Application of EOF to sea level forecasting

Håkan Törnevik

Swedish Meteorological and Hydrological Institute

Note:

The research work reported here has been carried out at the Research and Training Department of the Swedish Meteorological and Hydrological Institute by Ingemar Holmström, John Stokes, Arne Törnvall and Håkan Törnevik.

A detailed report on the first part of the work is being published as a SMHI Report, Holmström and Stokes (1977). Reports on further work will be published during 1978.

INTRODUCTION

For shipping along the Swedish coast of the Baltic, information on sea level and its variation has become more and more important.

A major reason is here that merchant ships have increased in size, which means that access to certain ports is only possible at normal or high sea level. This is the matter especially with the ports in the Northern Sweden where sea level variation of 1 meter in one or two days is not uncommon.

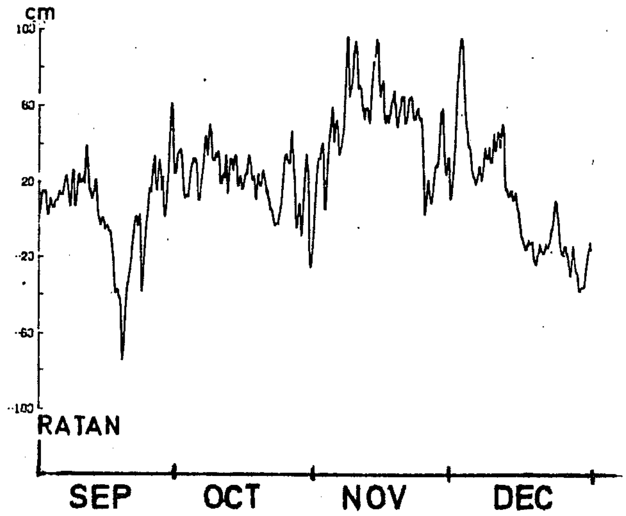


Figure 1 Sealevel-variations in Ratan 1969, September - December

Baltic tidal variations are of very little interest. They are only of the order of a few centimeters.

For planning the shipping and for operational purpose, forecast of sea level will obviously be of great economic value.

The dominating factors influencing sea level are surface pressure and surface wind stress and since these can be predicted 2-3 days ahead, it should be possible to make predictions of sea level changes.

A comparison between a numerical finite difference model and a statistical model both forced by surface stress and pressure for the Lake Vänern showed some interesting results.

Both of them gave about the same accuracy in predicted values of sealevel but the statistical model was considerably more economic and it was therefore decided to test a statistical model on the Baltic.

Available data

In the first experiment data were available from 6 Swedish coastal stations (see figure 2). A 3.5 year period, 1 June 1967 - 31 December 1970 was chosen as a basic data period. There were 6 observations a day (every 4th hour) which means 7860 observations from each station.

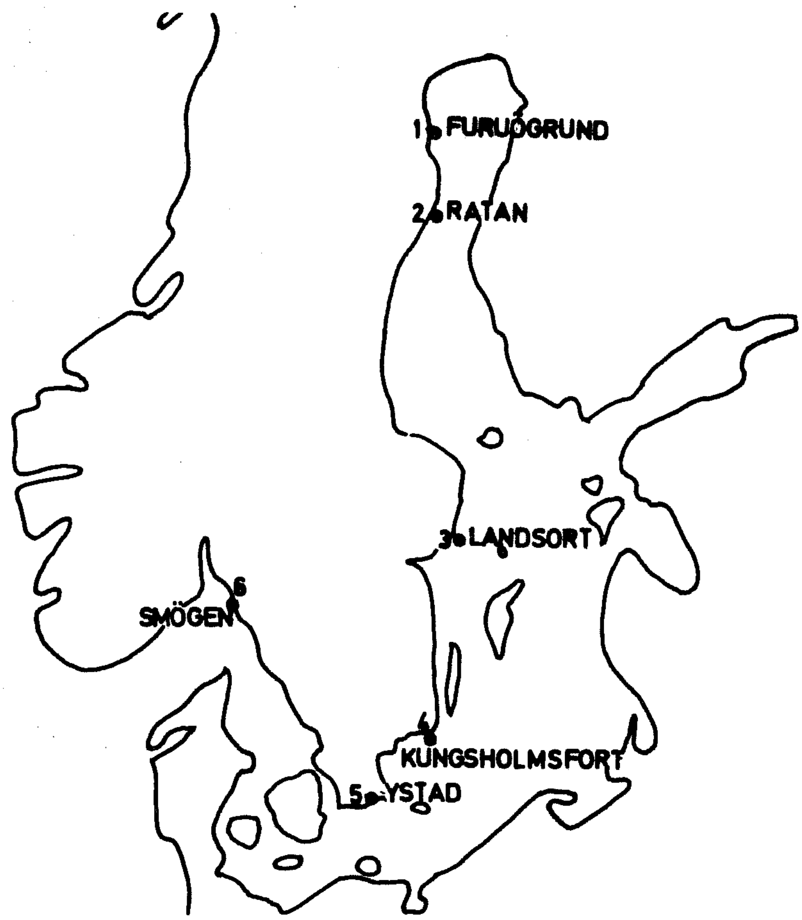


Figure 2 Distribution of stations used in the first experiment

Statistical approach

At each station an arithmethic mean (for the basic data period) was subtracted and the series were expanded into eof.

The expansion had the form

$$s_i(t) = \sum_{n=1}^{6} \beta_n(t) \cdot h_{ni} \qquad (1)$$

i = station index

s = waterlevel

$\beta_n(t)$ = a time dependent amplitude function common to all stations

h_{ni} = a local responsefunction for the station i (independent of time)

(1) is doubly orthogonal:

$$\int_0^T \beta_n(t)\beta_m(t)\ dt = \delta_{nm} \int_0^T \beta_n^2(t)\ dt$$

$$\sum_{i=1}^{6} h_{ni} \cdot h_{mi} = 6\delta_{nm}$$

where we also have applied a normalizing condition on the response function h_{ni}.

Figure 3 shows the normalized responsefunction h_{ni} for the different modes where the index number represents the stations from north to south and h_6 is Smögen on the west coast. The numbers to the right give the relative variance in each mode and below the numbers the accumulated relative variance.

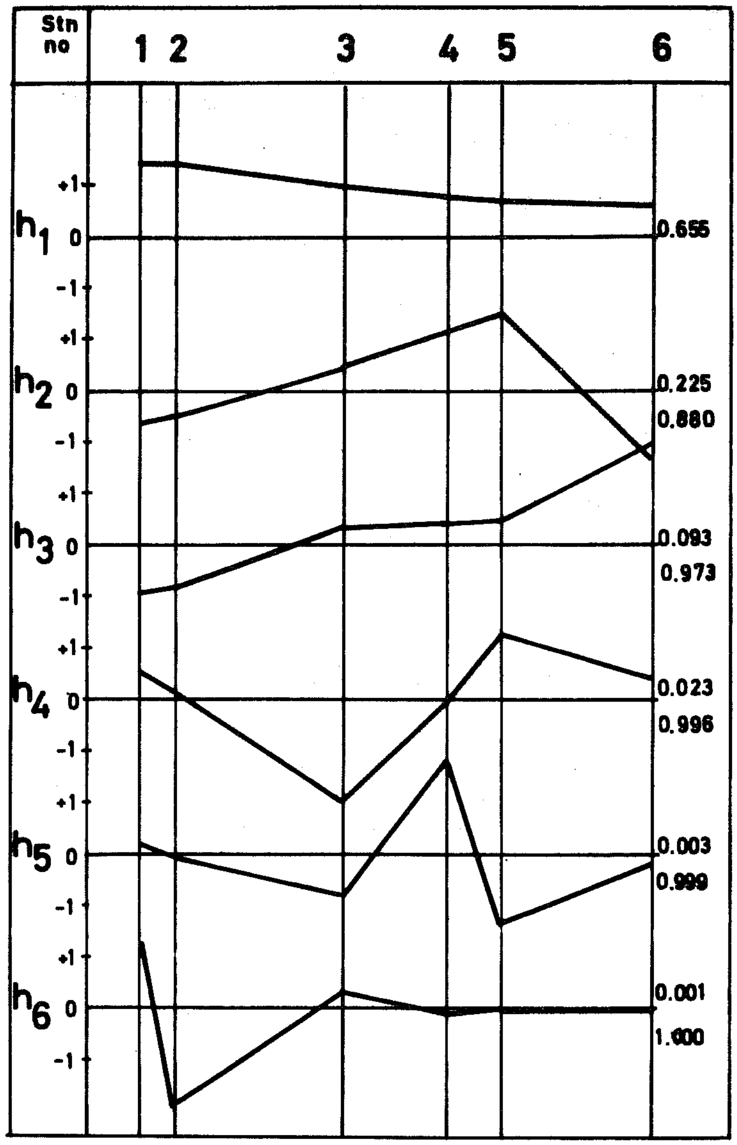


Figure 3 Local responsefunctions. The numbers correspond to the stations in figure 2.

The first function covers no less than 65.5 per cent of the total variance.

It consists of a simultaneous rising or lowering most pronounced in the northern part of the Baltic and less pronounced on the west coast.

A covariance as large as this must naturally depend on a forcing characterized by a very large scale, a fact that will have to be taken into account when atmospheric predictors in the regression scheme are to be selected.

The next two functions describe a north-south tilting of the Baltic, h_{2i} with the west coast in phase and in h_{3i} out of phase.

The last three modes correspond to very small variations, together just 2.7 per cent of the relative variance, mainly small scale variations in the centre and southern part of the Baltic for the fourth and the fifth mode and in the Sea and the Bay of Bothnia for the sixth.

In figure 4 we see the amplitude functions for June and July during the period 1967 - 1970.

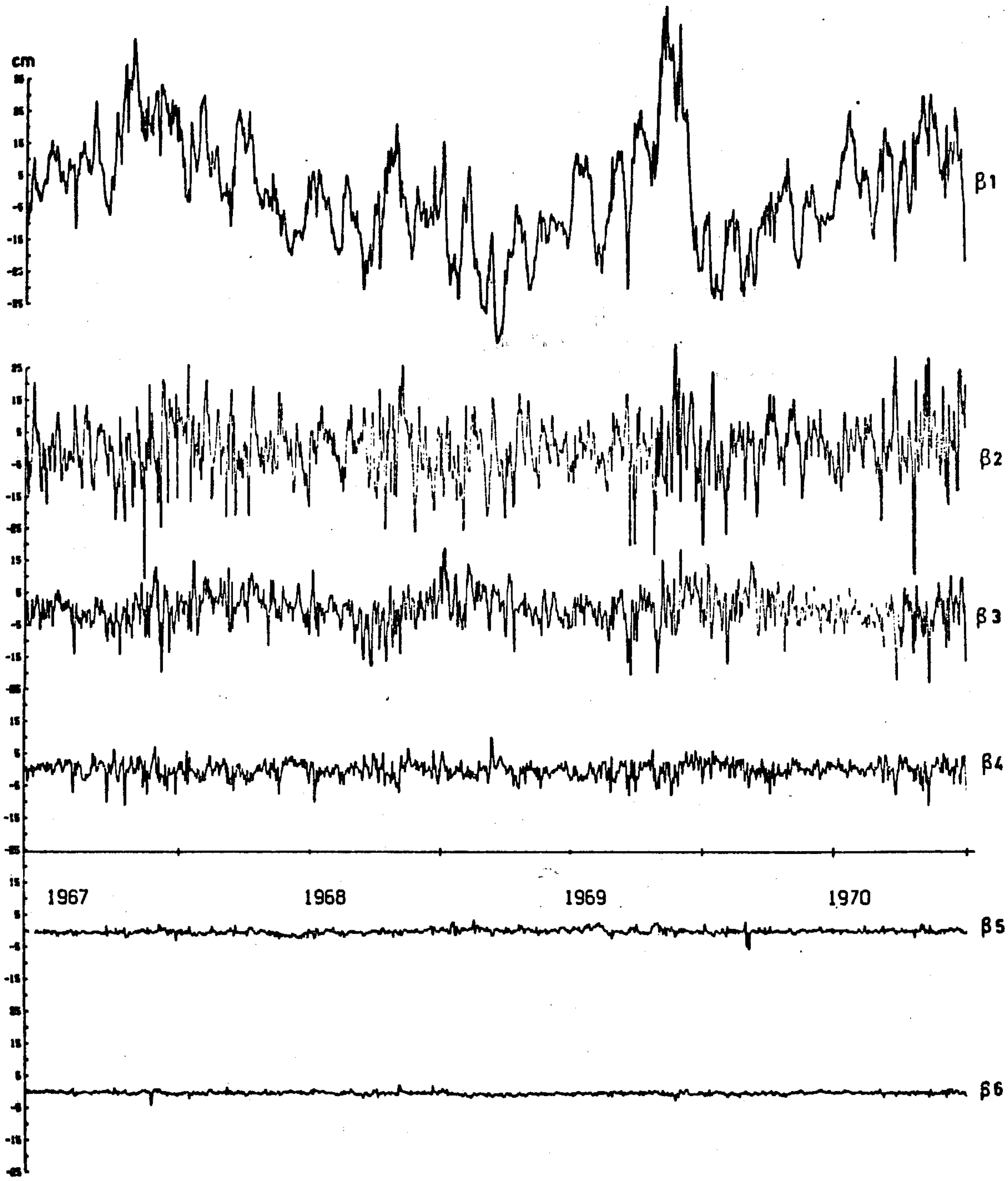


Figure 4 Amplitudefunctions for the basic data period.

Since the response functions are normalized, the magnitude of the variations of the amplitudes is representative for the contribution of each term in the series expansion.

The amplitudes decrease and the frequencies increase with higher modes. This indicates that small and rapid changes in the weather are becoming more important as forcing for higher modes.

If we look at the long term variation we see that even for as short periods as 1 to 2 months the mean value of amplitudes of order 2 or higher is almost zero.

But for the first mode we can, except for 1968, see a yearly trend. For determination of a typical yearly trend we need a longer period.

In order to investigate characteristic features of the amplitude functions $\beta_n(t)$ a Fourier analysis has been made.

The spectral distributions are shown in figure 5 and 6 where we can see that the main part of the energy lies in frequencies, corresponding to typical meteorological periods.

External forcing is obviously of major importance for the system.

The peak at 12 hours most pronounced in β_3 and β_4 is due to the tidal variations.

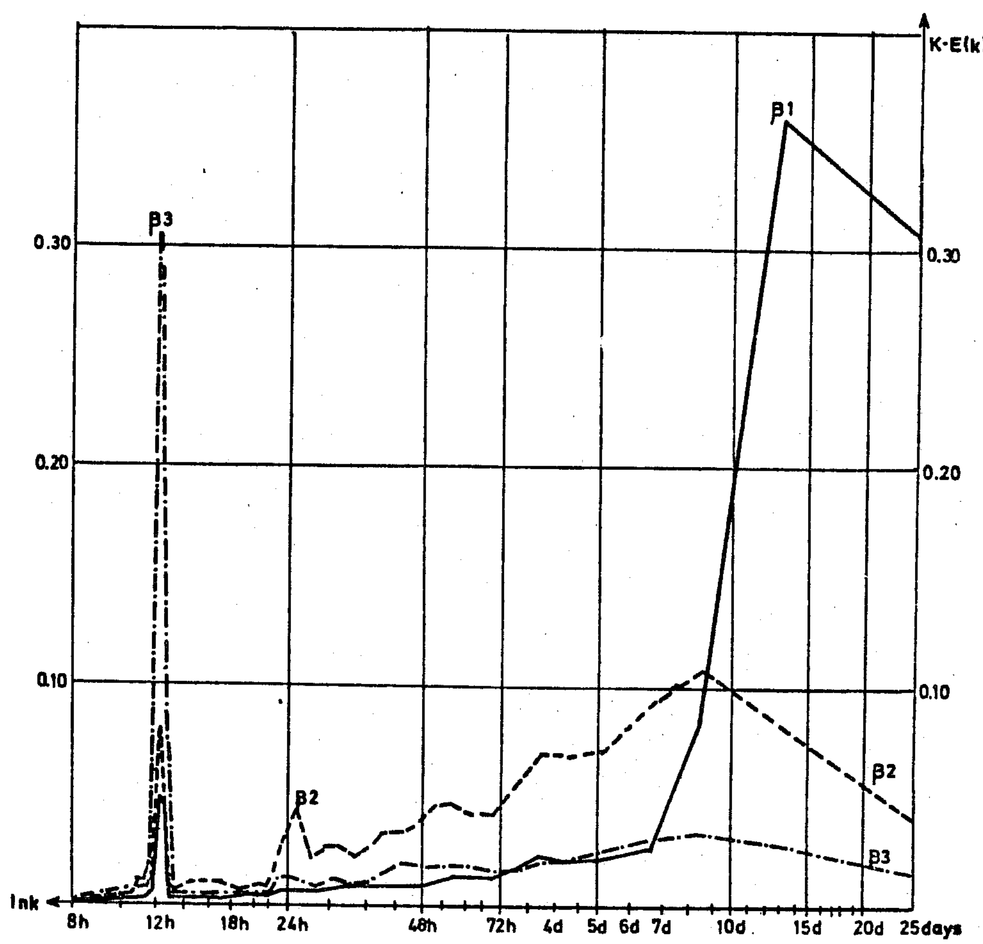


Figure 5 Spectral distribution of the first three modes.

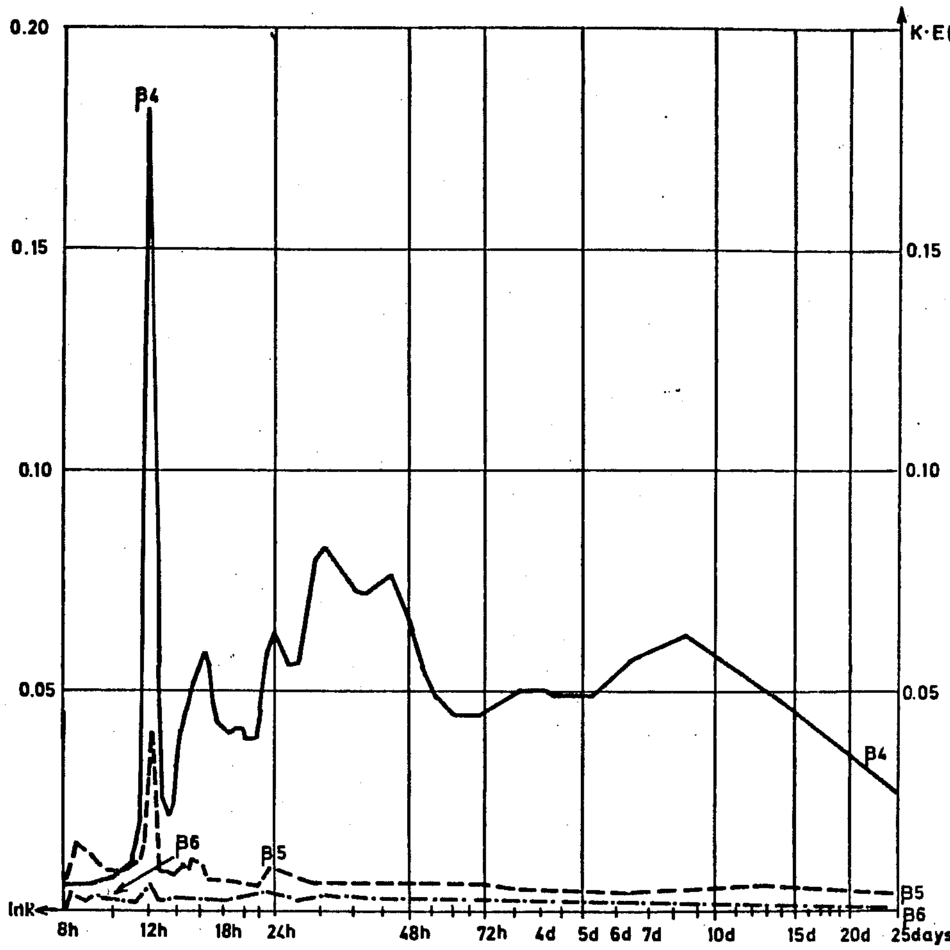
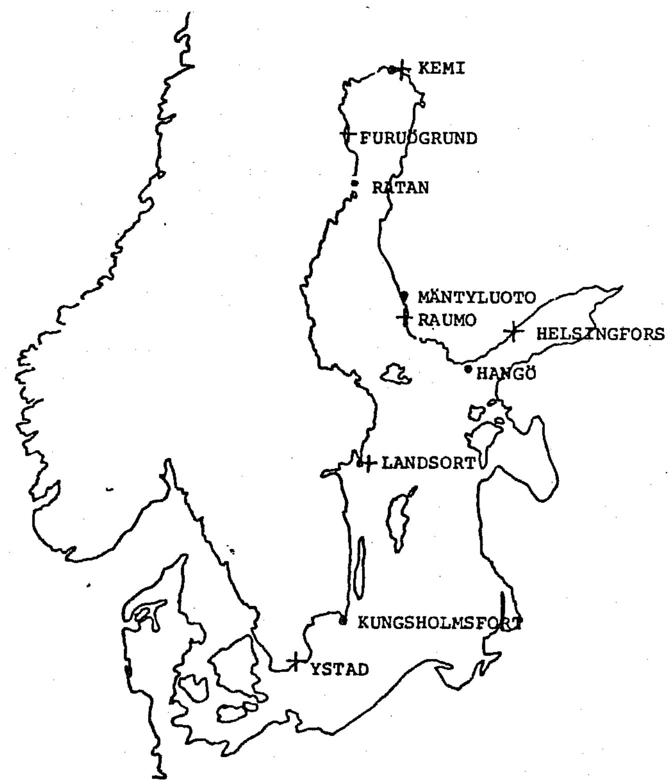


Figure 6 Spectral distribution of the last three modes.

FURTHER EXPERIMENTS

As we now have seen, the general behaviour of the Baltic seems to be rather simple. But can we draw such a conclusion when the expansion is based on just 6 Swedish coastal stations and the basic data period is only 3.5 years. Would we get similar results by using observed data from stations all around the Baltic? Another questions is; how long a period of basic data is required in order to obtain local response functions that do not depend on the length of the period.

To get answers to these questions two new expansions into eof were made including also Finnish data. The stations are given in figure 7. Dots indicate stations in the first expansion and crosses in the second. The different modes in these expansions showed the same characteristics as in earlier results.



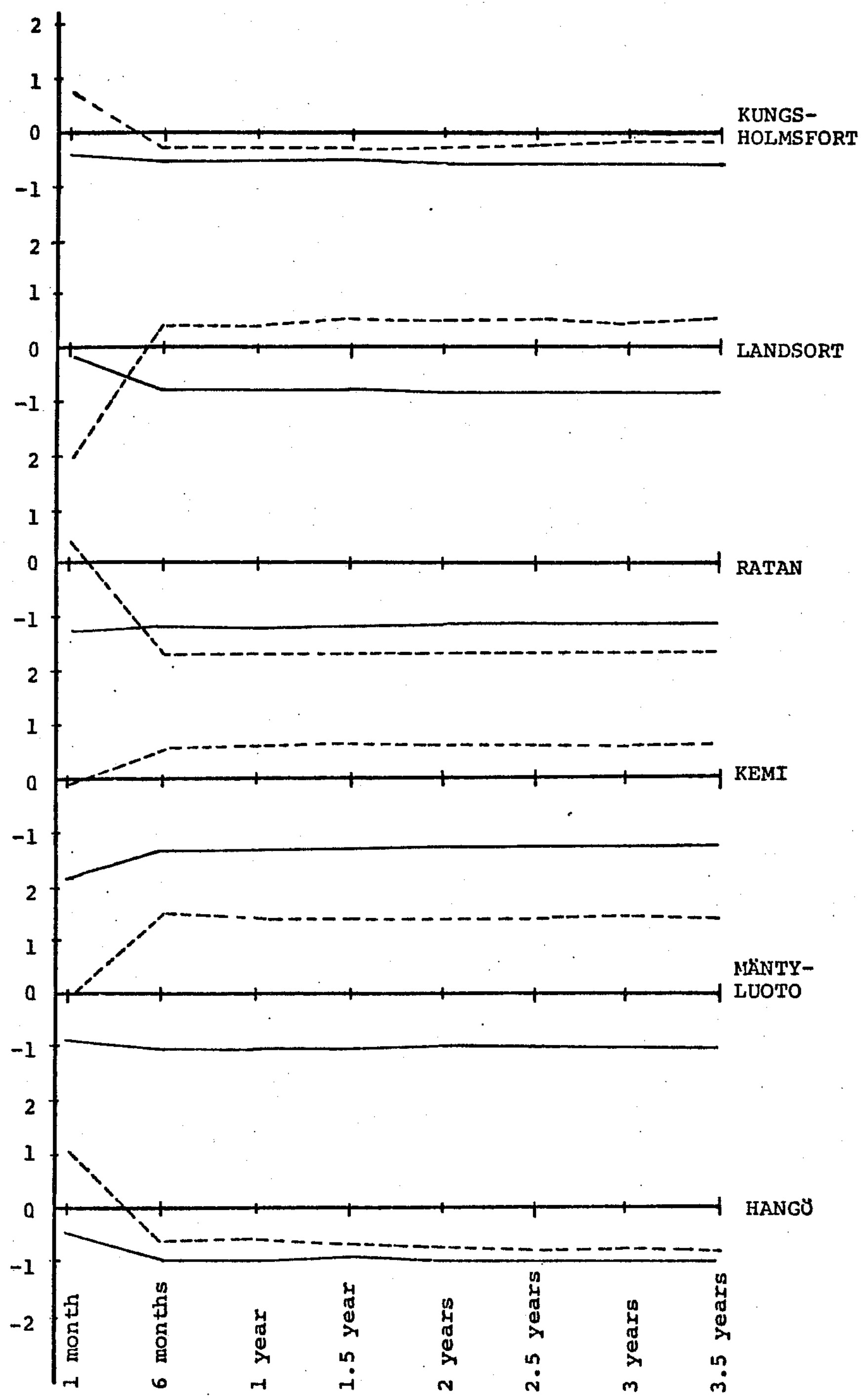
β_1: 1.00	β_4: 0.57
β_2: 0.97	β_5: 0.61
β_3: 0.94	β_6: 0.65

Figure 7 Correlation between β:s determined from two expansions:
1. YSTAD, LANDSORT, FURUÖGRUND, KEMI, RAUMO, HELSINGFORS
2. KUNGSHOLMSFORT, LANDSORT, RATAN, KEMI, MÄNTYLUOTO, HANGÖ

In the figure is also given the correlation between the corresponding β-functions in the two expansions. We see that the correlation is very high for the three first modes. Together they cover not less than 98.5 per cent of the total variance.

But what about the basic data period? Is a 3.5 year period sufficient for determining the time independent response function? To get an answer to this question, data were expanded from 3 Swedish and 3 Finnish stations.

We started with a period of 1 month for determining the h-functions, then added 6 months and redetermined the response-functions and then added further 6 months and so until finally we had a 3.5 year period. We see in figure 8 that after 6 months the values are very stable. This is true for the first mode, but even for the 6th mode half a year is sufficient for determining the response-function.



. Figure 8 Stability in response functions

——— 1:st mode (h_{1i})

- - - - 6:th mode (h_{6i})

Looking again at a set of response-functions based on an expansion of 6 Baltic stations (2 Finnish and 4 Swedish) during a 3.5 year period with data measured every 4th hour (see figure 9).

The first function covers here 76.4 per cent of the relative variance and as mentioned earlier this mode represents a general behaviour of the Baltic influenced by a very large scale forcing. In fact this must represent a general rising or lowering of the surface of the Baltic mainly due to a total inflow or outflow in the Baltic.

The other modes have response functions with an average very close to zero and they therefore represent internal effects. One would then expect that a similar expansion based on sea level data averaged over longer periods would give low order functions a larger variance.

Comparing figures 9-11 we see that this is true. But what is remarkable is: the response functions do not change in appearance.

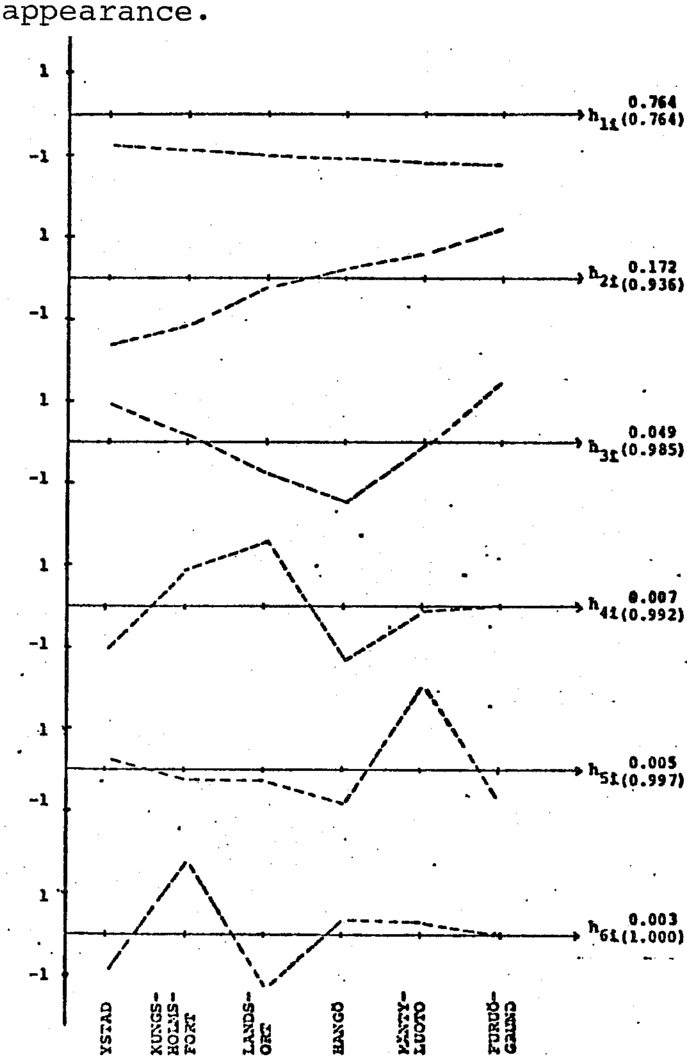


Figure 9 Response functions. The expansion based on 7860 observations from each station 1967, June - 1970 December (every 4th hour). The numbers in the right of the figure are the relative contributions in % from the different modes to the total variance. Those in parenthesis are accumulated values.

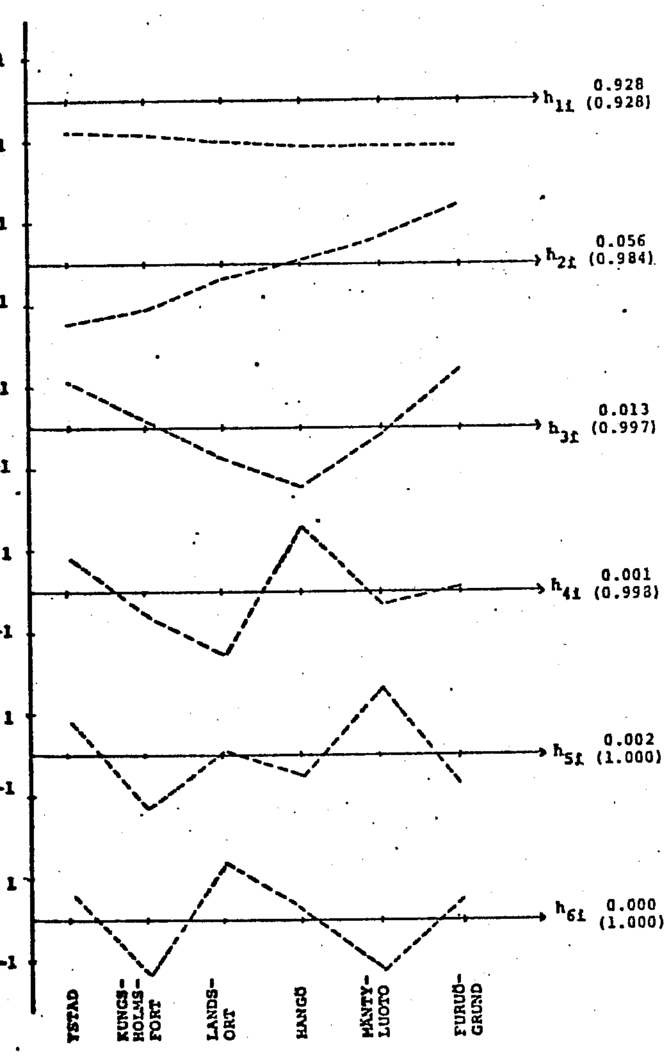


Figure 10
The same as in figure 9 but the expansion is based on observations daily averaged.

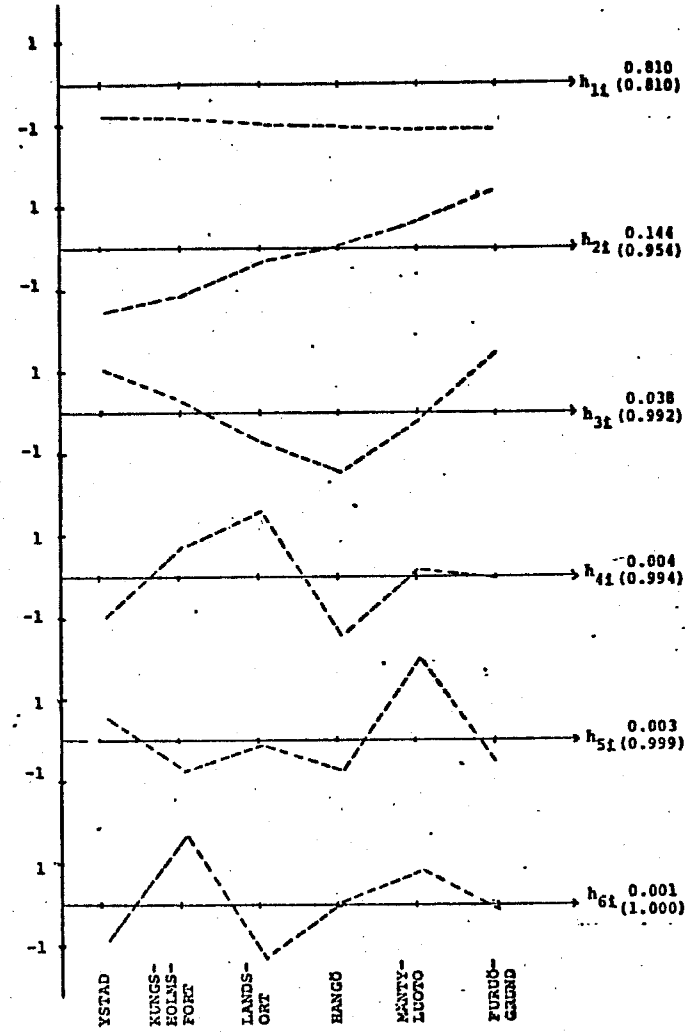


Figure 11
The same as in figure 10 but the expansion is based on observations 10 days averaged.

Volumes

For a 20 year period 1950 - 1970 Lazarenko [1977] has computed monthly mean values of the total water volume of the Baltic. Comparing his results with the β_1-function we see from figure 12 that the results are almost identical. The correlation between his and our results is 0.99.

Lazarenko's results are based on data from almost 80 Baltic coastal stations, while the β_1-series is determined from 4 Swedish and 2 Finnish coastal stations.

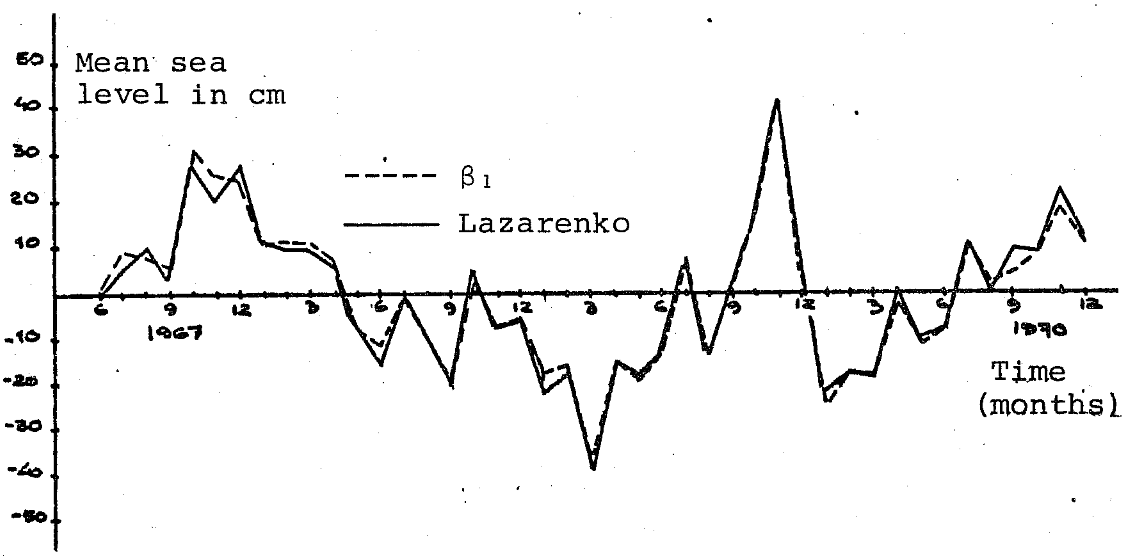


Figure 12

Redundant information?

As we have seen (figure 7) the modes, at least the first three modes seem to be identical from one expansion to another. These modes do really represent a general structure of the Baltic. In other words: knowing the amplitudes (determined from a set of observations) we are able to describe the waterlevel situation fairly well at an arbitrary point on the Baltic coast. There are two methods available for determining the local responsefunctions for an arbitrary point.

1. A simple linear interpolation between the local responsefunction from two surrounding stations. This method would not guarantee a good result because the variation in h along the coast does not need to follow a straight line.

2. The other method takes advantage of the orthogonal properties of the β:s. If we have a period of observations from the "arbitrary point"(a period of 6 months would probably be enough), the data could be expanded into eof i.e.

   $$S_{AP}(t) = \sum_{n=1}^{6} \beta_n(t) \cdot h_{n,AP}$$

   (AP denotes "arbitrary point")

   where the β:s already have been determined from a set of stations.

   Multiply $S_{AP}(t)$ by β_k and integrate over the period

   $$\int_0^T S_{AP}(t)\, \beta_k(t)\,dt = \int_0^T \beta_k(t) \sum_{n=1}^{6} \beta_n(t) \cdot h_{n,AP}\,dt$$

   But for this period β_k and β_n, $k \neq n$ are approximately orthogonal so we get

   $$\int_0^T S_{AP}\, \beta_k(t)\,dt = h_{k,AP} \cdot \int_0^T \beta_k^2(t)\,dt$$

   from which we can determine $h_{k,AP}$

By using the β:s from the very first expansion (based on stations in fig 2) responsefunctions for additional 11 stations were determined. In figure 13 we can see to what extent the total variance can be covered in the different stations by using response functions determined with method 2.

As we can sẹe the method is very successful except for the stations in the Gulf of Finland (Hamina, Helsingfors) and those in and near the narrow straits (Klagshamn, Björn, Degerby).

How many observation stations do we really need in the Baltic to describe the behaviour of the sea sufficiently well? Various estimates have been made all insisting on a need for more observation stations to be built.

Since mareograph-stations are expensive it is of interest to note that this investigation does not confirm such demands. On the contrary it seems that part of the network could be excluded after say one or two years of measurements.

stn/β	β_1	$+\beta_2$	$+\beta_3$	$+\beta_4$	$+\beta_5$	$+\beta_6$
HAMINA 67-06--64-03	59.3	59.3	59.8	69.1	71.1	71.1
HELSINGFORS 67-06--70-13	69.3	69.4	70.0	81.2	82.0	82.0
HANGÖ 67-06--70-12	76.8	77.2	78.1	90.7	91.3	91.4
RAUMA 67-06--70-12	89.8	90.0	90.2	94.7	95.0	95.3
MÄNTYLUOTO 67-06--70-12	91.2	91.7	92.1	95.1	95.3	95.6
VAASA 68-09--70-12	87.6	88.6	90.4	90.9	91.0	90.9
KEMI 67-06--70-12	75.7	84.5	93.0	94.9	94.9	95.8
STOCKHOLM 68-01--69-12	79.5	84.8	84.5	97.3	97.7	97.5
KLAGSHAMN 68-01--69-12	28.8	79.0	79.0	81.5	81.6	81.7
BJÖRN 67-06--68-10	69.0	69.2	69.1	76.2	76.7	79.2
DEGEBY 67-06--68-11	71.5	70.7	70.4	83.7	83.8	84.4

<u>Figure 13</u> The relative contribution (accumulated in %) from the different modes to the total variance. The β:s determined from the stations in figure 2.

Predictors

The surface pressure data required for prediction were taken from grid point values in a 7 x 7 grid in the routine numerical analysis and forecast model of the Swedish Weather Bureau (see fig 13).

For the basic data period (June 1967 - December 1970) these pressure data were expanded into eof.

The expansion had the form:

$$p_j(t) = \sum_{m=1}^{49} \alpha_m(t) \cdot g_{mj} + \overline{p_j} \qquad (2)$$

j = gridpoint index

$\overline{p_j}$ = local average over the period

The reason for subtracting a local mean of the pressure is that it should correspond to local mean values of the sea level which was separated out earlier.

The response functions for the 8 first modes are showed in fig 13 and 14.

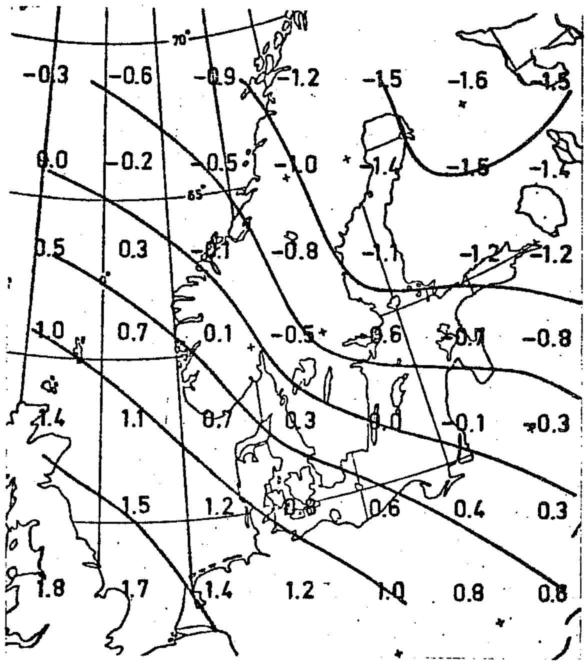
The convergence is here much slower than in the sea level example and in order to arrive at a variance reduction of 98 per cent not less than 13 terms in the expansion were needed.

An advantage of expanding the surface pressure field into EOF is seen if we apply a finite difference gradient operator to both sides of the relation

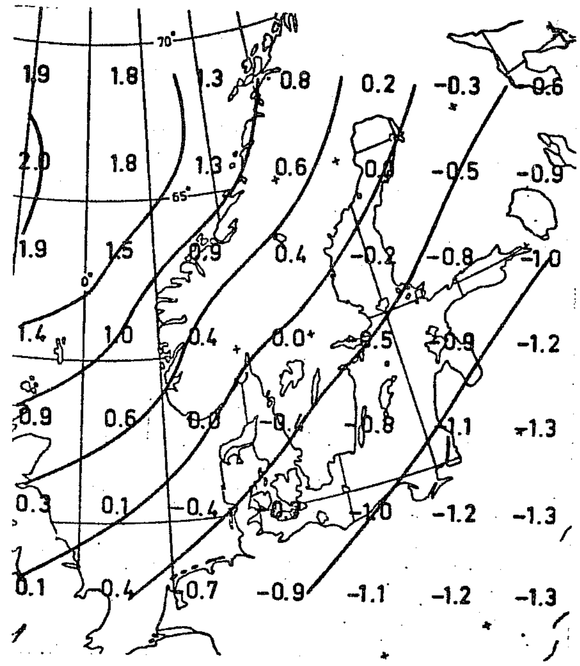
$$\nabla p_j = \nabla \left(\sum_{m=1}^{49} \alpha_m(t) \cdot g_{mj} \right)$$

$$\Rightarrow \quad \nabla p_j = \sum_{m=1}^{49} \alpha_m(t) \cdot \nabla g_{mj}$$

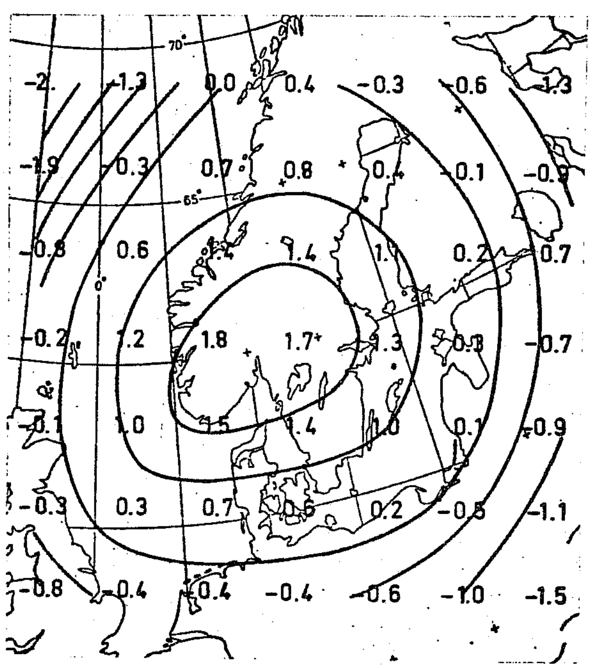
This means that the amplitudes α_m not only represent the pressure field but also the wind field (at least the geostrophic wind).



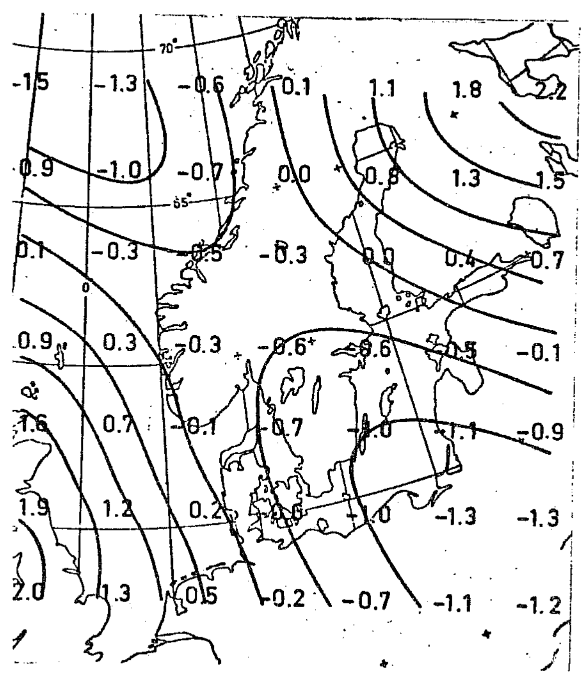
g_{1j} 38.2 (38.2)



g_{2j} 33.2 (71.4)

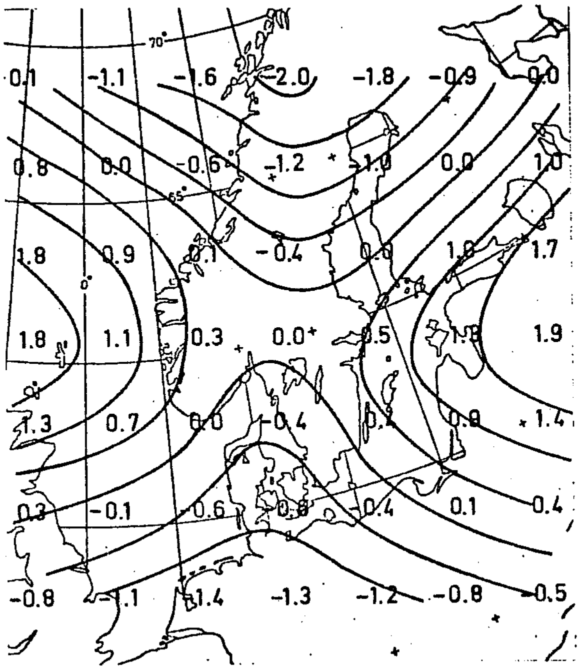


g_{3j} 8.5 (79.9)

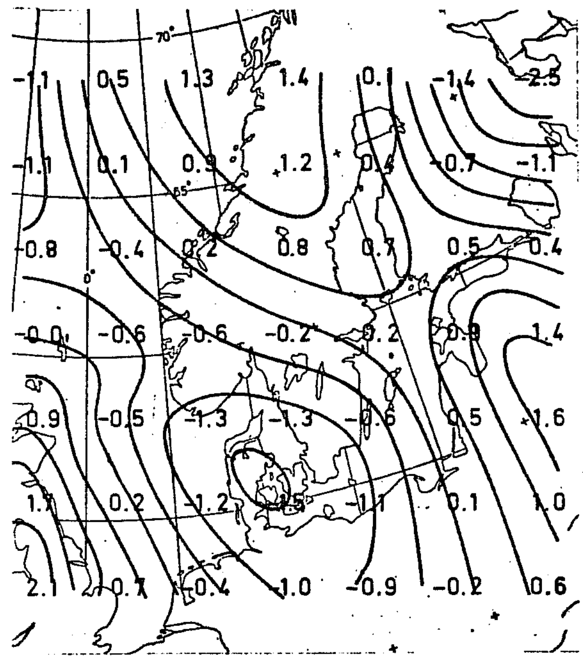


g_{4j} 7.3 (87.1)

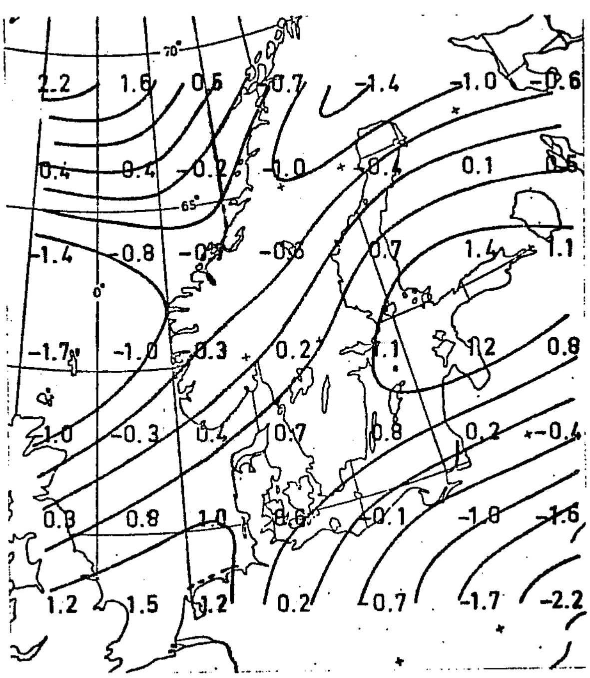
<u>Figure 13</u> The response functions g_{ij} - g_{4j}, j = 1.49. The numbers below the figures give the relative variance in each mode. (In paranthesis the accumulated relative variance).



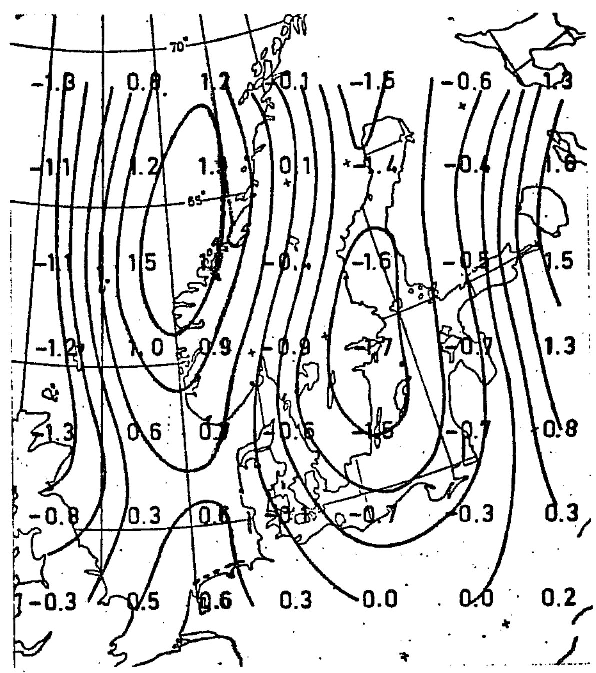
g_{5j} 4.3 (91.4)



g_{6j} 1.6 (93.0)



g_{7j} 1.4 (94.5)



g_{8j} 1.0 (95.4)

Figure 14

The response functions g_{5j} - g_{8j}.

The numbers below the figures give the relative variance in each mode. (In parenthesis accumulated relative variance)

The regression method

Since sea level variations in the Baltic may imply transports of large masses of water through narrow passages such as Öresund and the Belts and the Åland sea one would expect that the response to varying pressure and winds is not immediate.

The sea level changes are to be considered as an integrated effect of wind and pressure variations over a period of a certain length.

As we have seen the α:s represent both wind and pressure and so we use α:s over a period as predictors (see fig 15).

One may also expect that the Baltic has a certain inertness, and the system needs a certain time to go back to normal from a distribution caused by a forcing. That means we ought to use a series of β:s also as predictors.

However, if forecasts of $\beta_n(t)$ are to be made, actual values of these functions will only be known up to the time when the forecast is issued, while predicted values of $\alpha_n(t)$ will be available up to the time for which the sea level forecasts should be valid.

In fig 15 we have the predictors used in the regression scheme for the 24-hours forecast.

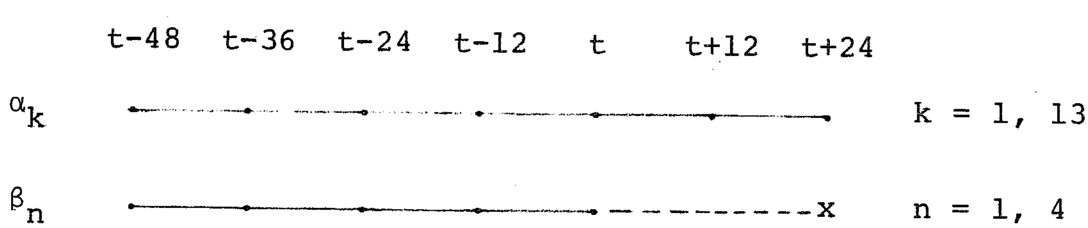


Fig 15. Predictors used for the 24-hours forecast. The α:s and the β:s are truncated at numbers 13 and 4 respectively.

$\alpha_k(t+12)$ and $\alpha_k(t+24)$ $(k=1,13)$ are taken from the numerical forecasts.

The predictands are $\beta_n(t+24)$ $(n=1,4)$ and the sea level at time t+24 in station i is described by

$$s_i(t+24) = \sum_{n=1}^{4} \beta_n(t+24)h_{ni}$$

Of course these predictors carry a considerable amount of redundant information. This will create difficulties when determining the regression coefficients.

To avoid this, it was necessary to orthogonolize the predictors. This was made in the most efficient way, i.e.by expanding the α:s and the β:s into a common set of empirical orthogonal functions.

In this series-expansion a sufficient number of terms were retained in order to cover about 99 per cent of the variance. Regression coefficients were then conveniently determined.

Results

Regression coefficients were determined for the basic data period. In this experiment the forecast method was (for several reasons) chosen to be a "perfect prog method". The forecast method was then tested for an independent data period 1971, 1 January - 1973, 31 December. The result is shown in figure 16. The first line represents the rms-error for a persistence forecast. The second line represents the rms-error for the statistical model when using predicted surface pressure from the numerical model at SMHI. Line number 3 represents the rms-error of the same model but now using observed pressure. It shows the result that would be obtained if the predicted pressure fields were correct.

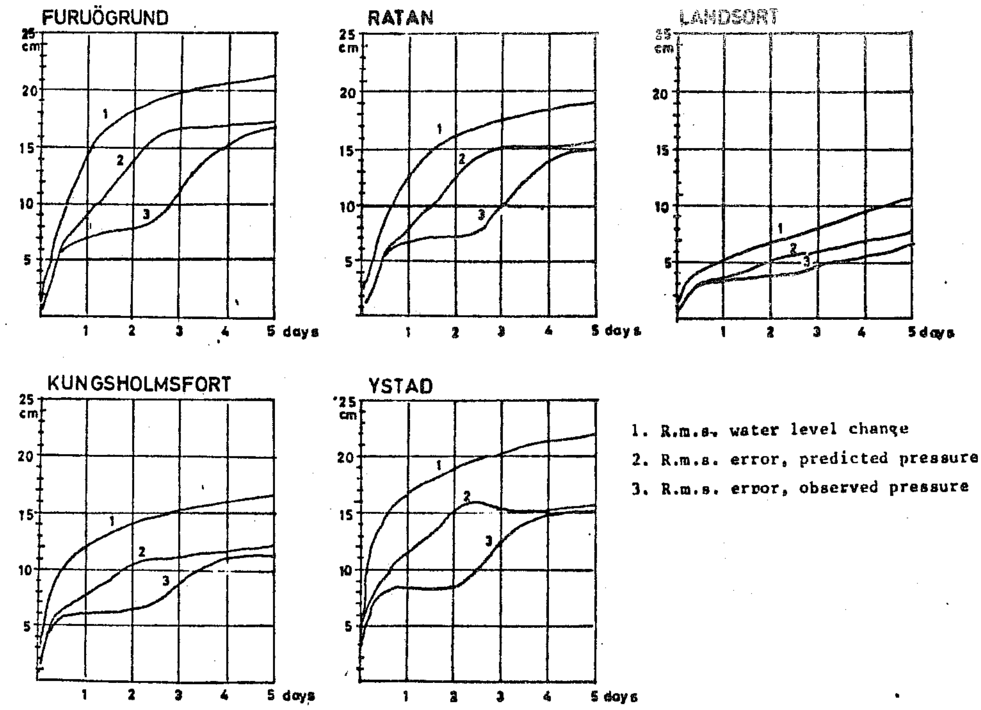


Figure 16

For a 24-hour-forecast the result is surprisingly good. The error is about 60% of the error for a persistence forecast. For a 48-hour forecast the rms-error is increasing rapidly. But if we look at line 3 we can see that this is mainly due to the increasing error in the predicted pressure fields. A MOS-technique would probably give better results in this case.

The predictions were extended up to 120 hours but since pressure data is not available for the period 48-120 hours this is a pure statistical extrapolation. The result is rather poor and there is no sense in making longer sea level forecasts than available pressure forecasts.

In February 1977 the model was tested operationally. The interpolation method 2 which was described earlier (see page 13) was used to extend the forecast to other stations than those in the model.

In figure 17 is given a comparison between observed sea level and 36-hours predicted sea level in the harbour of Gothernburg.

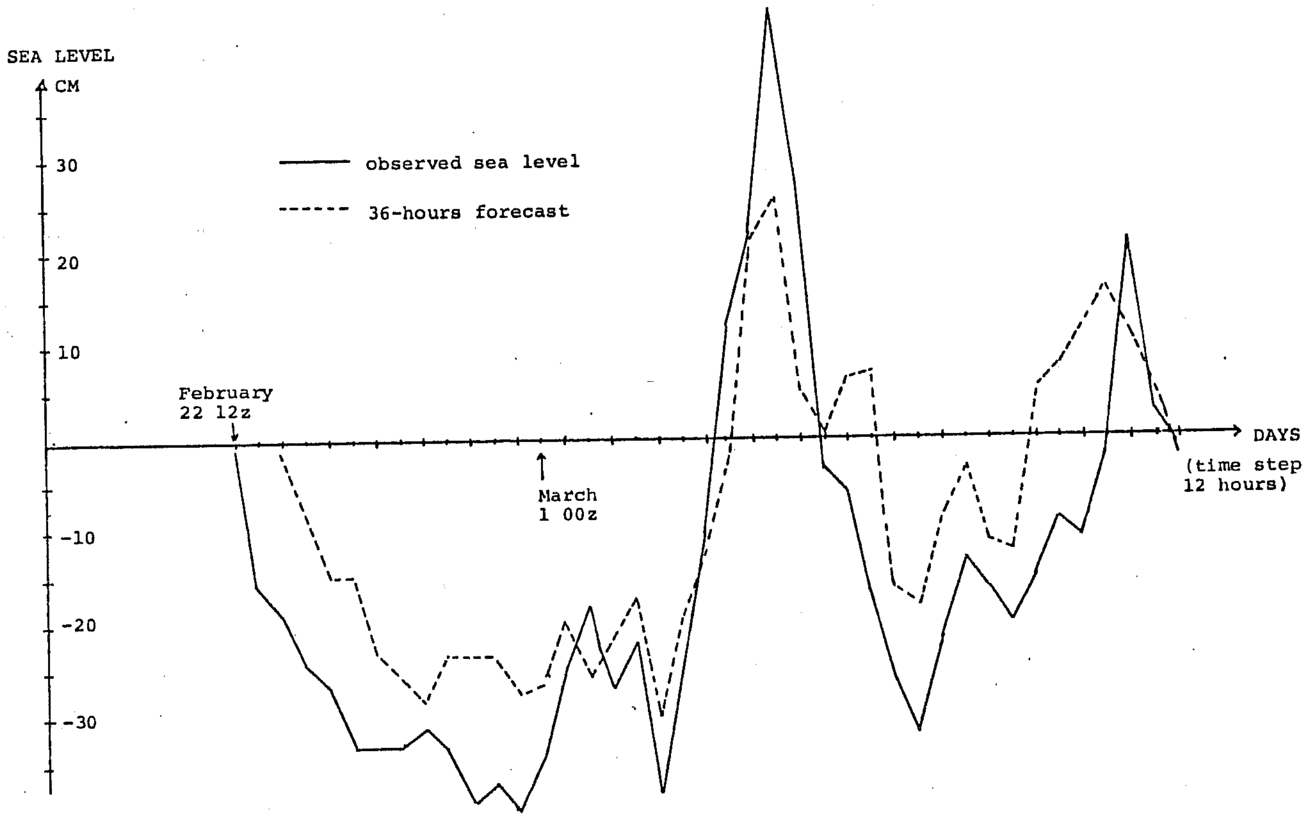


Figure 17

36-hours predicted sea level in Gothenburg compared to observations during the period February, 22 - March, 14.

Conclusions

The investigations reported here have shown that statistical methods in connection with forecasts of surface pressure over a large area can be utilized for predicting sea level variations in the Baltic.

A number of improvements on the model are possible and preparations for these are actually being made. In its present shape the model takes only 10 seconds CPU-time for a medium size computer like the UNIVAC 1100. A finite difference model giving the same forecasting possibility would require much longer computational time.

References

Holmström, I. & Stokes, J. Statistical forecasting of sea level changes in the Baltic. SMHI Rapporter, Nr RMK 8 (1977).

Lazarenko, N.N. Mean Sea Level and Volume Variations of the Baltic Sea for 1951 - 1970 Period. Report from the 5th meeting of experts on the waterbalance of the Baltic, Rostock, May, 23-27 1977.