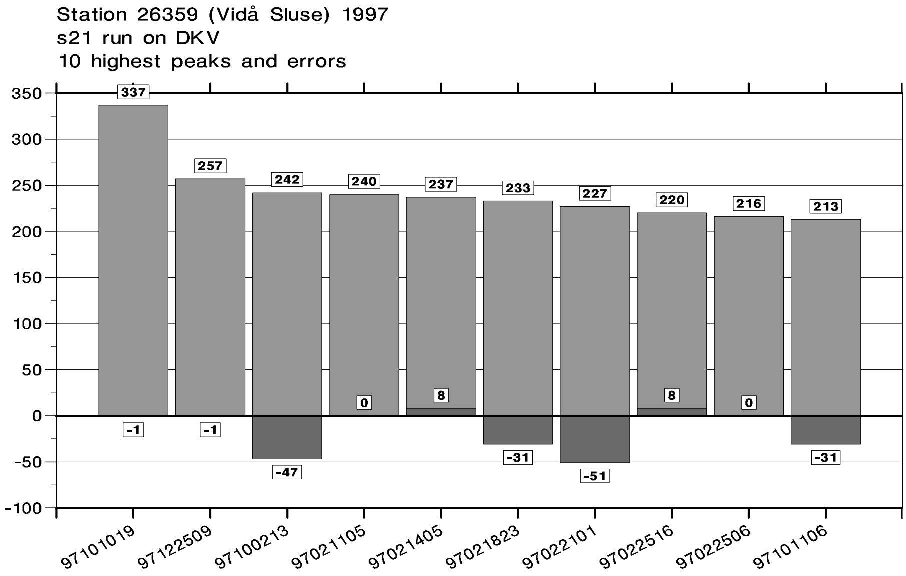


Figure 16: Esbjerg 10 highest peaks and errors, 1997.



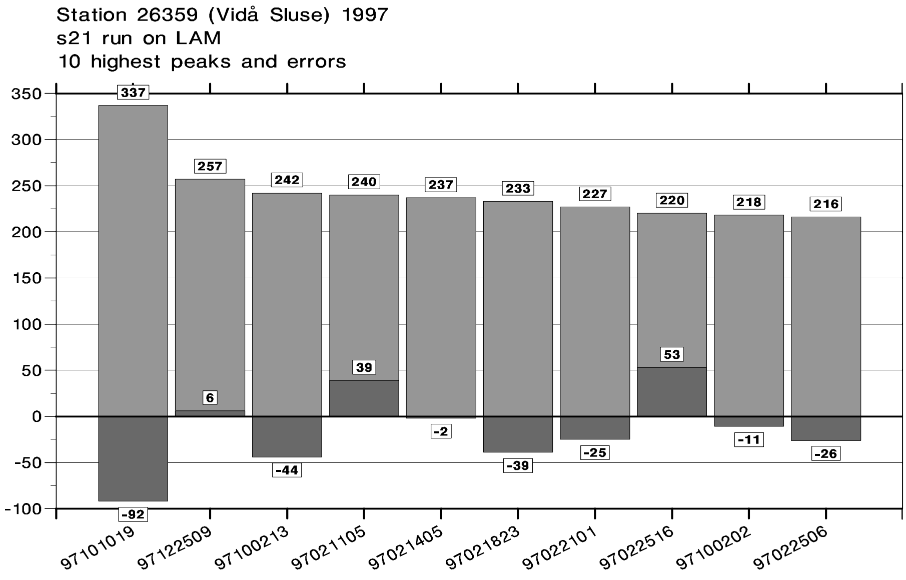
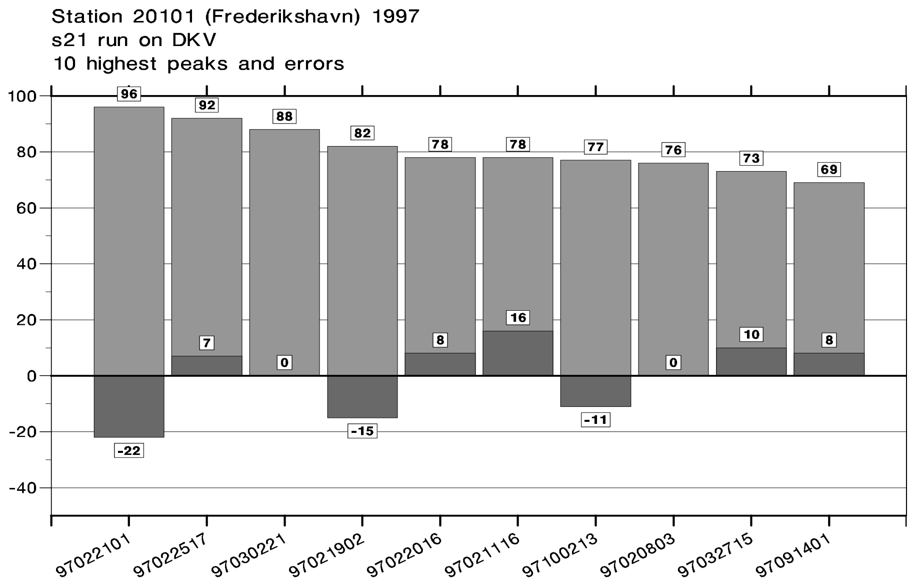


Figure 17: Vidå Sluse 10 highest peaks and errors, 1997.



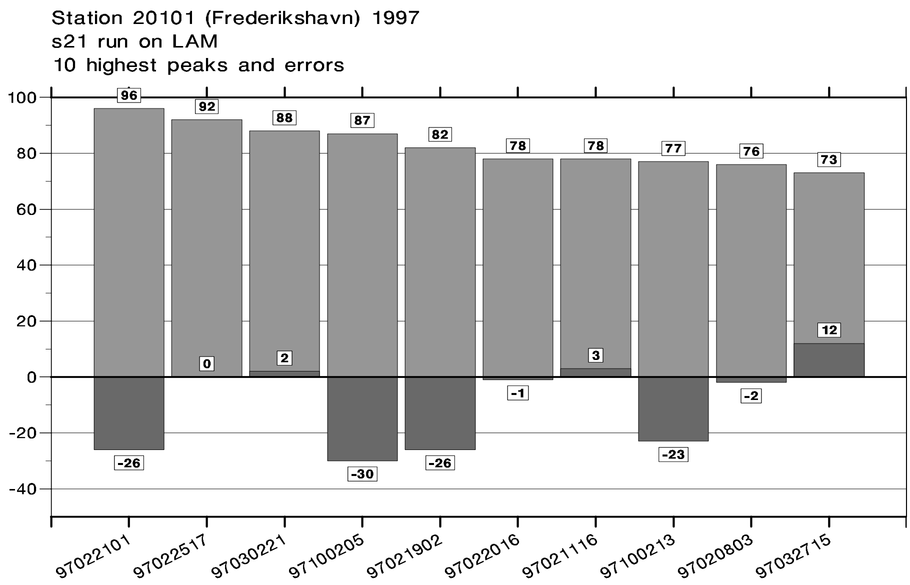
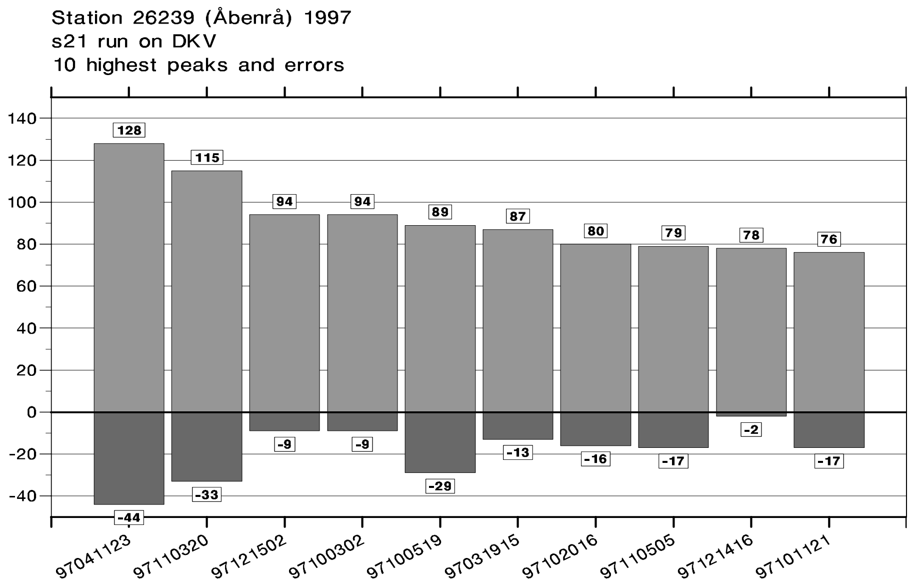


Figure 18: Frederikshavn 10 highest peaks and errors, 1997.



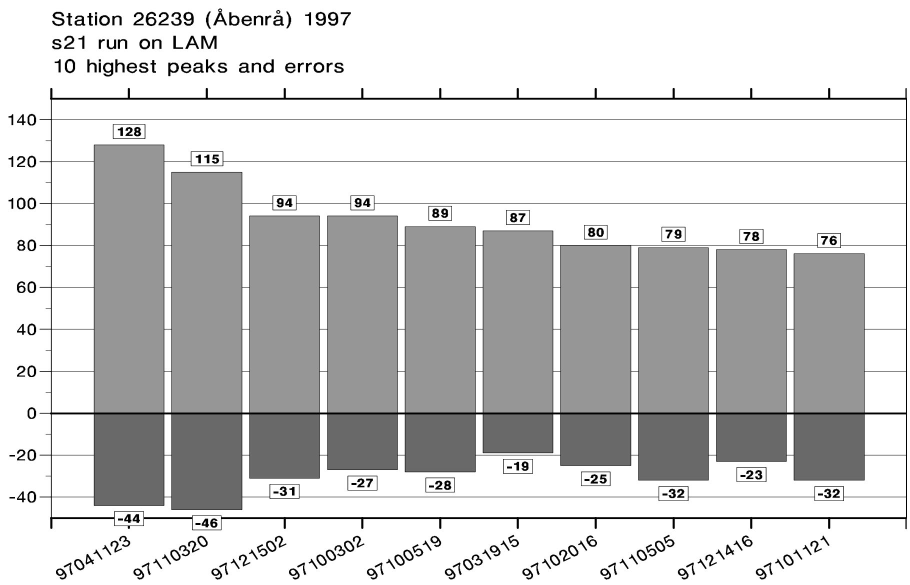
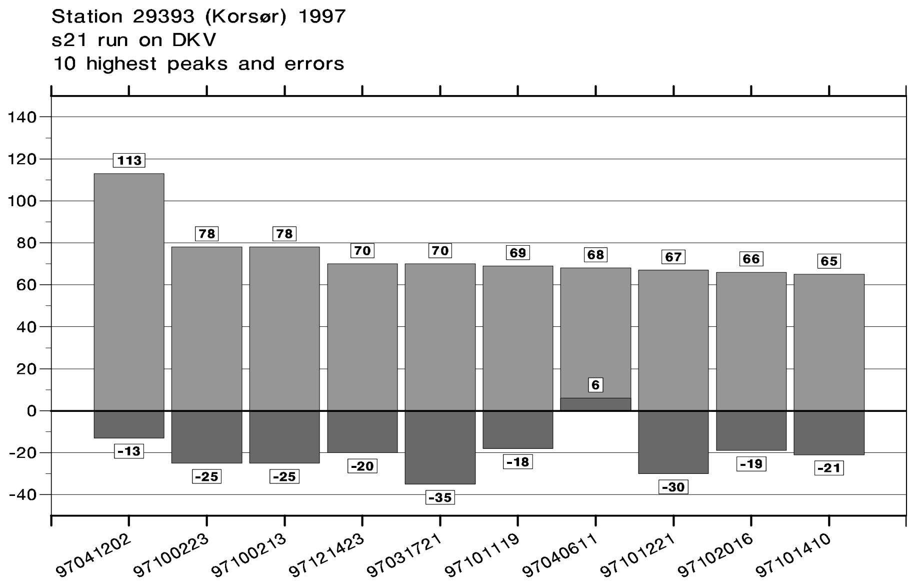


Figure 19: Åbenrå 10 highest peaks and errors, 1997.



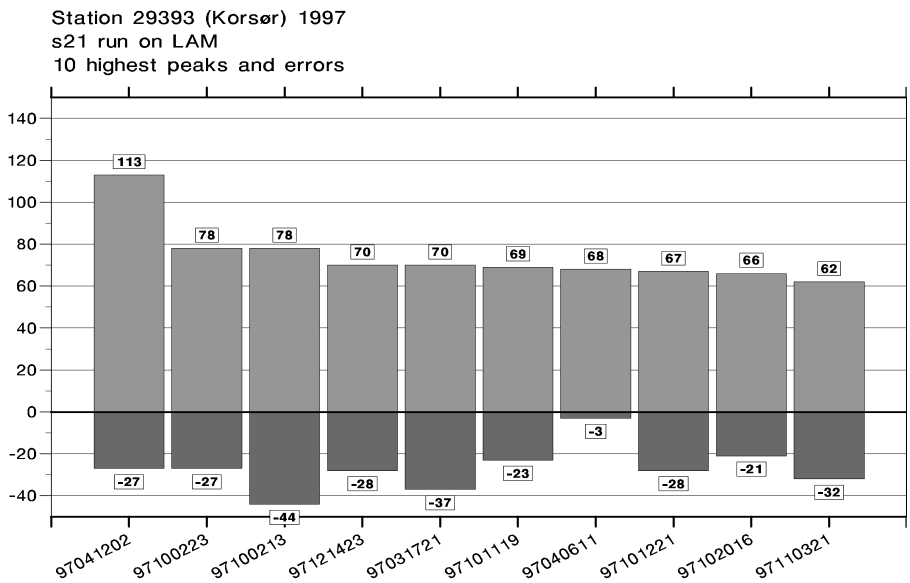
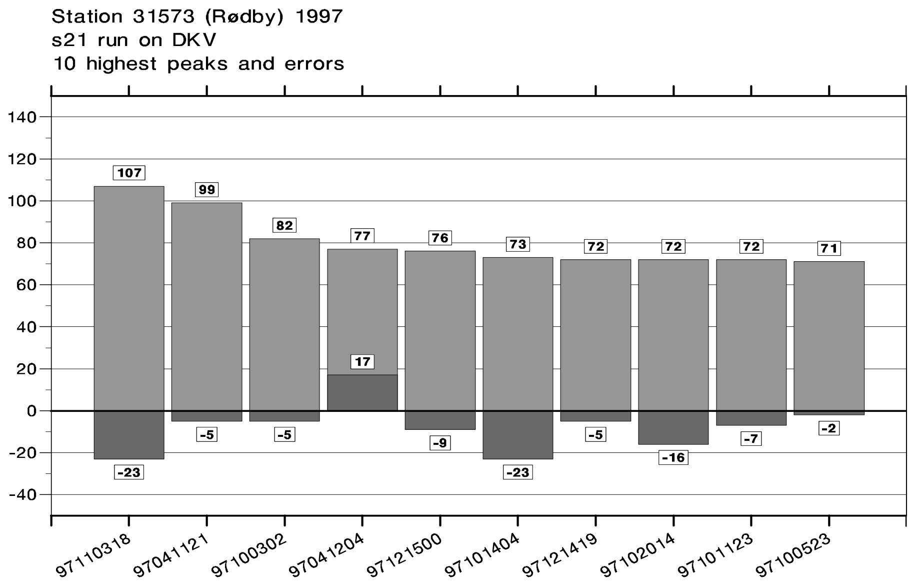


Figure 20: Korsør 10 highest peaks and errors, 1997.



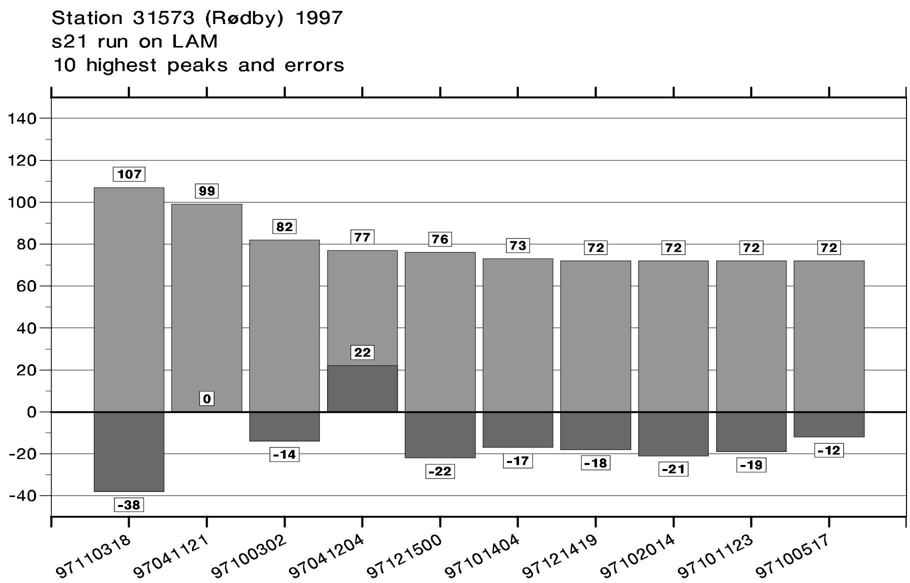
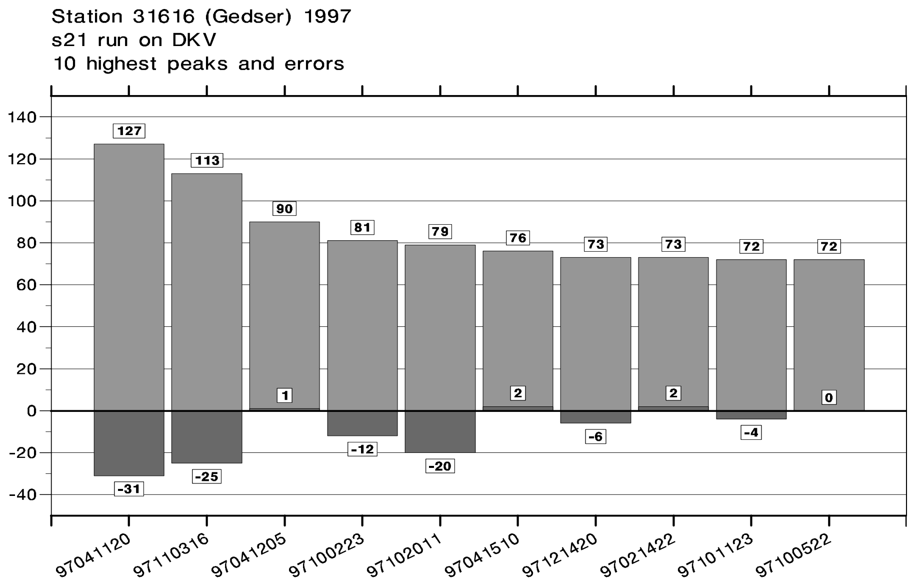


Figure 21: Rødby 10 highest peaks and errors, 1997.



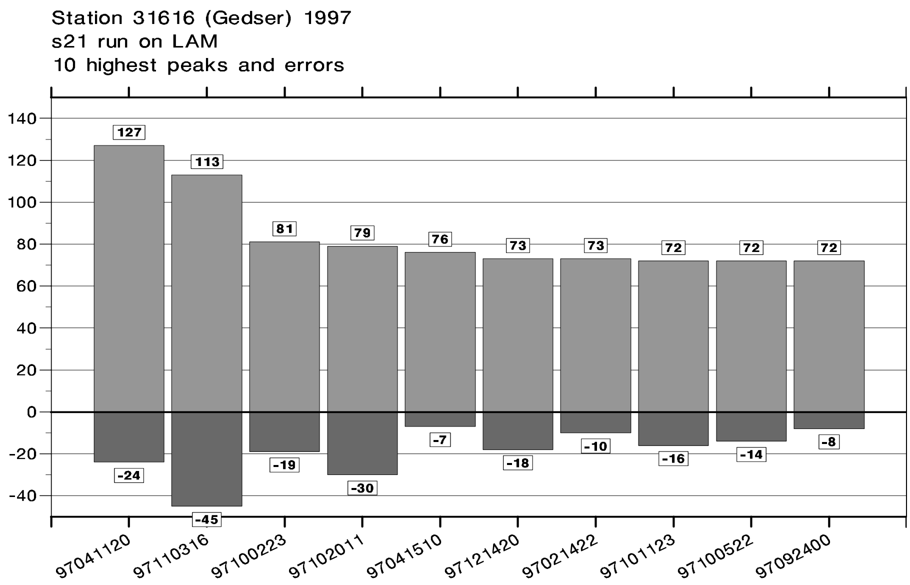


Figure 22: Gedser 10 highest peaks and errors, 1997.

7 Unbiased forecasts

Unbiasing is done by subtracting a mean error. The mean error at individual stations varies considerably from year to year and from month to month, and precisely what value to employ may be subject to improvement. For the present report, we use 1994-96 mean error averages (Table 7). The effect of unbiasing is described in average error measure terms (Table 8), in terms of the range of the error measures at individual stations, and by counting the number of stations where error measures has bettered or worsened due to unbiasing.

The *average ME* is unchanged but the range has narrowed considerably, with major ME present only at Åbenrå (-8 cm) and Rødby (+8 cm). The tide gauge at Åbenrå is rather new and no bias has yet been calculated. At Rødby unbiasing increases the ME. All other stations have ME below 5 cm.

The *average AME, MAE and RMS* improve slightly, mainly due to the large bias corrections at Hirtshals and Fynshav. Unbiasing reduces the MAE at 9 stations, increases it at only 3 stations. Major MAE improvements result at Hirtshals (5 cm) and Fynshav (8 cm), with changes below 2 cm at the rest of the stations. The RMS improves much in the same way. The *average AME* decreases from 5 cm to 3 cm.

The *average MAX* decreases by 4 cm to 65 cm. Unbiasing reduces the MAX by 2-15 cm at 13 stations and increases it slightly at only 3 stations.

The *average AMPE* is inproved by 2 cm, and the MPE is now negative at all stations ranging, from 0 cm to -25 cm. This error measure has improved at 12 stations and worsened at 9 stations. The *average MAPE* decreases by 1 cm, improving at 5 stations and worsening at 2 stations. The *average MXPE* also decreases by 1 cm but its range is unchanged, improving at 11 stations and worsening at 9.

The *peak hit rate* increases by 2-4 %, depending on the success criterion.

Unbiasing leads to better forecasts in on average, when the bias value is selected carefully. This holds both for residual error measures and for peak error measures. The improvement is substantial at a few stations (those with a large bias correction) but the forecasts get slightly worse at other stations, when the mean error has a large year-to-year variation. It should be noted that the above discussion retain to average conditions and at a given instant there is no guarantee that unbiasing improves the forecast.

8 Comparison with previous years

The 1994-97 time development of nine station-averaged statistical measures is shown in Fig. 23. In addition to this, some direct comparison at individual stations with 1994-96 results is made (cf. [4], [6] and [9]). The following inhomogeneities are noted:

- In 1994, 17 stations were used, and the months January-March and missing
- In 1995, 21 stations were used (Havneby, Ribe, Vidå and Ballum included)
- In 1996, 21 stations were used, and the months August-September are missing
- In 1997, 22 stations were used (Åbenrå included), and the month of July is missing

Keeping this in mind, we may conclude the following:

ME: The average ME stays within the $\pm$ 2 cm range and it seems that 1996 was an odd year. The bias at individual stations (not shown) displays rather large year-to-year variations.

MAE: The average MAE stays constant, about 12 cm for both set-ups. A slight decrease may be noted in the DKV set-up.

RMS: The average RMS error is constant about 15 cm for both set-ups.

MAX: The average MAX error has decreased from 75-80 cm in 1995-96 to just below 70 cm in 1997, for both set-ups.

E.V.: The 1997 average E.V. is higher than any of the previous years, for both set-ups. In 1996, the average E.V. reached a low point.

MPE: Very high peaks are under-predicted by 10-15 cm on average. In 1997 least so by the DKV set-up.

MAPE: The average MAPE is usually about 25 cm. In 1997, the DKV error decreased by about 8 cm compared to previous years. 1994 results are not representative because 4 Wadden Sea stations were not included yet.

MXPE: The average maximum peak error has decreased by more than 10 cm in both set-ups. In 1997, no station has MXPE in excess of 80 cm. As for the MAPE, 1994 results are not representative.

PGOOD: The 1997 DKV peak hit rate increased to 50 %, while the UKM peak hit rate stays constant about 45 % (10 cm success criterion).

In 1997, DKV peak forecasts improved compared to previous years, while the UKM forecasts did not. Residual error measures stay very much the same from year to year for both set-ups, except for a slight decrease of the maximum error and an increase of the average explained variance.

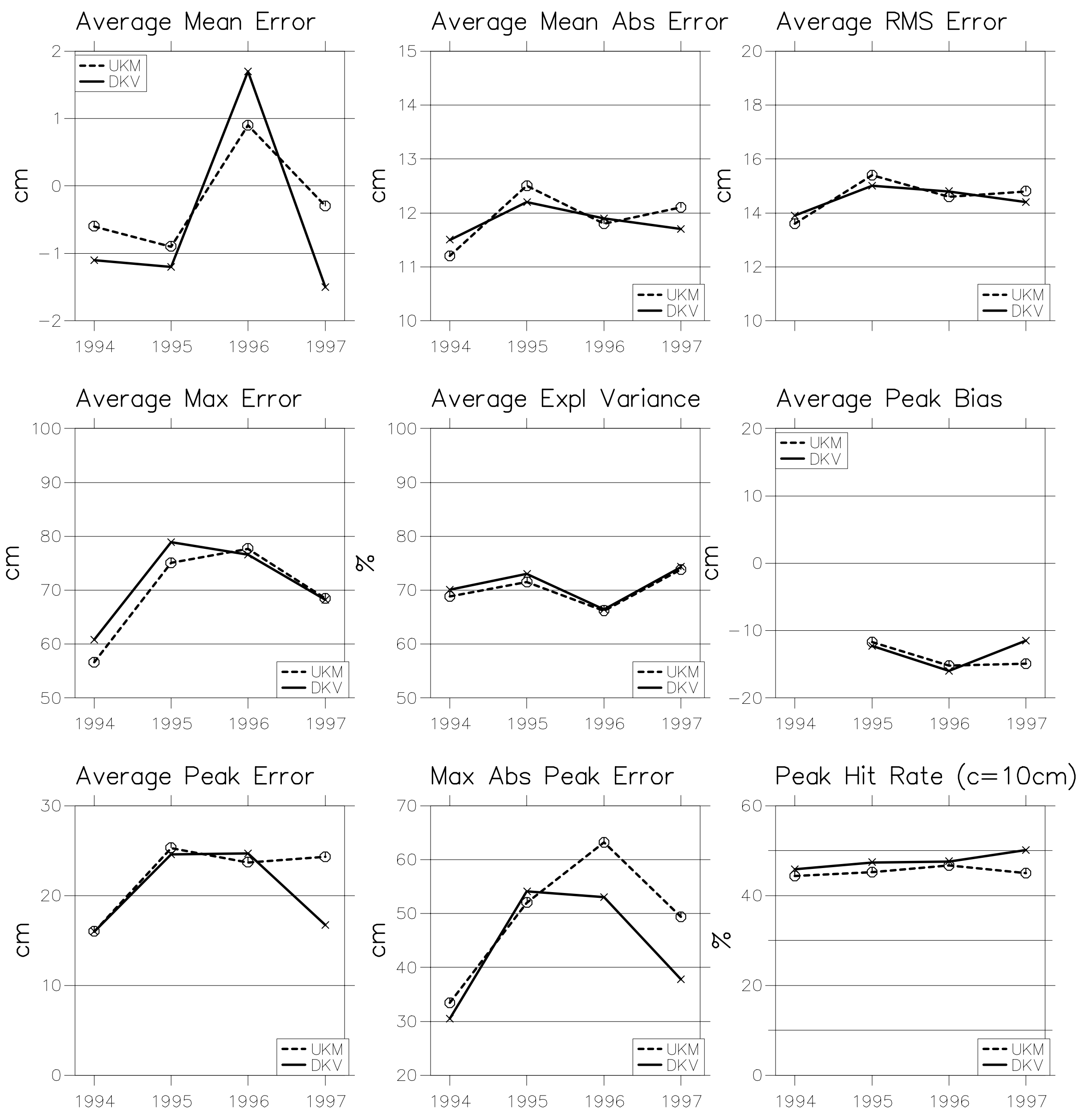


Figure 23: Station-averaged error estimates, 1994-97.

9 Comparison with test set-ups

The upgrade set-up *DKSS96*:

In DKSS96, all stations (Table 2) are archived, making a direct comparison of average error measures possible. This is shown in Table 8, last column. All residual and peak error measures are noticeable worse than both the DKV and the UKM set-up. Comparing with the DKV set-up, we find that

- the average mean error is about -5 cm
- average residual error estimates (AME,MAE,RMS) are 2-3 cm worse
- average MAX error is 5 cm worse
- average peak error estimates are (AMPE,MAPE,MXPE) 15-20 cm worse
- peak hit rates are 10 % lower

The MAE and the EV are examined station by station on Fig 24. We find that the upgrade MAE (upper panel) is worse than operational MAE (Fig. 8) at a majority of the stations (11), with the largest error increase at the Wadden Sea stations and in the Western Baltic. The upgrade MAE is less along the northern part of the Jutland West Coast, and in the Skagerrak. The upgrade EV (lower panel) is worse at 14 stations, and very much worse at some Belt Sea stations, most notably København. At Esbjerg, the EV is reduced by 5.5 %, compared with operational results (Fig. 11). Considerable improvement over the operational set-up is found only at Torsminde and Thyborøn.

The MAPE and surge errors at Esbjerg are shown on Fig 25. Compared with the operational set-up (Fig. 13, the MAPE improves only at Hirtshals, and is very much worse at many stations. In the Wadden Sea, average peak errors are of the order of 50 cm, compared with 15-20 cm for the DKV set-up. Other stations are also worse, especially Købanehavn, where average peak errors exceed 50 cm. The top ten surges at Esbjerg (lower panel) are all underestimated by the upgrade set-up, with a maximum error of -88 cm. The operational set-up (Fig. 16) may either under- or overestimate very high surges.

The DKSS96 set-up performs worse than the present operational set-up in all respects, but mostly in terms of peak errors. An improvement may be noted in some respect in the Skagerrak region, and at the northernmost part of the Jutland West Coast. Very high sea level in the wadden Sea are badly predicted, as is sea level at København in general.

The model needs further calibration.

The assimilation set-up *DMIsurge*:

DMIsurge only covers the North Sea - Kattegat region, and sea level is archived at the 8 stations marked by a d) i Table 2, making a direct comparison with the operational station averaged error measures invalid. Restricting ourselves to the 8 stations archived (Skagerrak and West Coast stations, plus Esbjerg and Havneby) we may conclude the following:

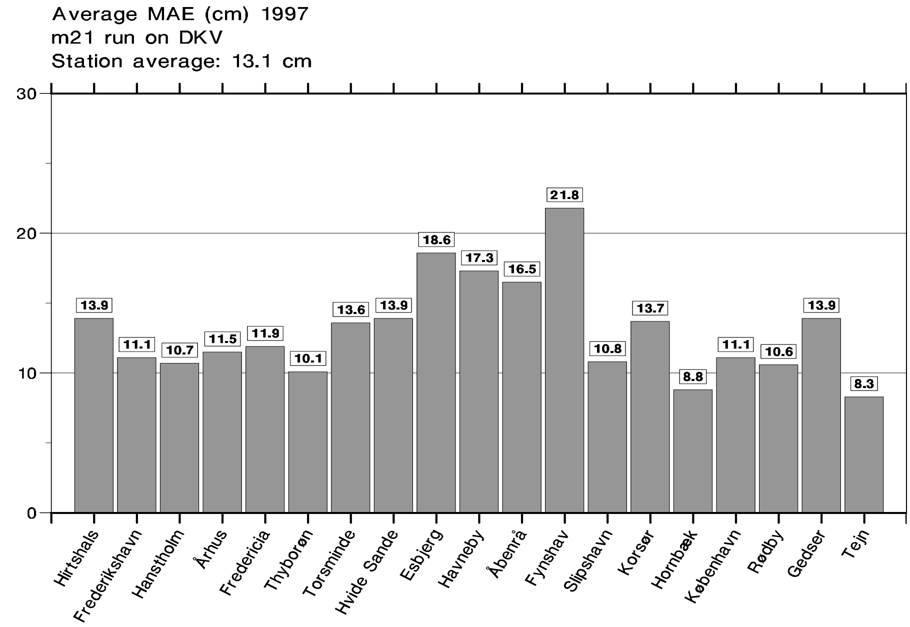
The data assimilation scheme has used observed sea level at 14 tide gauge stations until analysis+03 hours. The model is expected to perform best on short ranges, in contrast to un-initialised models where no dependence on forecast range is found. However, for sake of comparison we choose the same forecast time window as for the un-initialised models.

Selected results are shown in Table 9. The MAE is everywhere higher than for the operational set-up (fig. 8). Results are just a few cm worse at the West Coast/Skagerrak stations, but the MAE reaches 40 cm in the Wadden Sea. The E.V. is everywhere lower than for the operational set-up (Fig. 11). At the Skagerrak station and at Hvide Sande, the E.V. falls below 50 %. Only at Torsminde is the E.V. comparable with the opearational result. The MAPE is everywhere higher than for the operational set-up (fig. 13). At Torsminde and Thyborøn, the highest surges are grossly underestimated. Better results are found at Hvide Sande, Esbjerg and Havneby, while the errors at the Skagerrak stations should be compared with the less high surges in this region.

Station	MAE	E.V.	MAPE
Hirtshals	15.1 cm	41.1 %	27.9 cm
Frederikshavn	14.6 cm	25.1 %	22.5 cm
Hanstholm	15.6 cm	42.2 %	44.5 cm
Thyborøn	15.7 cm	66.6 %	69.6 cm
Torsminde	18.5 cm	74.8 %	61.7 cm
Hvide Sande	22.9 cm	45.7 %	25.1 cm
Esbjerg	36.6 cm	53.7 %	26.4 cm
Havneby	40.0 cm	55.9 %	23.1 cm

Table 9: Data-assimilation results, 8 stations, 1997

The data-assimilation set-up is certainly no replacement for the operational set-up on forecast ranges beyond 6 hours.



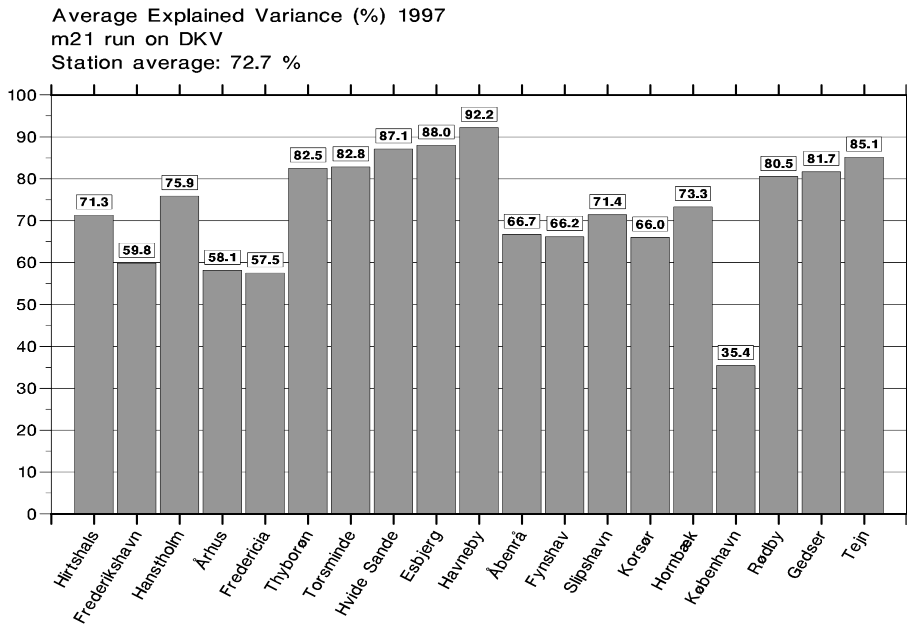
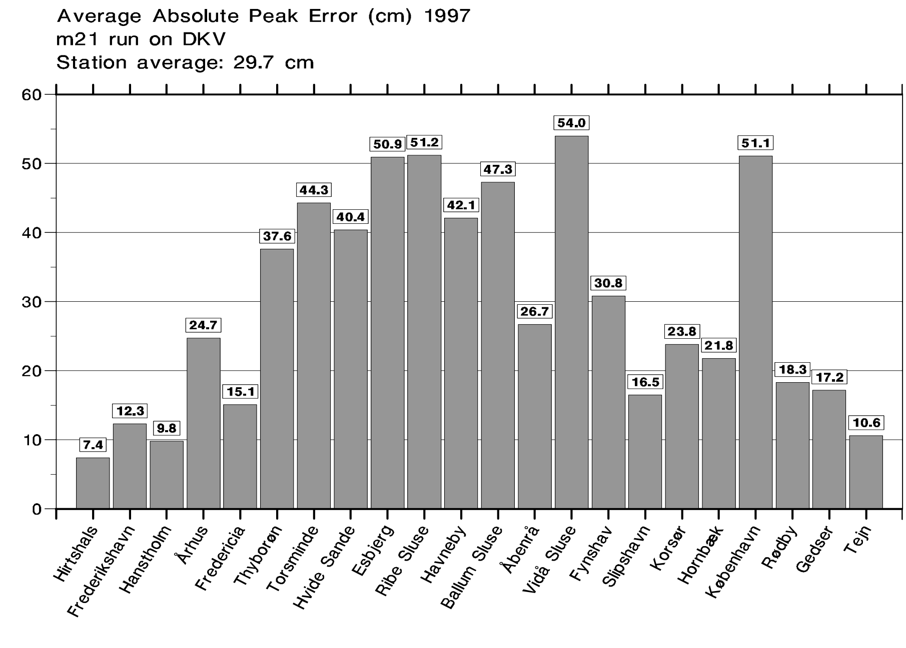


Figure 24: Mean absolute error (upper) and explained variance (lower), by the upgrade set-up (DKSS96/Mike 21). 22 stations, 1997.



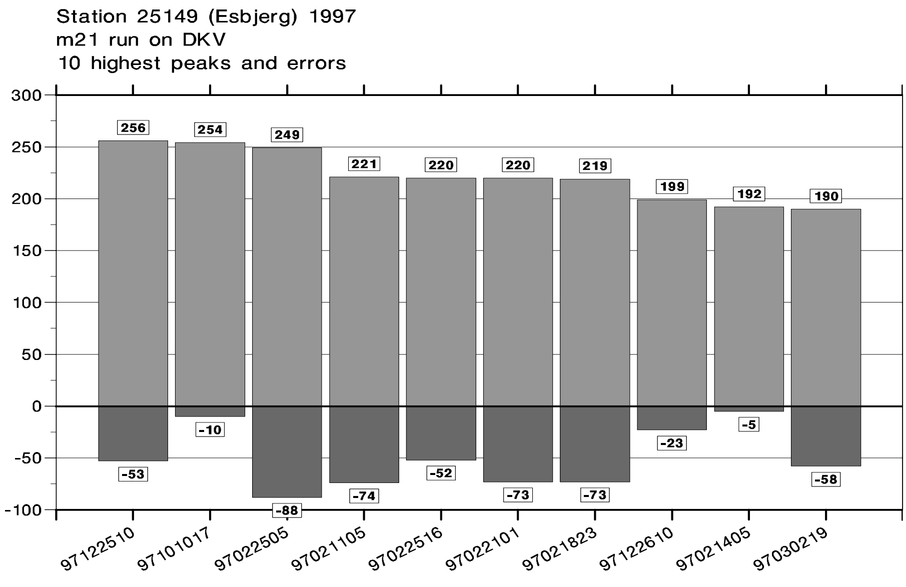


Figure 25: Absolute peak error (upper) and Esbjerg surge errors (lower), by the upgrade set-up (DKSS96/Mike 21), 1997.

10 Conclusion

DMI-HIRLAM (in short: DKV) based and UK-LAM based sea level forecasts from 1997 have been verified, using observations from 22 Danish tide gauge stations. Three of these stations (Vidå, Ballum and Ribe) are only used for peak statistics. Four stations (Hvide Sande, Ribe, Åbenrå and København) have low data coverage but were used anyway. Only the time window analysis +06 hours to analysis +18 hours is considered. Comparison is made with unbiased results, with results from previous years and with results from two test set-ups.

In terms of residual error measures, the DKV and the UKM set-ups are of very similar quality, both on average and at individual stations.

In terms of peak-to-peak error measures, the DKV set-up is noticeably superior to the UKM set-up. The peak hit rate is 4% higher, and average peak errors are 10cm lower with the DKV set-up.

Unbiasing the DKV forecasts using 1994-96 average mean errors leads to some improvement of both residual and peak-to-peak error estimates.

DKV peak error measures have improved considerably since 1996 and this must be ascribed to an improved performance of the DKV-HIRLAM itself. Compared with previous years, average peak error on the 10 highest peaks has decreased by 8cm to 18cm, and the peak hit rate has increased by 2% to 50%.

Neither of the two test set-ups are quite up to the mark. Both need further calibrating before forecasts from these models may replace the current operational forecasts. The Upgrade set-up has considerably lower quality than the DKV set-up in the Wadden Sea and at some Belt Sea/Sound stations, while forecasts at the Skagerrak and northern Jutland West Coast stations are of operational quality. Th e Data-assimilation set-up is very much worse than the DKV set-up at all 8 stations at this forecast range. That set-up is designed to improve forecasts at ranges below 7 hours, but since the model mirrors observations approximately until analysis +03 hours, it was not expected to drift so fast.

References

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

[2] A.W. Heemink. Two-dimensional shallow-water flow identification. *Appl. Math. Modelling*, 12:109–118, april 1988.

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

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

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

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

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

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

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

[10] 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.

[11] Bent Hansen Sass. The DMI Operational HIRLAM Forecasting System. version 2.3 - a short summary. Technical Report 94-8, Danish Meteorological Institute, 1994.

[12] Anita Sielecki. An energy-conserving difference scheme for the storm surge equations. *Monthly Weather Review*, 96(3):150–156, 1968.

[13] 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 Figures