STATISTICAL FORECASTING OF SEA LEVEL CHANGES IN THE BALTIC

by Ingemar Holmström and John Stokes

SMHI Rapporter

METEOROLOGI OCH KLIMATOLOGI

Nr RMK 9 (1978)

SVERIGES METEOROLOGISKA OCH HYDROLOGISKA INSTITUT

STATISTICAL FORECASTING OF SEA LEVEL CHANGES IN THE BALTIC

by Ingemar Holmström and
John Stokes

SMHI Rapporter

METEOROLOGI OCH KLIMATOLOGI

Nr RMK 9 (1978)

SVERIGES METEOROLOGISKA OCH HYDROLOGISKA INSTITUT

Norrköping 1978

SMHI, RMK 9 (1978)

STATISTICAL FORECASTING OF SEA LEVEL CHANGES IN THE BALTIC

by

Ingemar Holmström and John Stokes
Swedish Meteorological and Hydrological Institute

SUMMARY

By expanding sea level data from 6 Swedish observation stations into empirical orthogonal functions a very simple picture of the response of the Baltic to atmospheric forcing is obtained. It is found that not less than 65.5 per cent of the total variance is due to a general rise or lowering of the whole surface. The time scale corresponds to the time scale of large scale atmospheric disturbances. This interdependence has been used in order to establish a regression equation between surface pressure fields and sea level variations which is used for prediction. In the statistical treatment extensive use is made of the empirical orthogonal function technique.

SAMMANFATTNING

Genom utveckling av vattenståndsdata från 6 svenska observationsstationer i empiriskt ortogonala funktioner har erhållits en mycket enkel bild av Östersjöns respons på atmosfärens inverkan genom vind och tryck. Inte mindre än 65.5 procent av den totala variansen representeras av en allmän höjning eller sänkning av hela ytan. Tidsskalan är här av samma storlek som tidsskalan hos storskaliga atmosfäriska störningar. Detta beroende har utnyttjats för att bestämma sådana regressionsekvationer mellan lufttrycksfält och vattenståndsvariationer som sedan utnyttjas för prognoser. I den statistiska behandlingen har tekniken med empiriskt ortogonala funktioner utnyttjats i hög grad.

1. 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 generally have increased in size with the result that for the larger ones, access to certain ports is only possible at normal or high sea level. At some passages ships pass over thresholds with only 20-30 cm free water under the keel. On such occasions wind, waves and swell will naturally also influence the possibility for safe passage. To some extent the difficulties are also due to the secular rise of the land in the major part of Scandinavia. It has its maximum on the Swedish coast of the Sea of Bothnia where it is of the order of 1 cm per year. It will over not a too long period of time have considerable effect on the shipping.

For planning the shipping and for operational purpose, forecast of sea level and sea level variations will obviously be of great economic value. For the Baltic, tidal variations of sea level are in this connection of little interest, they are only of the order of a few centimeters. The same is probably true also for sea level changes due to variations in river discharge. The dominating factors influencing sea level are surface pressure and surface wind stress and since these can be predicted 2 to 3 days ahead, it should be possible also to make predictions of sea level changes.

Two different methods are available for this purpose. One is a numerical finite difference model of the Baltic, forced by surface stress and pressure. The second is a purely statistical regression method. Both methods were tested for Lake Vänern and gave about the same accuracy in predicted values of sea level. However, with regard to complexity and to computer time required for routine calculations, the statistical model was considerably more economic and it was therefore decided to test a statistical model on the Baltic.

2. COVARIANT SEA LEVEL CHANGES

In order to limit the initial statistical treatment, data from only five sea level stations in the Baltic were included in the first experiment. Taken from north to south, the stations are (see fig 1) Furuögrund, Ratan, Landsort, Kungsholmsfort and Ystad. In order to investigate at the same time a possible interdependence with conditions on the Swedish west coast, a sixth station, Smögen, was added.

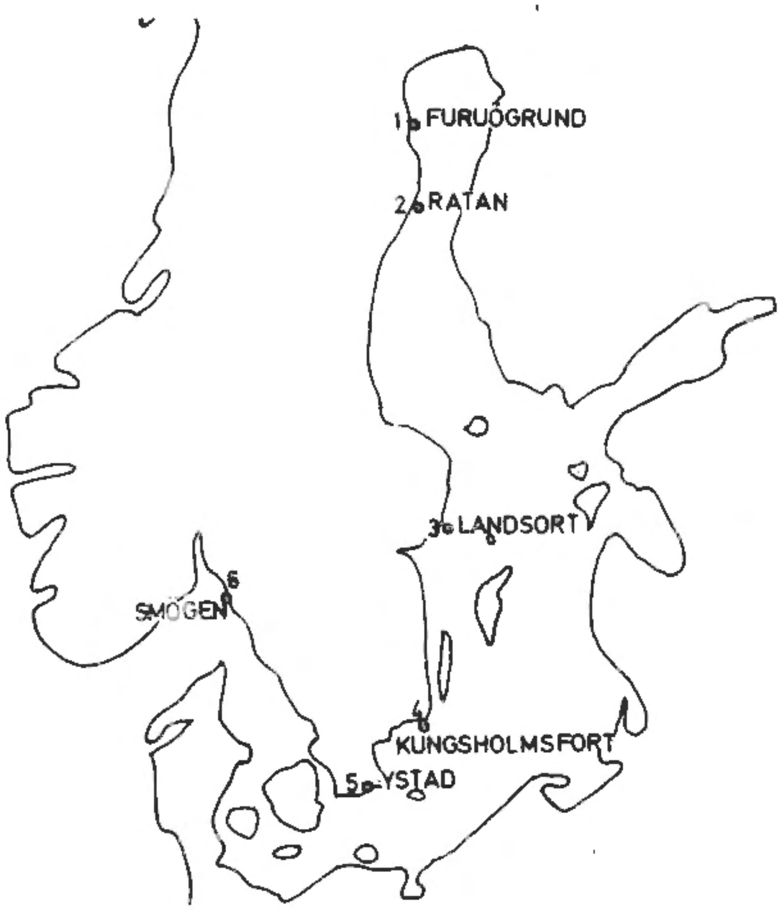


Fig 1 Position of sea-level observation stations.

Data were available for 00, 04, 08, 12, 16 and 20 hours local time.

The period 1 June 1967 - 31 December 1970 or 1310 days with 7860 observations at each station was taken as a basis for the statistical analysis of the behaviour of the Baltic and for calculating regression coefficients. A second period, 1 January 1971 - 31 December 1973 or 1096 days with 6576 observations at each station was retained as a test period.

The basic data period from 1967 to 1970 was sufficiently short for secular changes to be neglected. Substracting an arithmetic mean, taken over all observations at each station, gave deviations in cm from normal sea level. These are denoted $s_i(t)$, $i(=1, 2, \ldots 6)$ being the station number counted from north to south (see fig 1).

In order to analyze the systematic behaviour of the Baltic with respect to sea level changes, the functions $s_i(t)$ were expanded into empirical orthogonal functions (e.o.f.).

The expansion had the form

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

where the amplitudes $\beta_n(t)$ and the local response functions h_{ni} were determined from basic data to give the series an optimized convergence. It should be noticed here that the amplitudes $\beta_n(t)$ do not depend on i and therefore are common for all stations. The expansion thus determines the extent to which sea level variations at different locations are covariant.

Due to the requirement of optimized convergence the expansion (1) is doubly orthogonal so that

$$(2) \qquad \int_o^T \beta_n(t)\beta_m(t)\, dt = \delta_{nm}\int \beta_n^2(t)\, dt;$$

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

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

Taking the square of (1), summing over i and integrating over t we have, due to the orthogonality

$$(3) \qquad \frac{1}{6T}\sum_{i=1}^{6}\int_o^T s_i^2(t)\, dt = \frac{1}{T}\sum_{n=1}^{6}\int_o^T \beta_n^2(t)\, dt$$

where on the left hand side we have the mean variance of all data and on the right hand side the contribution to this variance from the different terms in the series (1). It is thus easy to calculate, in per cent, the relative contribution to the total variance from each term in the series.

Results from these calculations are shown in fig 2 where values of the normalized functions h_{ni} are plotted at intervals corresponding to the geographical distance between the stations. For these functions only point values are obtained. The lines drawn between the points do not indicate that linear interpolation is everywhere possible. They are drawn purely for illustrative purpose.

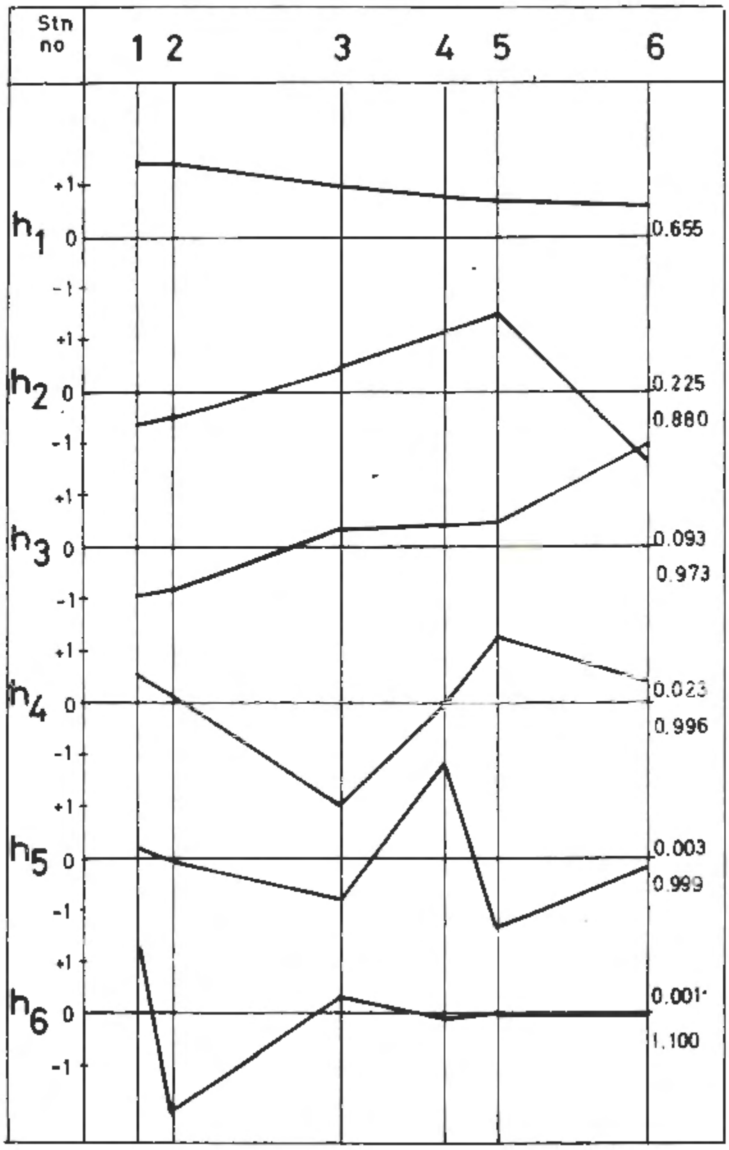


Fig 2 Normalized h_{ni}-functions.

To the right in the figure is also given the relative variance in per cent for each of the terms in the series as well as accumulated relative variance.

It is seen that the first function h_{1i} covers not less than 65.5 per cent of the total variance. Thus, with reservation for the small number of stations, it seems that almost two thirds of sea level variations in the Baltic and on the west coast of Sweden consist of a simultaneous rising or lowering, most pronounced in the northern parts of the Baltic and least pronounced on the west coast. A covariance as large as this must naturally depend on a forcing that is characterized by a very large scale, a fact that will have to be taken into account when atmospheric predictors in the regression scheme are selected.

The functions h_{2i} and h_{3i}, accounting for 22.5 and 9.3 per cent respectively of the total variance, are very similar for the stations in the Baltic. The main difference is found at station nr 6 on the west coast which in h_{2i} is out of phase and in h_{3i} in phase with stations 4 and 5. Thus, if station 6 had been excluded from the data, h_{2i} and h_{3i} would have been approximately replaced by one function only, representing somewhere around 30 per cent of the total variance and corresponding to a general north-south tilt of the Baltic.

The functions h_{4i} and h_{5i} are small on the west coast and also small in the north. They represent 2.3 and 0.3 per cent of the total variance and correspond to small scale variations in the southern part of the Baltic.

Finally h_{6i}, representing only the remaining 0.1 per cent of the total variance, corresponds to even smaller scale variations taking place in the Sea and Bay of Bothnia.

It is in this connection instructive to look at the behaviour of the amplitude functions $\beta_n(t)$. A sample over a period of two months is shown in figure 3. Since the functions h_{ni} are normalized the magnitude of the variations of the functions $\beta_n(t)$ as seen in the figure is representative for the contribution from each term in the series expansion (1) to sea level variations. It is seen that the amplitudes decrease considerably with increasing n but that occasionally amplitudes in higher modes may be quite large. It is also seen that the frequency increases with n, indicating that the influence of small scale and more rapid changes in the weather situations become more important as a forcing for the higher modes.

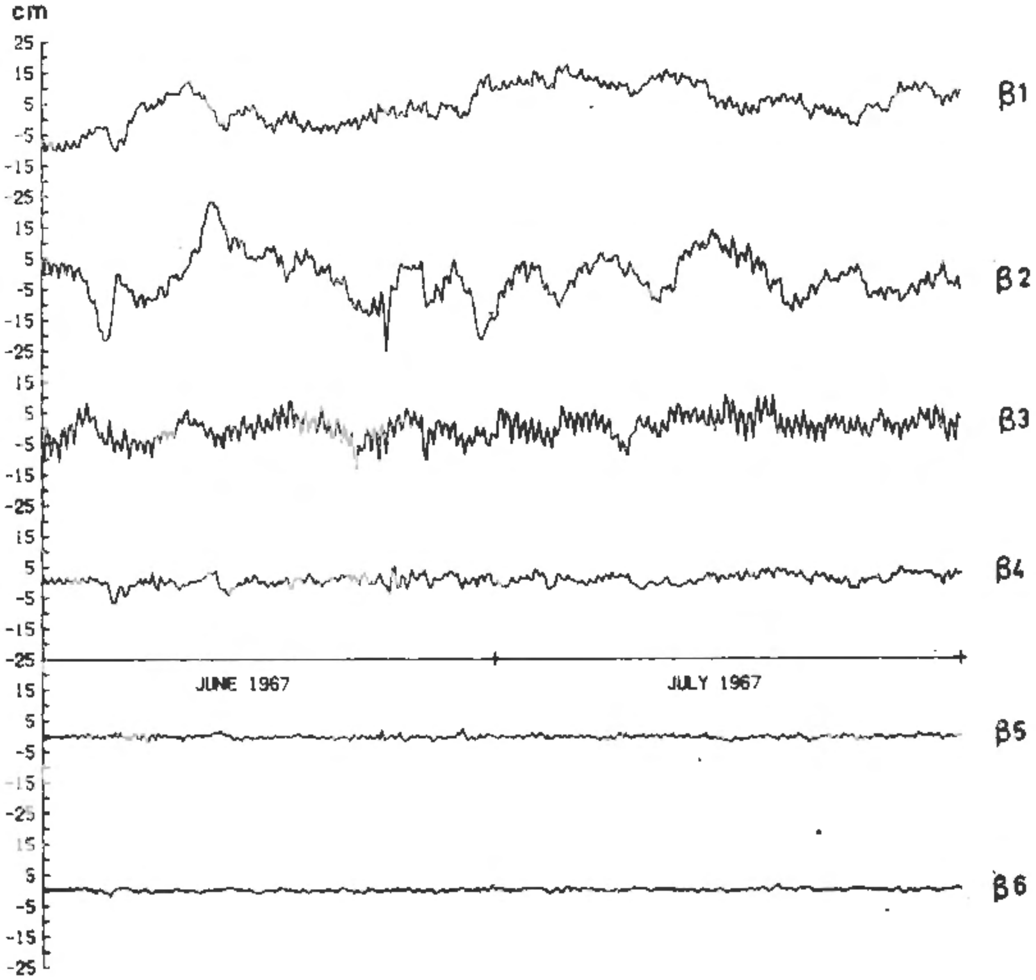


Fig 3 $\beta_n(t)$-functions over the period June-July, 1967.

It is also instructive to look at long term variations of the amplitude functions.

For this purpose daily mean values of $\beta_n(t)$ for the entire period 1 June 1967 - 31 December 1970 are presented in fig 4.

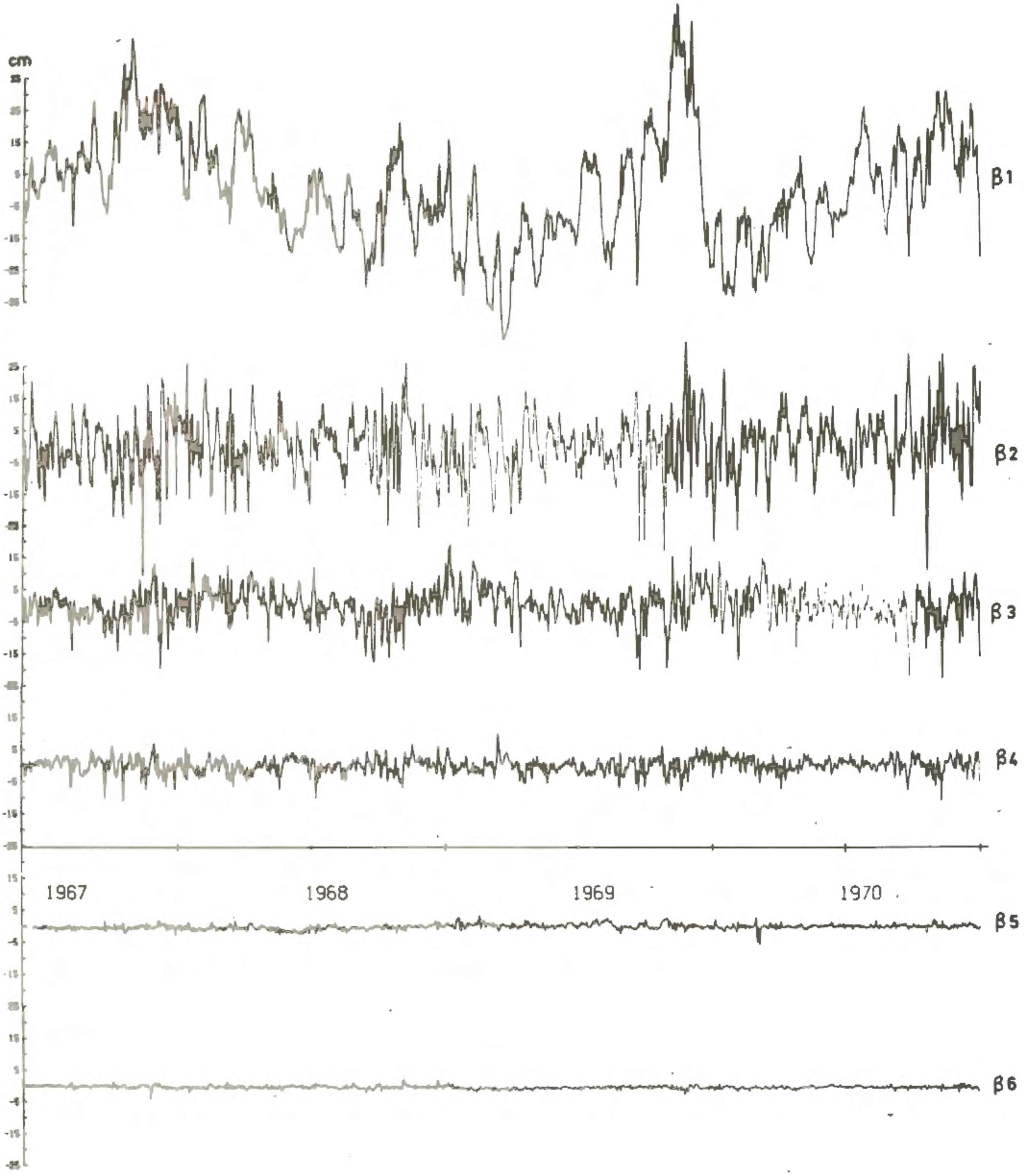


Fig. 4 $\beta_n(t)$ for the period 1 June 1967 - 31 December 1970.

Except for the year 1968 one can in $\beta_1(t)$ see the same yearly trend, as indicated by the broken line. The trend is, however, reversed during 1968 so that a longer series will be required for determination of a typical yearly trend.

In order to investigate characteristic features of the amplitude functions $\beta_n(t)$ energy spectra have been determined from auto-covariance functions, calculated for a maximum phase shift of 26 days. The limit here was determined entirely from practical reasons.

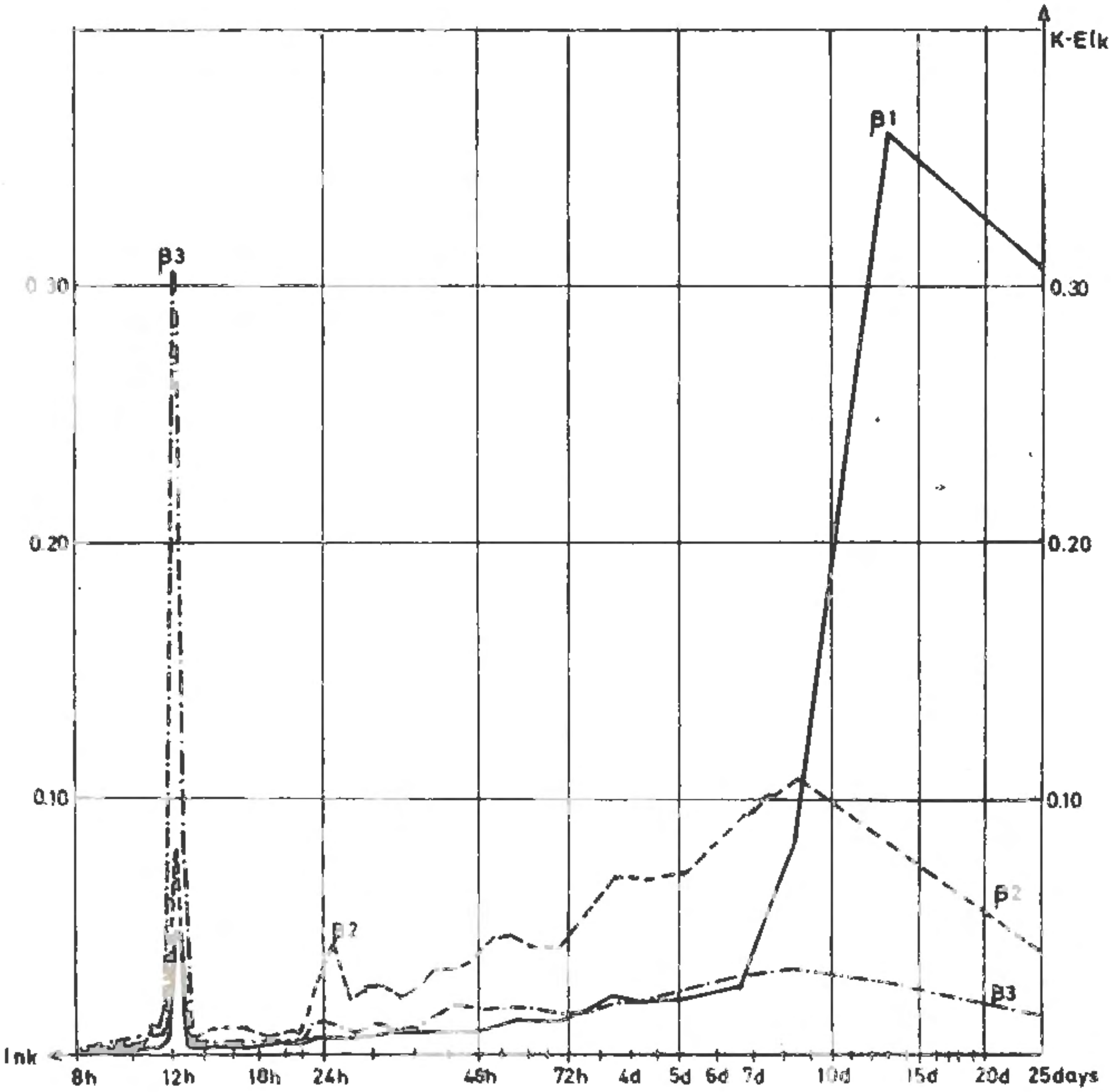


Fig 5a Frequency spectra for the functions $\beta_n(t)$.

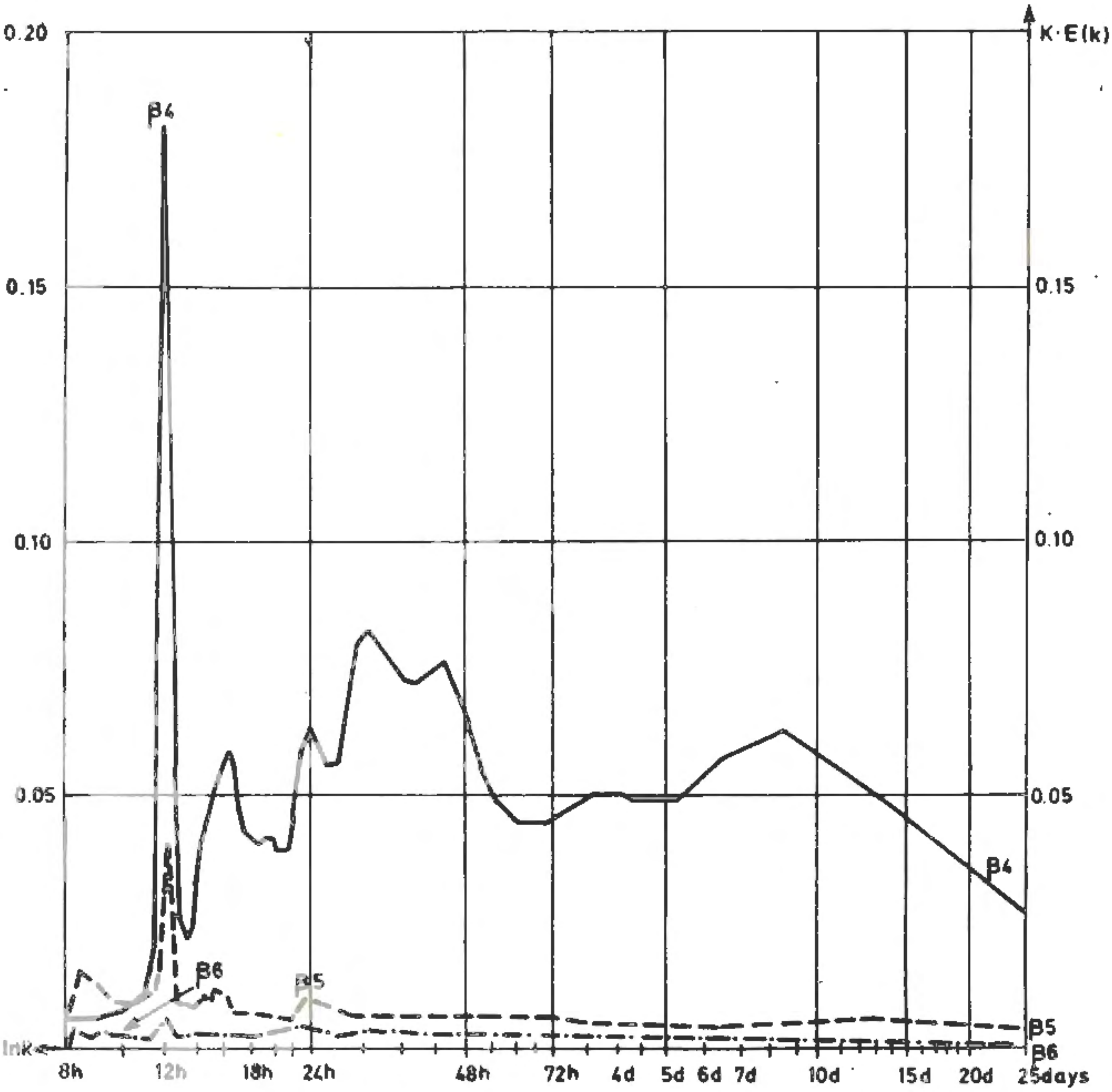


Fig 5b Frequency spectra for the function $\beta_n(t)$.

It is seen in fig 5a that the main part of the energy in β_1 and β_2 is found in frequencies that are typical for large or medium scale atmospheric processes. In forecasts of sea level variations it is therefore necessary to take these processes into account. In β_3 the same holds true except for a very narrow and pronounced peak at about 12h. Since h_{3i} has a large value at station 6, this peak is believed to be the result of tidal variations at the west coast of Sweden.

It is first in β_4 (note the difference in scale between 5a and 5b), representing 2.4 per cent of the total energy, that we find a large part of the energy also in frequencies that may be related to external seiches as well as rather rapid atmospheric forcing.

Since the observations utilized in this investigation are taken only every four hours, the resolution is evidently coarse in these intervals and it is not possible from the curves to determine typical frequencies with satisfactory accuracy. For this purpose a higher resolution in time is evidently required.

3. PREDICTORS

It has already been pointed out that the form of the function h_{1i} indicated the presence of a pronounced large scale forcing on the Baltic. It was therefore considered necessary, in the choice of predictors, to include a large area for pressure and wind. The surface pressure data required for prediction were taken from grid point values in a 7x7 grid in the routine numerical analysis and forecast model of the Swedish Weather Bureau (see fig 6 and 7).

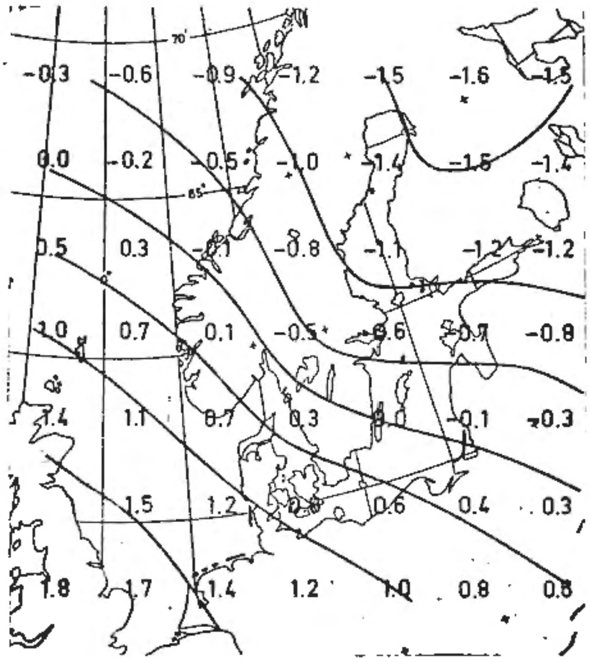
For the same period, 1 June 1967 - 31 December 1970, these pressure data were also expanded into empirical orthogonal functions. The expansion had the form

$$(4) \qquad p_j(t) = \overline{p_j} + \Sigma\alpha_n(t)\ g_{nj}$$

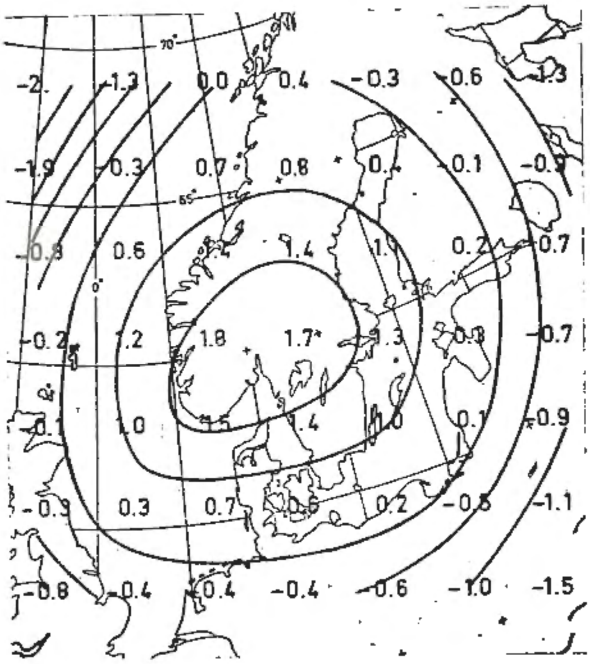
where j indicates grid point and $\overline{p_j}$ a local average over the period.

The reason for substracting out a local mean of the surface pressure is that the local sea level mean, which was separated out from sea level data, naturally should correspond to a mean pressure and wind field.

In the expansion (4) the g_{nj} are again normalized functions of position, determined from the pressure data, and after that considered as a known and unchanged orthogonal set. The first 8 functions are shown in fig 6 and 7. The figures in the lower corner to the right give in per cent the relative part of each field of the total variance as well as the accumulated relative variance. It is seen, as expected, that the convergence is here much slower than in the sea level data and in order to arrive at a variance reduction of 98 per cent not less than 13 terms in the expansion were needed. In the figures isolines are drawn for an interva of 0.5 units.



g_{1j} 38.2 (38.2) g_{2j} 33.2 (71.4)

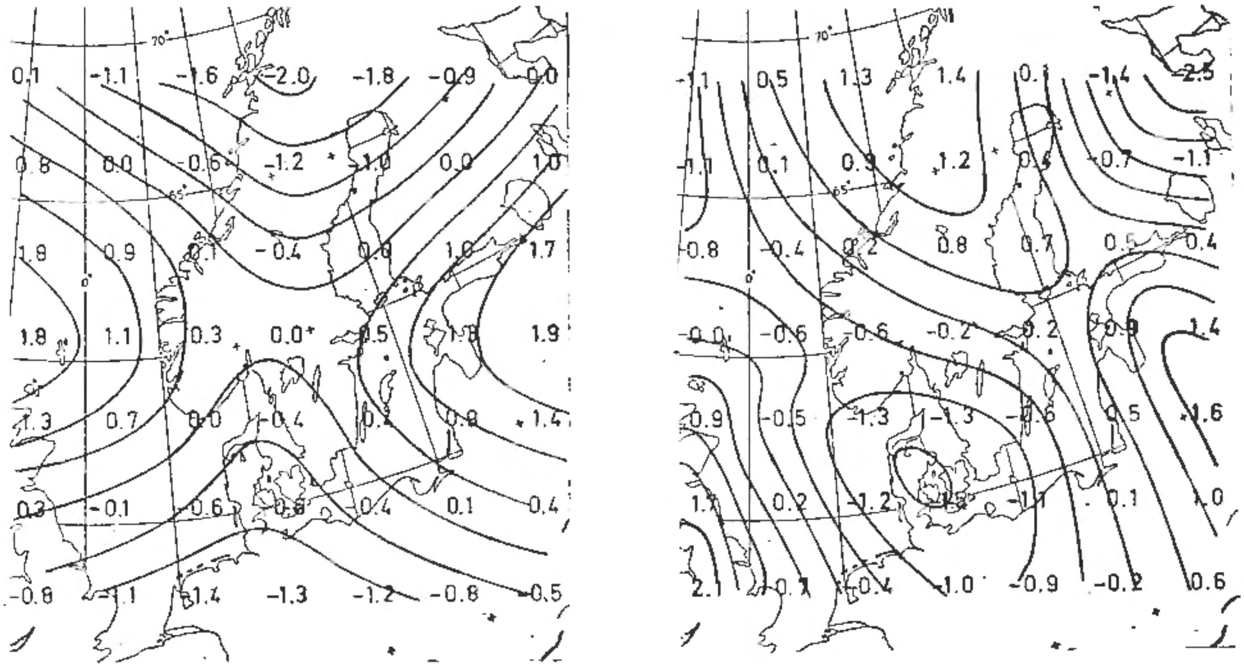


g_{3j} 8.5 (79.9) g_{4j} 7.5 (87.1)

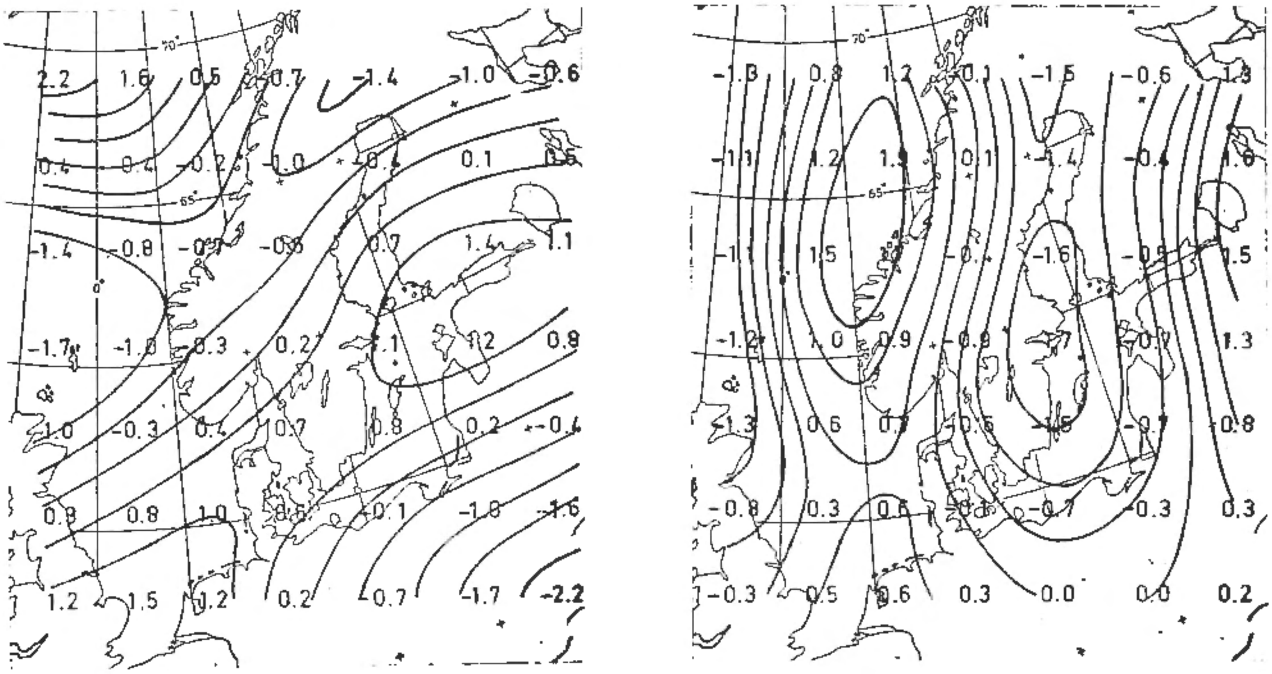
Figure 6

The functions g_{1j} - g_{4j}, j = 1-49.

The numbers below the figures give the relative variance in each mode. (In parenthesis 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 7

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

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

The advantage of expanding the surface pressure field into normalized surface functions and time varying amplitudes is easily seen if we apply a finite difference gradient operator to both sides of the relation (4).

We obtain

$$(5) \qquad \nabla[p_j(t) - \overline{p_j}] = \sum_m \alpha_m(t) \; \nabla g_{mj}$$

Since the fields g_{mj} are determined once and for all, their gradients are also fixed functions of the coordinates. Assuming a linear relationship between pressure gradient and local wind as influenced by local topography it is seen from (5) that the amplitudes $\alpha_m(t)$ incorporate a determination of the wind field as well as the pressure field. They are therefore suitable predictors for the amplitude functions $\beta_n(t)$ in (1), from which local sea level changes may be calculated.

Normally the stress excerted by the wind is set proportional to the wind multiplied by its absolute value. This would imply that second order products of the type $\alpha_m(t)|\alpha_n(t)|$ should be used as predictors for stress, while first order values should be used in order to describe the influence of differential pressure. Since this would considerably complicate the calculation of the regression coefficients it was decided to test to what extent first order regression would be sufficient. An expected result would then be that small changes would be overpredicted and large changes underpredicted.

4. REGRESSION EQUATION

Since sea level variations in the Baltic may imply transport of large masses of water through comparatively narrow passages such as for instance, Öresund, the Belts and the Åland Sea it is to be expected that the response to varying pressure and wind is not immediate. Instead, sea level changes are probably to be considered as an integrated effect of wind and pressure variations over a period of a certain length. It was therefore considered necessary to include also past values of the α-functions in the regression scheme.

One may also expect that, given a specific sea level situation, the Baltic will require a certain time to go back to normal, possibly through damped oscillations. Due to the difficulties to determine response functions when the system is continuously subject to varying forcing it was considered easier to take this effect into account by including also past values of $\beta_n(t)$ among the predictors.

We denote by t the time at which the sea level forecast is issued and by $t+m_1\Delta t$ the time for which it is to be valid. Routine surface pressure analyses and forecasts are only available for 00 and 12 GMT and it was therefore necessary to take $\Delta t = 12$h. Furthermore, routine surface pressure forecasts are at present only made up to 48 hours ahead and m_1 could therefore only be given the values 1, 2, 3 or 4.

For prediction of $\beta_n(t+m_1\Delta t)$ the following predictors were used

a. $\beta_j(t-m\Delta t)$ where $0 \leqslant m \leqslant m_o$, corresponding to β-values over the past few days and derived from observed sea level data at the six stations. From the behaviour of covariances it was decided to make $m_o = 6$, independent of the value of m_1.

b. $\alpha_j(t-k\Delta t)$, where $0 \leqslant k \leqslant k_1$ corresponding to α-values derived from observed surface pressure fields during a past period. For different values of m_1, k_1 has also been given different values in order to save computer time.

c. $\alpha_j(t+r\Delta t)$ where $0 \leqslant r \leqslant m_1$, corresponding to α-values derived from predicted surface pressure fields. For large m_1, k_1 in b had to be reduced and the procedure adopted was to make $m_1 + k_1 = 7$. Thus for long forecasts, past pressure influence was only taken into consideration to a limited extent.

Preliminary tests clearly showed that linear regression would give very poor results in predicting β_5 and β_6. Since these functions correspond only to 0.4 per cent of the total variance, no further attempts were made to establish prediction equations for these. It was also indicated that they would be of little value as predictors and they were therefore excluded from further computations.

High order surface functions g_{nj} in the expansion (4) of the pressure field in general showed small scales and corresponding $\alpha_n(t)$ functions a high frequency. Their influence on sea level changes was therefore assumed to be very small and in the choice of predictors the expansion (4) was therefore truncated at n = 13, giving on basic data a residual variance of 2 per cent. With the limitations mentioned above the regression equation for which the coefficients A and B had to be determined from the period of basic data had the following form

$$\begin{aligned}\beta_n(t+m_1\Delta t) &= \sum_{i=1}^{13}\sum_{m=-k_1}^{m_1} A_{nim}\alpha_i(t+m\Delta t)\\ &+ \sum_{i=1}^{4}\sum_{m=0}^{m_o} B_{nim}\beta_i(t-m\Delta t)\end{aligned}$$

Of course all these predictors carry a considerable amount of redundant information and the system of equations for determination of A_{nim} and B_{nim}, both functions also of m_1, is almost singular. In order to arrive at a useful regression equations it was therefore necessary to orthogonalize the predictors. This was made in the most efficient way, i.e. by expanding $\alpha_i(t+m\Delta t)$ and $\beta_i(t-m\Delta t)$ into a common set of empirical orthogonal functions, preceded by a normalization. The amplitudes in this expansion, truncated at 99.1 per cent of the variance corresponding to 91 terms, were then used as predictors and corresponding regression coefficients determined from data for the period 1 June 1967 - 31 December 1970.

5. RESULTS

In calculating regression coefficients a choice had to be made between using predicted surface pressure fields (a MOS method) or using observed data. Since the model used at the Swedish Weather Bureau was planned to be changed considerably during the near future the MOS method was considered inappropriate at the time of this investigation.

Observed surface pressure analyses have therefore been used for calculating the necessary regression coefficients as well as for control calculations during the test period.

Since forecasts of sea level are based on prediction of the amplitude functions $\beta_n(t)$, to be introduced in the series expansion (1), it is of particular interest to test in the first instance the prediction of $\beta_n(t)$. Results are shown in fig 8.

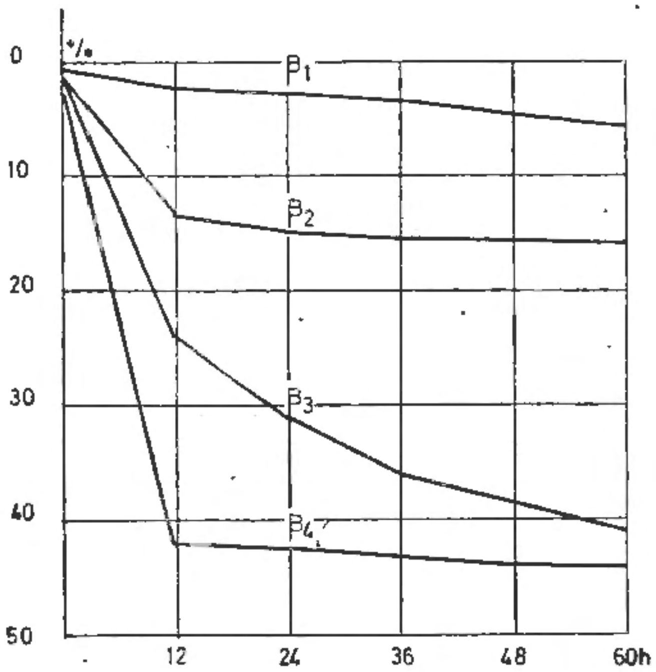


Fig 8 Increase of residual variance with prediction time for β_n, n=1-4.

Observed pressure data from the test period, 1 January 1971 - 31 December 1973, have here been used to simulate pressure forecasts.

The figure shows for forecast periods up to 60 hours the extent to which the observed variance in β_1, β_2, β_3 and β_4 is covered by the prediction and how the residual (unpredicted) variance increases. It is seen that β_1 is well predicted, the unpredicted variance being less than 6 per cent for a 60h forecast. As could be expected the accuracy in the prediction decreases for higher order β-functions.

It is also seen that the error in β_2, β_3 or β_4 levels out after 12 hours, indicating that the surface pressure influence on these functions is rather immediate and not very cumulative.

As a second test on the method, sea level forecasts have been made for the 5 stations in the Baltic for 00 and 12h on each day of the test period. The distribution of errors in 24-hour forecasts is shown in fig 9 for the station Furuögrund. The reason for the uneven distribution of the errors may be the fact that the same mean sea level value has been used for this period as for the basic data period while, in reality, a land rise of the order of a few centimeters should have been taken into account.

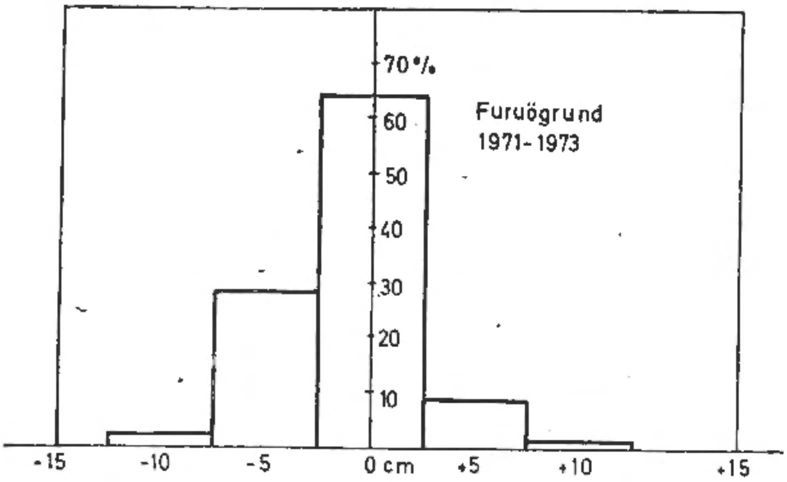


Fig 9

Error distribution 24-hour forecasts at Furuögrund based on correct pressure data.

For all five stations the results in the form of r.m.s. errors from the test period are shown in fig 10. The curves indicated by 1 give the increase in r.m.s. error for persistence forecasts. Curves 2 give the increase in error if forecast surface pressure fields are used to calculate α-values in the regression equation. Curves 3 give the error if observed pressures are used instead of predicted and thus presupposes correct pressure forecasts.

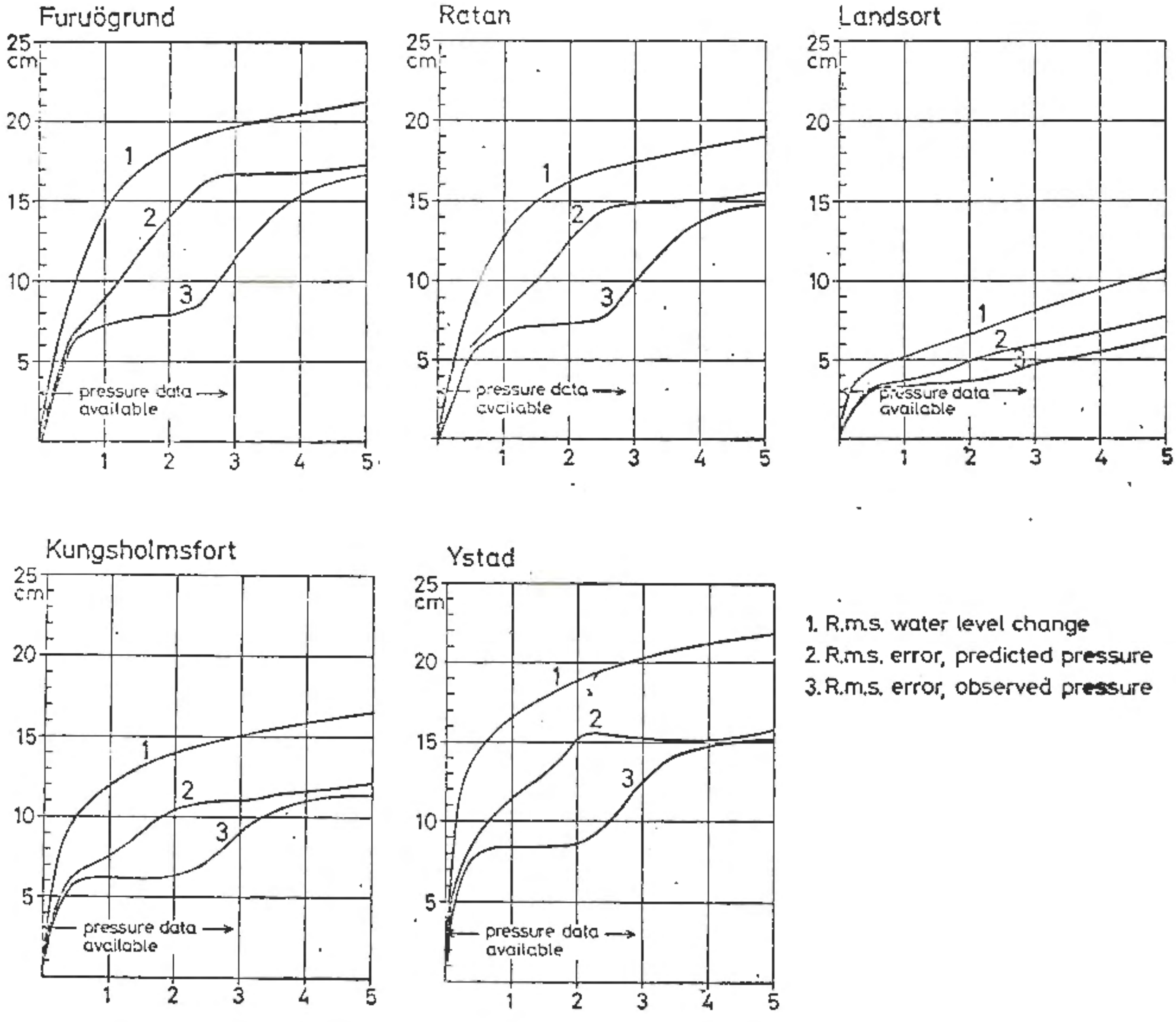


Fig 10 R.m.s. forecast errors from test period.

In these computations pressure data have been utilized up to a forecast period of 48 hours. The extension of the curves up to 5 days has been made with no pressure data available for the period 48-120 hours. The purpose has here been to determine if sea level forecasts for an extended period could be of any value even if surface pressure predictions were not available. The answer here is rather negative. It is seen that in general the r.m.s. after 2 days increases at almost the same rate as in the persistence forecast.

6. CONCLUSION

The investigation presented here has shown that it is quite possible to use a linear regression scheme for predicting sea level changes in the Baltic, utilizing forecast values of surface pressure over a rather large area. Since the total computer time required for a forecast is very small, (less than 20 seconds CPU-time on SAAB D23), the method can be used operationally. However, a number of improvements, simplifications and extensions are possible. A list of these is given below, in some cases together with a short discussion of the implications or the methods to be used.

a. In this first experiment the calculations have been made utilizing only five stations in the Baltic. An extension to other places will certainly be required if the method is to be used in routine forecasting. This will require a mapping of the functions h_{ni} along the coast lines of the Baltic. This can be done in various ways.

The regular behaviour of the functions, so far determined, indicates that linear interpolation should give rather satisfactory results except in the vicinity of large variations in bottom topography. As a second alternative one may use the β-functions determined in this investigation and their orthogonality in order to extend the h-functions to new stations. Finally, as a third alternative, one may make a complete recalculation of the eigenfunctions, utilizing data from all available stations. Since the influence of bottom topography and specially treshold areas on the form of the h-functions is not well known the linear interpolation method is not considered very satisfactory. The second method will therefore be tested since the third alternative will involve a considerable amount of calculation.

b. In the regression equations determined here the predictors certainly carry a considerable amount of redundant information and the resulting singularity was only avoided by orthogonalizing the predictors in a rather arbitrary way. The method also had the result that unimportant predictors could not easily be found and excluded. One indication of this may be seen in fig 10 where the curves 2, giving the r.m.s. error presuming correct pressure forecasts up to 2.5 days, after this point show an increase in error of about the same rate as the persistence forecast. A probable interpretation is that past sea level data are unimportant as predictors, except possibly for β_4, β_5 and β_6. It should therefore be possible to reduce considerably the number of predictors.

c. From the results, especially the form of h_{ij}, it seems likely that the Baltic to a certain extent behaves as part of a much larger system, i.e. a "fjord" with its boundary situated between southwestern Norway and Scotland. The scale of the system is then comparable to the synoptic one. It may therefore be necessary to include a larger area of surface pressure predictors in order to determine more accurately the sea level variations in the Baltic. If this extension is made it may also be possible to include sea level variations in the North Sea, an area which here has been left outside most of the computations. In such a case the program must take tidal variations into account.

d. For reasons already given, the "perfect prog method" has been used. When a new model has been introduced and a sufficient amount of forecast statistics has been collected, it would be more appropriate to use a MOS method. An interesting alternative to this is to separate the method into two parts,, the first one a statistical correction to the model output and the second a perfect prog method for the parameter of interest. A preliminary investigation a few years ago clearly indicated the feasibility for statistical corrections of model output and the obvious gain is that, in case of changes in the model, corrections are required in only one statistical program.

The work reported here has been sponsored in part by the Swedish Natural Science Research Council under contract G 2864-003 and by the National Swedish Environment Protection Board under contract no 7-156.

SMHI Rapporter, METEOROLOGI OCH KLIMATOLOGI (RMK)

Nr 1 Thompson, T, Udin, I och Omstedt, A, Sea surface temperatures in waters surrounding Sweden. (1974)

Nr 2 Bodin, S, Development on an unsteady atmospheric boundary layer model. (1974)

Nr 3 Moen, L, A multilevel quasi-geostrophic model for short range weather predictions. (1975)

Nr 4 Holmström, I, Optimization of atmospheric models. (1976)

Nr 5 Collins, W G, A parameterization model for calculation of vertical fluxes of momentum due to terrain induced gravity waves. (1976)

Nr 6 Nyberg, A, On transport of sulphur over the north Atlantic. (1976)

Nr 7 Lundqvist, J E och Udin, I, Ice accretion on ships with special emphasis on Baltic conditions. (1977)

Nr 8 Eriksson, B, Den dagliga och årliga variationen av temperaturer, fuktighet och vindhastighet vid några orter i Sverige. (1977)

Nr 9 Holmström, I och Stokes, J, Statistical forecasting of sea level changes in the Baltic.