551. 509. 314: 509. 5: 547. 5

INVESTIGATION OF THE FORECASTING ERROR OF A SIMPLE BAROTROPIC MODEL WITH THE AID OF EMPIRICAL ORTHOGONAL FUNCTIONS

PART II

BY

JUHANI RINNE

Department of Meteorology,
University of Helsinki

Abstract

The empirical orthogonal functions determined from a *small* sample of 500 mb height analyses, reducing the variance of this basic sample by $\sim$ 99%, have been in some adjacent samples capable of reducing the sample variance by $\sim$ 90%. Even the high-indexed components have been of importance. The deviations in the sample means played a minor role.

The method of forecast reproduction derived in Part I of this paper has been extended. A reproduced forecast will now include also information from the corresponding persistence forecast in addition to that of the conventional one. This new information seemed not to be of noticeable importance. The method, which theoretically could be expected to be a working one, has been shown empirically to be capable of reducing the forecasting error in the case of a longer forecatings period than three days, in spite of the smallness of the basic sample.

Forecasts of a new kind (probability variants) were prepared from the reproduced forecasts. In the case studied, neither the conventional nor the reproduced forecast showed a trace of a wave actually found in the verification analysis. However, it would have been possible to conclude from the simultaneous probability forecasts that the probability of the presence of such a wave had been very large, about 70%.

1. *Introduction*

In Part I, a method for eliminating non-atmospheric phenomena from a forecast utilizing empirical orthogonal functions has been presented. The sample used in the determination of the functions was realized to be too limited to generate functions including all of the 500 mb phenomena to be predicted. For this and other reasons it was impossible, when the method was applied to a sample of 24 hr forecasts, to draw any final conclusions concerning the efficiency of the method.

In Part II, the representativeness of the functions will be examined by the means of some samples which are adjacent to the sample mentioned above. The method of reproduction will be applied to barotropic forecasts of a longer period, *i.e.*, in conditions where forecasting errors are large when compared to the defects in the set of the determined orthogonal functions.

2. *The representativeness of the determined orthogonal functions*

The analysis samples used in this study are

S_{-1} consisting of the objective 00 GMT analyses of the 500 mb height of the 15th day of each month during 1965—68 (48 cases),

S_0	as above but	16th	days during	1965—68	(48 cases)
S_1	»	17th	»	1965—68	(48 cases)
S_2	»	18th	»	1966—68	(36 cases)
S_3	»	19th	»	1967—68	(24 cases) and
S_4	»	20th	»	1967—68	(24 cases).

The set of the functions consists of 33 empirical orthogonal functions determined from S_0, which in the following will be referred to as the basic sample. When the set of the orthogonal functions $f_\nu(i, j)$, $\nu = 1, 2, \ldots, 33$, is known, the height fields in S_0 may be expressed as

$$z(t, i, j) - \overline{z(t, i, j)} = \sum_{\nu=1}^{33} C_\nu(t) f_\nu(i, j) + \text{residual}, \qquad (2.1)$$

where $z(t, i, j)$ is the analyzed height in the grid point (i, j) and the bar indicates an averaging with respect to time t. The sample mean of S_0 will be denoted as $zm(i, j) = \overline{z(t, i, j)}$.

To give a more »familiar» description of the orthogonal functions their »wavelengths» are presented in Table 1. For the determination of the wavelengths, see the Appendix.

In the routine work, one does not usually know the actual mean conditions of the atmosphere. Hence an estimate of the true mean field should be used when a height field will be described by the functions

Table 1. Approximate wavelength of the functions $f_\nu(i, j)$.

Component number ν	»Wavelength» L [kms]	»Wavenumber» at 60° N
1	6900	2.9
2	7200	2.8
3	6700	3.0
4	5200	3.9
5	7900	2.5
6	6400	3.1
7	6300	3.2
8	6100	3.3
9	6300	3.2
10	5900	3.4
11	6100	3.3
12	5500	3.6
13	5200	3.9
14	5100	3.9
15	5000	4.0
16	5000	4.0
17	5200	3.9
18	4900	4.1
19	4400	4.6
20	4600	4.4
21	4000	5.0
22	4500	4.5
23	3200	6.3
24	4100	4.9
25	4000	5.0
26	4000	5.0
27	4000	5.0
28	3900	5.1
29	3600	5.6
30	3400	5.9
31	3200	6.3
32	3500	5.7
33	3300	6.1

determined from S_0. For instance, any height field of a sample may be given by

$$z(t\,,i\,,j) - zm(i\,,j) = \sum_{\nu=1}^{33} C_\nu(t)\, f_\nu\,(i\,,j) + \text{residual}, \tag{2.2}$$

where the actual sample mean has been estimated as $zm(i\,,j)$. The coefficients $C_\nu\,(t)$ will be determined from

$$C_\nu(t) = \sum_{i,j} [z(t\,,i\,,j) - zm(i\,,j)\,]\, f_\nu\,(i\,,j)\, dA(i\,,j)\,, \quad *) \tag{2.3}$$

where $dA(i\,,j)$ is the element of area. The coefficients $C_\nu(t)$, unlike those of the basic sample, may be mutually correlated. Averaging (2.2) with respect to the time yields

$$\overline{z(t\,,i\,,j)} - zm(i\,,j) = \sum_{\nu=1}^{33} \overline{C_\nu\,(t)}\, f_\nu(i\,,j) + \text{residual}\,. \tag{2.4}$$

Hence the quantities $\overline{C_\nu(t)}$ are the coefficients of the expansion of the difference between the true and the estimated sample mean. The variance of $z(t\,,i\,,j)$ becomes

$$\begin{aligned} \text{VAR} = & \sum_{i,j} \overline{(z(t\,,i\,,j) - zm(i\,,j))^2\, dA(i\,,j)} - \\ & - \sum_{i,j} \overline{(\overline{z(t\,,i\,,j)} - zm(i\,,j))^2\, dA(i\,,j)}\,. \end{aligned} \tag{2.5}$$

From the orthogonality of the functions $f_\nu\,(i\,,j)$, when substituting from (2.2) and (2.4), it follows that

$$\text{VAR} = \sum_{\nu=1}^{33} \overline{C_\nu(t)^2} - \sum_{\nu=1}^{33} \overline{C_\nu(t)}^{\,2} + \text{residual}\,. \tag{2.6}$$

The latter sum equals zero in the basic sample S_0.

Table 2 presents variance terms associated with the difference between the true and the estimated sample mean.**) The orthogonal functions have been determined to give an optimal representation of the 500 mb height *anomalies*. Obviously the representation is no longer so efficient when deviations of the sample *means* are depicted by the functions. In Table 2, the variance reduction does not exceed 75%. However, the unexplained part of the variance of $(\overline{z\,(t\,,i\,,j)} - zm(i\,,j))$

*) The erroneous areal elements determined in Part I will be applied throughout Part II except in the Appendix where the correct elements will be used.

**) The residual values in Table 2 are due to the residual term of (2.4).

Table 2. The variance reduction $\overline{C_\nu(t)}^2$ of the difference between the true and the estimated sample mean. Unit: gpm².

Component number ν	Sample S_{-1}	S_0	S_1	S_2	S_3	S_4
1	0.35	0.00	0.77	10.40	17.86	14.23
2	0.08	0.00	1.19	3.65	11.76	19.06
3	0.87	0.00	5.70	0.09	15.71	0.01
4	0.56	0.00	0.08	11.28	0.35	4.02
5	7.35	0.00	0.11	5.91	40.66	47.18
6	0.12	0.00	6.34	0.16	9.10	55.92
7	0.23	0.00	0.12	0.06	2.42	41.58
8	2.09	0.00	0.85	9.32	119.30	53.25
9	0.04	0.00	0.07	2.07	5.09	29.24
10	2.45	0.00	0.12	5.99	1.14	0.71
11	0.26	0.00	0.26	0.12	21.20	10.71
12	4.86	0.00	0.65	11.65	8.29	0.72
13	1.21	0.00	0.82	3.38	30.95	54.17
14	0.40	0.00	3.41	4.67	0.41	1.33
15	1.47	0.00	0.04	7.40	5.48	6.10
16	0.00	0.00	0.61	3.56	10.77	18.24
17	1.67	0.00	4.79	1.62	0.40	1.60
18	0.82	0.00	0.44	4.34	26.79	1.32
19	0.03	0.00	0.47	1.61	0.03	2.63
20	0.15	0.00	0.37	0.95	1.07	0.08
21	1.16	0.00	1.88	0.03	0.12	0.45
22	0.10	0.00	0.11	1.96	40.77	18.98
23	7.83	0.00	0.09	0.15	1.68	2.53
24	0.00	0.00	1.21	0.37	7.27	2.08
25	1.42	0.00	0.10	1.97	3.39	0.18
26	0.00	0.00	0.54	0.23	1.39	5.67
27	0.18	0.00	1.00	5.55	4.58	6.70
28	0.92	0.00	2.16	5.60	4.48	0.00
29	1.43	0.00	1.55	0.31	0.71	0.03
30	0.02	0.00	0.01	0.65	1.20	0.01
31	1.38	0.00	0.00	7.41	23.80	11.07
32	0.01	0.00	2.34	0.58	3.82	0.06
33	0.06	0.00	0.14	0.01	1.82	0.93
Σ	39.56	0.00	38.35	113.05	423.82	410.77
residual	47.06	0.00	51.24	102.14	160.12	205.03
total var.	86.62	0.00	89.59	215.19	583.94	615.80
variance reduction	45.7%	—	42.8%	52.5%	72.6%	66.7%

seems to be rather small, for instance, only 205 gpm² in the sample S_4. Some low-indexed components seem to be more important in S_3 and S_4, indicating large scale deviations in the mean state. Otherwise the distribution of the variance reduction is rather irregular. This irregularity and the inefficiency in the variance reduction show that the eigenfunctions are sensitive to the nature of the field to be built up.

Let the two components, $f_\nu(i, j)$ and $f_\mu(i, j)$, be of similar importance in the atmosphere. Let further f_ν be strongly and f_μ weakly represented in a certain sample. Then the variance reduction due to f_ν will be larger, whereas that due to f_μ will be smaller, *i.e.* $\nu < \mu$. Hence the importance of low-indexed components (f_ν) determined from a sample tends to be overestimated, whereas the high-indexed components (f_μ) probably are more important than the sample shows. This conclusion is well illustrated in Tables 3, 4, 5. The low-indexed components and the high-indexed components have in S_0 larger and smaller reduction values, respectively, than elsewhere. For instance, with the aid of S_0 less important components $\nu = 31$ and $\nu = 32$ have been found. However, according to the values in S_3 and S_4, these components seem to be more intense.

Comparison between Tables 2 and 3 shows that the contribution of the deviations of the mean field to the total variance has not been very great. Especially, the unexplained part of the variance of the oscillations of the mean state (»residual» in Table 2) is of minor importance.

3. *An application to forecasts*

In analogy to (2.1), the expansion of a forecast field $z'(t, i, j)$ will be

$$z'(t, i, j) - \overline{z'(t, i, j)} = \sum_{\nu} D_\nu(t) f_\nu(i, j), \tag{3.1}$$

where

$$D_\nu(t) = \sum_{i,j} [z'(t, i, j) - \overline{z'(t, i, j)}] \; f_\nu(i, j) \, dA(i, j). \tag{3.2}$$

The variance of the forecasting error is here defined as

$$ERR = \Sigma \overline{\left\{ [z(t, i, j) - zm(i, j)] - [z'(t, i, j) - \overline{z'(t, i, j)}] \right\}^2} dA(i, j). \tag{3.3}$$

Table 3. The variance reduction $\overline{C_\nu(t)^2}$ associated with the ν^{th} component in different samples. Unit: gpm^2.

Component number ν	Sample					
	S_{-1}	S_0	S_1	S_2	S_3	S_4
1	16505.46	16724.95	16821.54	16768.00	17101.03	16760.07
2	1250.68	1533.30	1422.34	1313.46	1281.93	1174.27
3	1141.82	1364.76	1320.42	1354.57	1138.26	805.69
4	752.28	1078.12	1073.61	936.49	789.34	479.30
5	812.61	922.00	873.25	804.30	553.18	694.49
6	686.96	839.01	695.47	413.42	321.20	500.88
7	630.73	623.68	522.32	450.66	468.51	478.15
8	524.08	582.20	430.83	556.79	1005.71	763.92
9	435.72	549.07	566.24	517.78	634.82	666.53
10	367.76	461.54	440.57	394.92	338.45	309.88
11	358.58	443.56	439.49	485.65	672.26	728.98
12	300.74	356.67	291.98	375.41	351.49	417.49
13	265.48	313.72	186.76	179.37	237.50	305.43
14	276.26	287.13	208.36	199.51	185.36	290.31
15	218.70	247.71	173.92	205.05	200.23	229.31
16	147.81	230.70	200.13	139.78	163.46	217.06
17	159.93	197.82	181.41	194.01	298.56	289.39
18	178.38	181.88	167.77	163.09	140.27	119.04
19	74.33	145.18	162.68	177.97	207.73	193.67
20	133.01	131.95	100.97	103.40	141.25	219.80
21	111.29	122.92	117.40	101.61	102.86	171.15
22	115.70	112.62	141.05	158.75	246.74	264.03
23	108.75	100.69	89.10	92.85	152.09	226.61
24	92.77	93.57	87.05	100.04	118.59	112.64
25	100.29	83.95	105.63	108.37	203.50	151.31
26	86.91	73.76	58.95	63.81	93.18	102.69
27	122.95	61.08	80.61	128.84	141.32	80.48
28	68.11	56.59	72.95	77.83	82.84	85.16
29	49.44	54.13	68.35	105.38	75.91	81.06
30	52.74	48.00	31.03	51.79	60.22	55.55
31	49.53	41.53	31.68	75.00	132.54	113.61
32	92.04	41.82	55.46	68.78	104.56	110.74
33	35.93	36.23	23.72	28.56	62.88	79.82
Σ	26307.79	28136.81	27243.02	26895.23	27807.74	27278.50
residual	2010.55	272.75	2008.54	2796.31	2895.72	3124.52
total var.	28318.34	28409.56	29251.56	29691.54	30703.46	30403.02
variance reduction	92.9%	99.0%	93.1%	90.6%	90.6%	89.7%

Table 4. As Table 3 but per cent.

Component number ν	Sample					
	S_{-1}	S_0	S_1	S_2	S_3	S_4
1	58.29	58.87	57.51	56.47	55.70	55.13
2	4.42	5.40	4.86	4.42	4.18	3.86
3	4.03	4.80	4.51	4.56	3.71	2.65
4	2.66	3.79	3.67	3.15	2.57	1.58
5	2.87	3.25	2.99	2.71	1.80	2.28
6	2.43	2.95	2.38	1.39	1.05	1.65
7	2.23	2.20	1.79	1.52	1.53	1.57
8	1.85	2.05	1.47	1.88	3.25	2.51
9	1.54	1.92	1.94	1.74	2.07	2.19
10	1.30	1.62	1.51	1.33	1.10	1.02
11	1.27	1.56	1.50	1.64	2.19	2.40
12	1.06	1.26	1.00	1.26	1.14	1.37
13	0.94	1.10	0.64	0.60	0.77	1.00
14	0.98	1.01	0.71	0.67	0.60	0.95
15	0.77	0.87	0.59	0.69	0.65	0.75
16	0.52	0.81	0.68	0.47	0.53	0.71
17	0.56	0.70	0.62	0.65	0.97	0.95
18	0.63	0.64	0.57	0.55	0.46	0.39
19	0.26	0.51	0.56	0.60	0.68	0.64
20	0.47	0.46	0.35	0.35	0.46	0.72
21	0.39	0.43	0.40	0.34	0.33	0.56
22	0.41	0.40	0.48	0.53	0.80	0.87
23	0.38	0.35	0.30	0.31	0.50	0.75
24	0.33	0.33	0.30	0.34	0.39	0.37
25	0.35	0.30	0.36	0.36	0.66	0.50
26	0.31	0.26	0.20	0.21	0.30	0.34
27	0.43	0.21	0.28	0.43	0.46	0.26
28	0.24	0.20	0.25	0.26	0.27	0.28
29	0.17	0.19	0.23	0.35	0.25	0.27
30	0.19	0.17	0.11	0.17	0.20	0.18
31	0.17	0.15	0.11	0.25	0.43	0.37
32	0.33	0.15	0.19	0.23	0.34	0.36
33	0.13	0.13	0.08	0.10	0.20	0.26
Σ	92.90	99.04	93.13	90.58	90.57	89.72
>33 (= residual)	7.10	0.96	6.87	9.42	9.43	10.28

Table 5. As Table 3 but cumulatively per cent.

Component number ν	Sample					
	S_{-1}	S_0	S_1	S_2	S_3	S_4
1	58.29	58.87	57.51	56.47	55.70	55.13
2	62.70	64.27	62.37	60.90	59.87	58.99
3	66.73	69.07	66.88	65.46	63.58	61.64
4	69.39	72.87	70.55	68.61	66.15	63.22
5	72.26	76.11	73.54	71.32	67.95	65.50
6	74.69	79.07	75.92	72.72	69.00	67.15
7	76.91	81.26	77.70	74.23	70.52	68.72
8	78.76	83.31	79.17	76.11	73.80	71.23
9	80.30	85.23	81.11	77.85	75.87	73.42
10	81.60	86.85	82.62	79.18	76.97	74.44
11	82.87	88.41	84.12	80.82	79.16	76.84
12	83.93	89.67	85.12	82.08	80.30	78.21
13	84.87	90.77	85.77	82.69	81.08	79.22
14	85.84	91.78	86.46	83.36	81.68	80.17
15	86.61	92.65	87.06	84.05	82.33	80.93
16	87.14	93.74	87.75	84.52	82.87	81.64
17	87.70	94.16	88.37	85.17	83.84	82.59
18	88.33	94.80	88.94	85.72	84.30	82.99
19	88.59	95.31	89.50	86.32	84.97	83.62
20	89.06	95.78	89.84	86.67	85.43	84.35
21	89.46	96.21	90.24	87.01	85.17	84.91
22	89.87	96.61	90.72	87.55	86.57	85.78
23	90.25	96.96	91.03	87.86	87.07	86.52
24	90.58	97.29	91.33	88.20	87.45	86.89
25	90.93	97.59	91.69	88.56	88.11	87.39
26	91.24	97.85	91.89	88.78	88.42	87.73
27	91.67	98.06	92.17	89.21	88.88	87.99
28	91.91	98.26	92.41	89.47	89.15	88.27
29	92.09	98.45	92.65	89.83	89.40	88.54
30	92.27	98.62	92.75	90.00	89.59	88.72
31	92.45	98.77	92.86	90.25	90.02	89.10
32	92.77	98.91	93.05	90.49	90.36	89.46
33	92.90	99.04	93.13	90.58	90.57	89.72

In (3.3) the mean field of the forecasts has been removed and replaced by zm, resulting in »forecast fields» $z' - \overline{z'} + zm$, the latter having zm as the mean field.

Hence it will be assumed that the mean field of the forecasts is known. It would be too crude, though perhaps more correct, to make use of

$zm(i, j)$ as an estimate of $\overline{z'(t, i, j)}$. The mean field settings cause a difference between ERR and the conventional concept of error variance.

From (2.1) and (3.1) one finds that

$$ERR = \overline{\sum_{i,j} [\sum_{\nu} C_\nu(t) \; f_\nu(i, j) - \sum_{\nu} D_\nu(t) f_\nu(i, j)]^2 \, dA(i, j)} \, .$$

From the orthogonality properties it follows that

$$ERR = \sum_{\nu} \overline{[C_\nu(t) - D_\nu(t)]^2} \, .$$

The term $\overline{[C_\nu(t) - D_\nu(t)]^2}$ may be referred as the variance of the forecasting error due to the ν^{th} component.

If a persistence forecast will be considered, then instead of (3.2) one has to make use of

$$D_\nu(t) = \sum_{i,j} [z(t_0, i, j) - \overline{z(t_0, i, j)}] \, f_\nu(i, j) \, dA(i, j) \, , \qquad (3.4)$$

where t_0 refers to the initial conditions.

If the ν^{th} component is eliminated, *i.e.* no forecast of it is given, then $D_\nu(t) = 0$. In the extreme case where all the quantities D_ν equal zero, the mean field $\overline{z'(t, i, j)}$ corresponds to the »forecast». Then we have $ERR = \overline{\Sigma \, C_\nu^2(t)}$.

The forecasts have been prepared from the objective 00 analyses of the 500-mb height of the 15th day of each month during 1967—68 (24 cases). Except for the sample period, which in every case is 1967—68, the initial conditions correspond to the sample S_{-1} and the 24, 48, 72, 96, 120 hour forecasts correspond to the samples S_0, S_1, S_2, S_3 and S_4, respectively.

There have been two forecasting models: the conventional simple barotropic model and a short-circuited model, where the Rossby parameter equals zero and the Cressman correction is rather large (*cf.* Part I).

A detailed analysis of error variance of four different forecasting »models» is given in Fig. 1. In the case of the 24 hour forecasts, the different forecasting models may be ordered according to their skill as follows: the conventional barotropic model (BT), the short-circuited model (SCBT), the persistence (P) and the elimination of a certain component (E). Exceptions are, for instance, the components $\nu = 1$, 11, 18, where the SCBT model is better than the BT model. The 24 hour forecasts have been discussed in more detail in Part I.

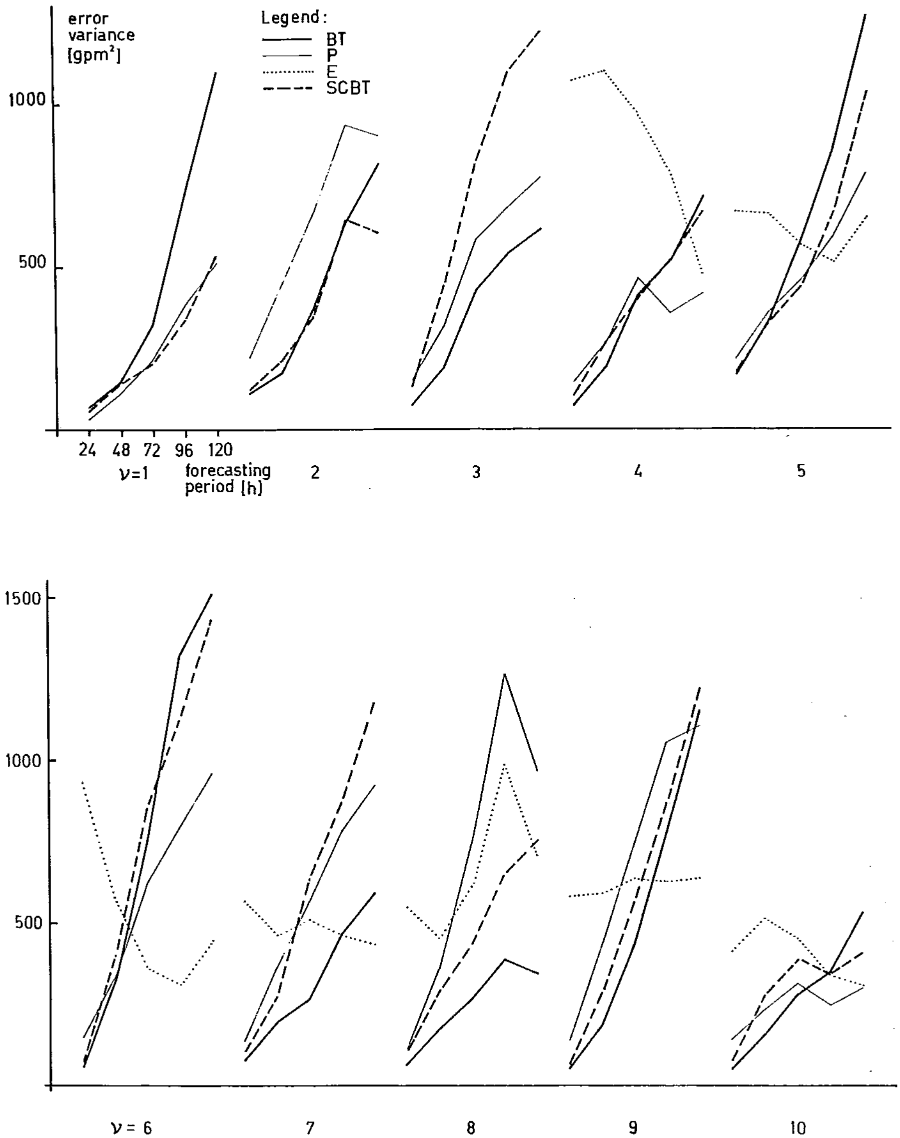


Fig. 1. Error variances associated with the ν^{th} component versus the length of the forecasting period in the case of the barotropic model, the short-circuited barotropic model, the persistence forecast and the elimination. The curves of E have not shown when $\nu < 4$ because of too large values.

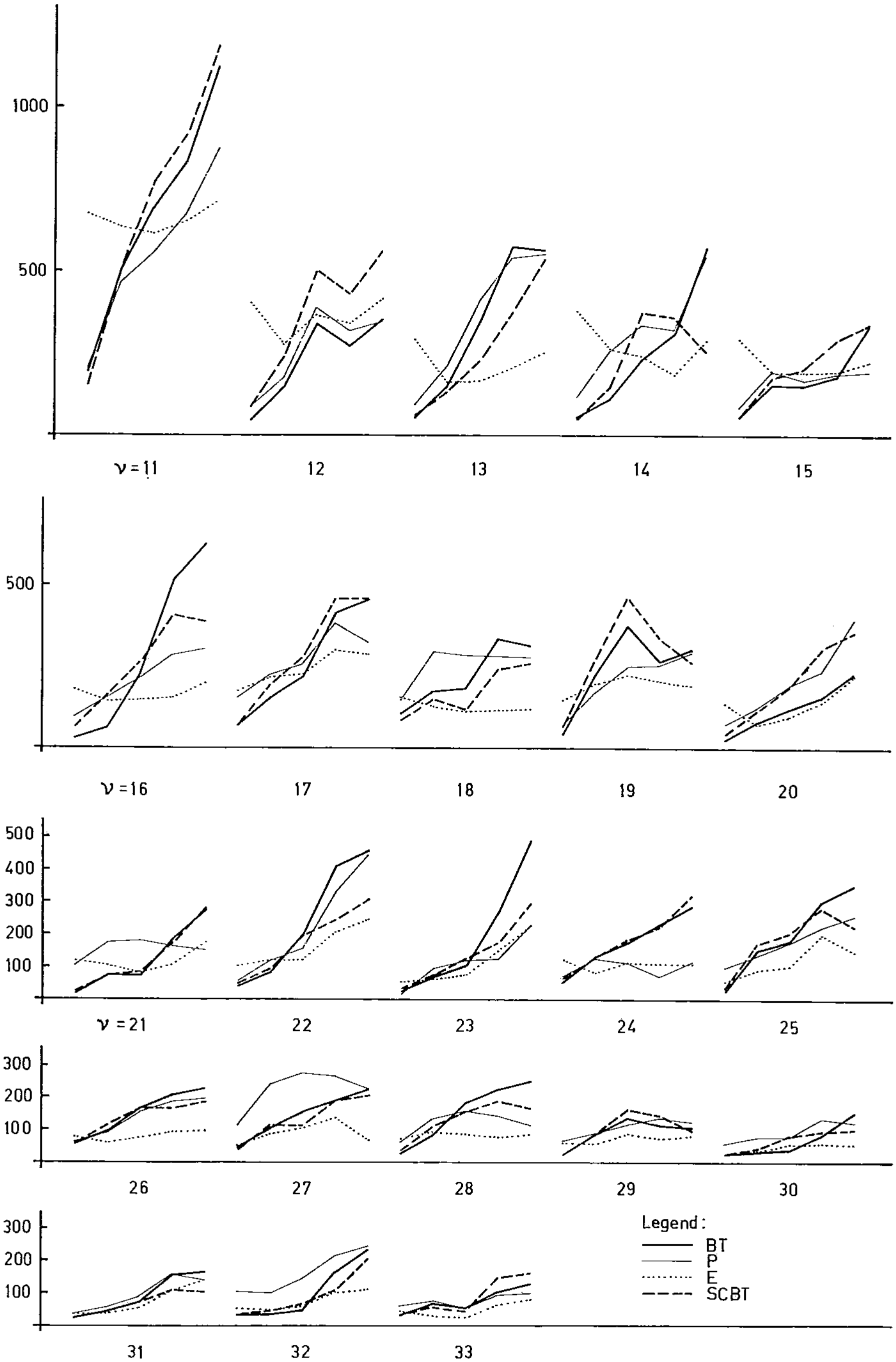


Fig. 1. (continued)

Table 6. The forecasting error of the unknown components or the sum of the error variances associated with the components $\nu > 33$ in the case of the barotropic model, the short-circuited barotropic model, the persistence forecast and the elimination. Unit: gpm².

Model	Forecasting period (hours)	24	48	72	96	120
BT		1053	2868	3895	4402	5333
SCBT		1081	3000	4028	4435	5067
P		2060	4047	4536	4680	4849
E		236	2211	2891	3342	3733

For forecasts of a longer period, the BT model is usually better than the SCBT model. Exceptions are the cases $\nu = 1, 5, 13, 18$. When $\nu > 21$, there is a tendency of the SCBT model to be better than the BT model in the case of the longest forecasts (see also Table 6). Perhaps the introduced meteorological error source, the short-circuiting of the β-term, has been counterbalanced by some other error sources.

In the routine applications, where the BT, P and E models are available, the persistence »model» seems to be rather useless except in the case $\nu = 1$. For instance, in the case of $\nu = 5 / 72$ h the persistence forecast has been the best one, but already in the cases $\nu = 5 / 96$ h and $\nu = 5 / 120$ h the elimination method would have been more useful. This reasoning leads to choosing rules by which a forecast may be rebuilt. As an example, in the case $\nu = 5$ one has to choose the final amplitude according to the BT, BT, P, E, E model in the case of a 24, 48, 72, 96, 120 hour forecast, respectively. A simple rule is found: when $\nu = 1$, then always use the amplitude given by the P model. In the case of 48 hour forecasts, nearly all of the BT amplitudes D_ν, $\nu > 22$, could be removed. Hence one only needs forecast information of the components $\nu \leq 22$. The corresponding truncation points are $\nu \sim 16$ and $\nu \sim 10$ in the 72 hour and 96 hour forecasts. In the case of a 120 hour forecast the rebuilt forecast would simply be

$$\begin{aligned} z''(t, i, j) &= zm(i, j) + D_1^P(t) f_1 (i, j) + D_2^{BT}(t) f_2(i, j) \\ &+ D_3^{BT}(t) f_3(i, j) + D_4^P(t) f_4(i, j) + D_8^{BT}(t) f_8(i, j) \\ &+ D_{10}^P(t) f_{10}(i, j) + D_{12}^P(t) f_{12}(i, j) + D_{15}^P(t) f_{15}(i, j) \\ &+ D_{21}^P(t) f_{21}(i, j) \,. \end{aligned}$$

Thus only the BT amplitudes D_2, D_3 and D_8 should be known; some amplitudes could be determined from the initial conditions (persistence) but most should be eliminated.

The behaviour of the components may be rather characteristic. For instance, the changes in the case $\nu = 10$ have been small when compared with those in the case $\nu = 11$.

Also the prognostic features of the components differ. For instance, the components $\nu = 11$ and $\nu = 18$ were in Part I identified as »boundary error components».

Fig. 2 presents a summarizing verification statistics. The climatological forecast has been defined as $zm\,(i,j) + C_1(t)f_1(i,j)$ and the mean-field-forecast as $zm\,(i,j)$. The corresponding error variances are $\sum_{\nu=2} \overline{C_\nu(t)^2}$ and $\sum_{\nu=1} \overline{C_\nu(t)^2}$. The other variance terms have been computed with the aid of formulae analogous to (3.3). The amplitudes $C_\nu(t)$, $\nu > 33$, are always assumed to be eliminated from the reproduced forecasts, *i.e.*, the variance of the terms $\nu > 33$ has been determined as the residual sum $\sum_{\nu=34}^{\infty} \overline{C_\nu(t)^2}$ (Table 6). The rapid increase in the r.m.s. error of the forecasts is mostly due to the boundary error caused by the limited grid size.

In everyday work, the information included in and transferred by the orthogonal functions should be rather fresh. In other words, the functions and the rules of choice derived from a sample should be applied to forecasts of an adjacent sample. In the present case, perhaps the samples S_3 and S_4, when referred to the basic sample S_0, best correspond to adjacent samples of everyday work. Therefore in the present case one should prefer the 96 h and 120 h, perhaps even 72 h, forecasts when verifying the efficiency of the reproduction method. On the other hand, the 24 h forecasts are too close to S_0. For this reason, the r.m.s. error of the reproduced 24 h forecasts is not shown in Fig. 2.

The statistical parameters required by the reproduction method are not precisely known quantities like regression coefficients. On the contrary, the necessary statistical information is found in the form of rules for choosing between some alternatives. These rules may be rather simple. For instance, in the case of the 48 h forecasts all components with $\nu > 22$ might be neglected, while all others might be determined from the BT forecasts. An exception is the first one which should be taken from the P forecasts. Thus the sample sizes of the statistical procedure can be smaller than in the case of a more complicated regres-

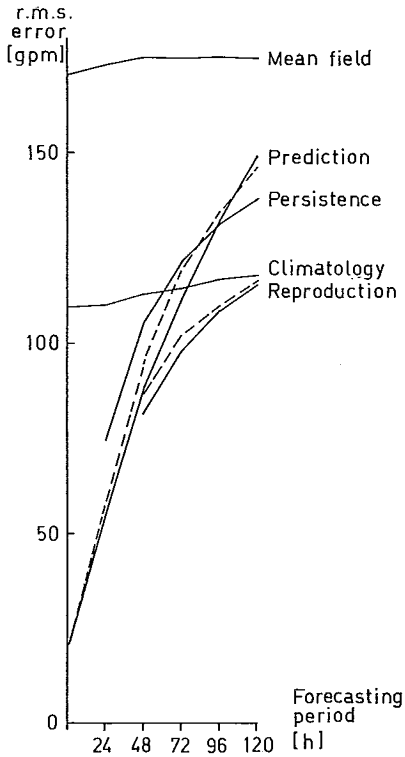


Fig. 2. The root mean square error of the forecasts versus the length of the forecasting period. The forecasts (»prediction») have been prepared by two models, the barotropic model (solid line) and the short-circuited barotropic model (broken line).

sion analysis. Nevertheless, one should keep in mind the smallness of the samples used when making conclusions.

Fig. 2 shows a remarkable decrease in the r.m.s. error (computed, as before, for the whole grid area) caused by the reproduction method, especially in the cases of forecasts of a longer period. An additional experiment might still be needed. Also a meteorologist should express his opinion of the quality of the reproduced forecasts. A statistical test quantity like the r.m.s. error may be too biased.

After the reproduction the SCBT forecasts have still larger values of the r.m.s.e. than the BT forecasts. However, one has to compare not the r.m.s.e. values but the variances. Hence the error of the SCBT forecasts has been reduced more than that of the BT forecasts (except

120 h). The reproduction has partly damped out the influence of the incorrect short-circuiting of the SCBT model.

The amplitudes $D_\nu(t)$ and the variances $\overline{D_\nu(t)^2}$ of the reproduced forecasts are known. Then it is possible to construct probability forecasts

$$\hat{z}(t, i, j) = zm(i, j) + \sum_{\nu}^{33} \hat{D}_\nu(t) f_\nu (i, j) ,$$

where $\hat{D}_\nu(t)$ is a random number distributed normally with the mean $\overline{D_\nu(t)}$ and the standard deviation $\sqrt{\overline{D_\nu(t)^2}}$. The number of such probability forecasts is naturally infinite.

In this approach, the meteorologist does not get a single forecast but a book of forecasts valid at a certain moment on a certain surface. The reproduced forecast will be the basic forecast while the probability forecasts will illustrate possible alternatives. Fig. 3 shows a failed forecast. The reproduction has of course decreased the boundary error. However, there are no marks of a wave over Scandinavia. On the contrary, small oscillations of the initial forecast have been eliminated in the reproduced forecast (REPRO). Obviously the method has interpreted these oscillations to be noise. Therefore it is very interesting that a wave over Scandinavia can be traced in many probability forecasts (Fig. 4). A more pro-

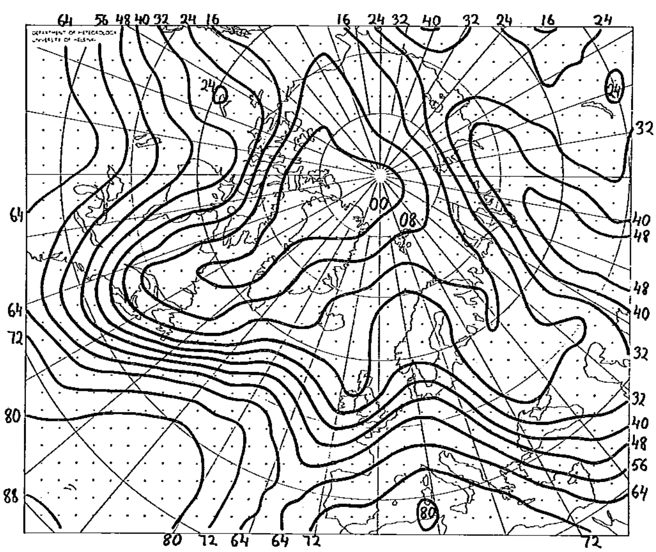


Fig. 3a. The objective analysis, Jan. 17. 1968, 00 GMT (500 mb).

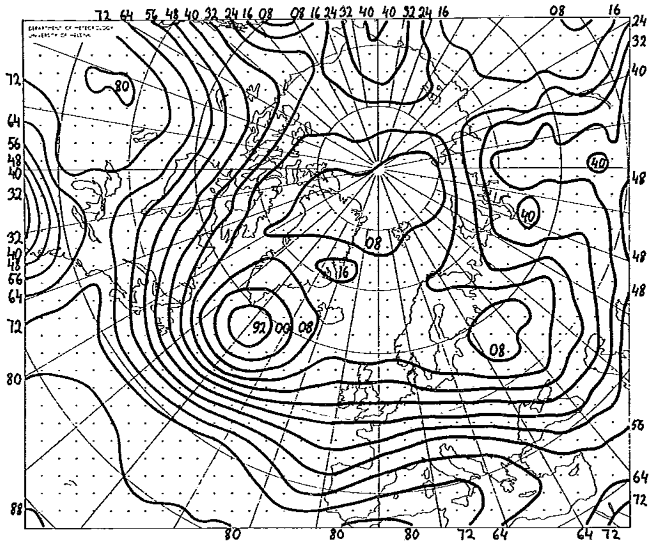


Fig. 3b. The routine 48 hour barotropic forecast (BT), valid on Jan. 17, 1968, 00 GMT.

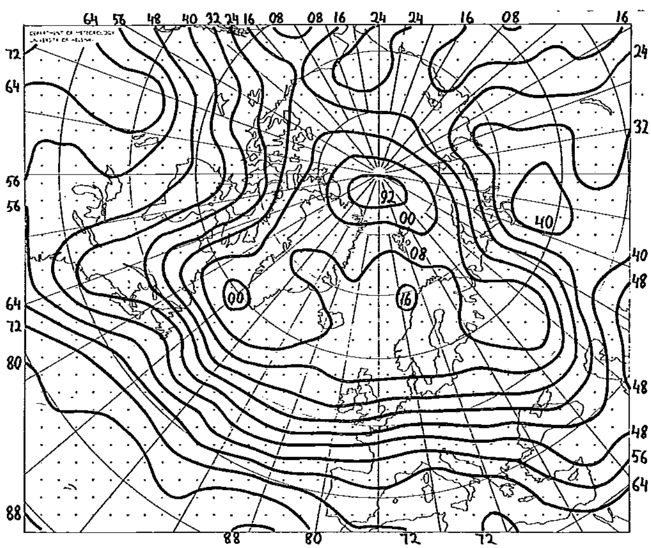
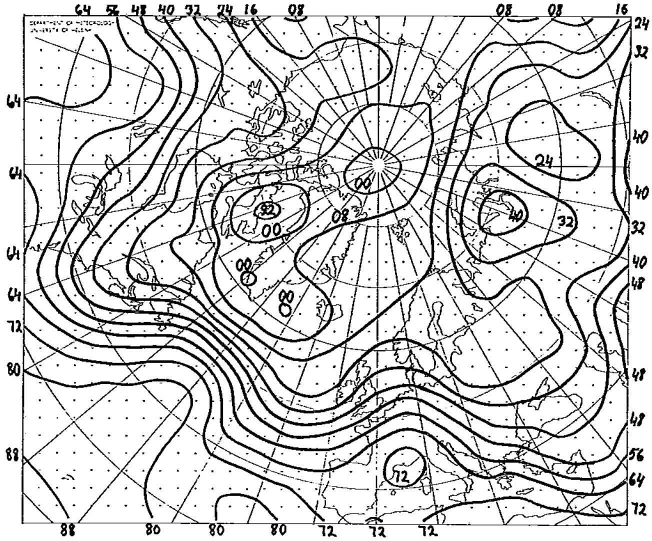
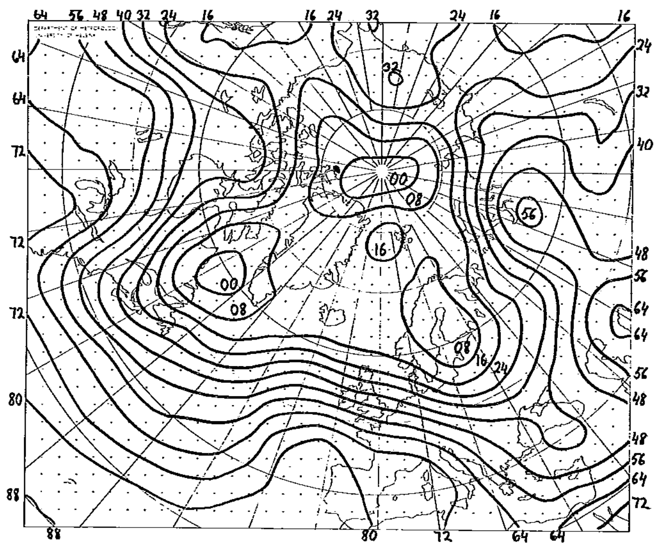


Fig. 3c. The reproduced forecast corresponding to that in Fig. 3b (REPRO).

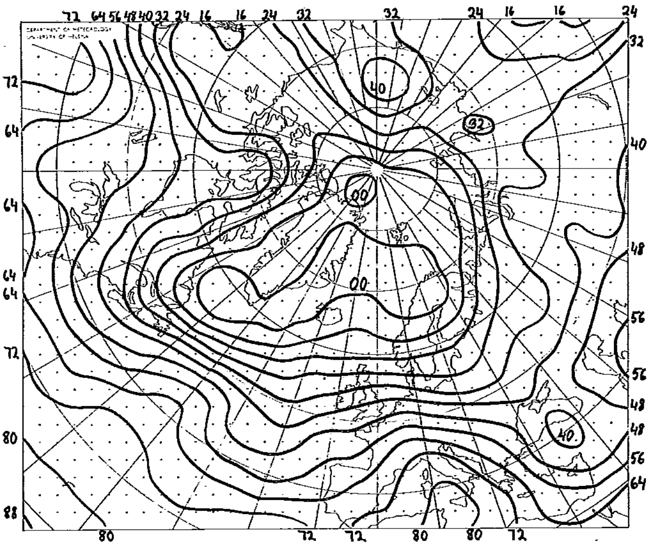
I.


64 56 48 40 32 24 16 08 08 08 16
24
32
40
40
32
40
48
48
48
56
64
72
64
64
64
64
72
80
88
88 80 80 80 72 72 72
24
00
32
00
08
00
00
40
32
72

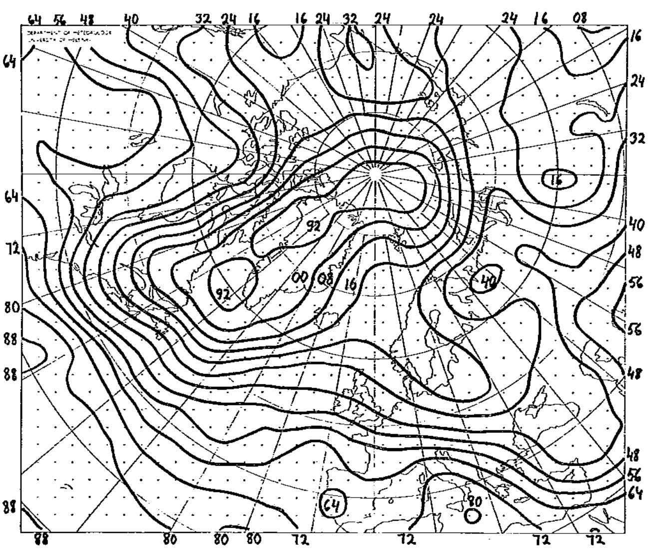
II.


64 56 48 40 32 24 16 16 24 32 24 16 16
24
32
40
48
56
64
64
56
48
48
56
64
72
64
64
72
72
72
80
88
88 80 72 64 64
32
00
08
56
16
00
08
08
16
24

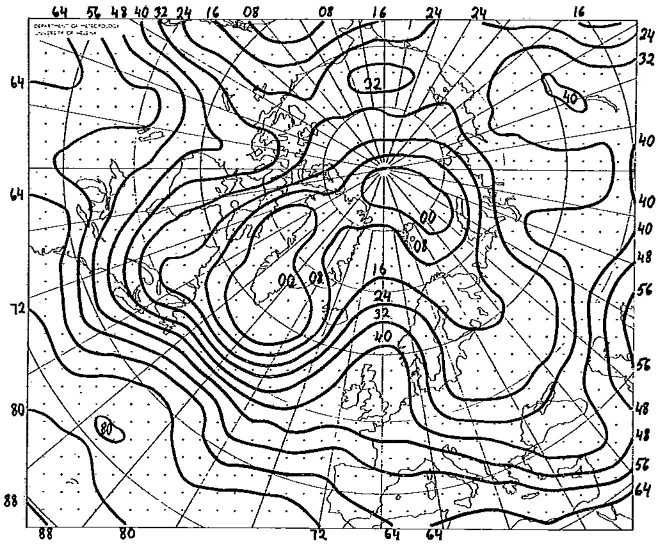
III.



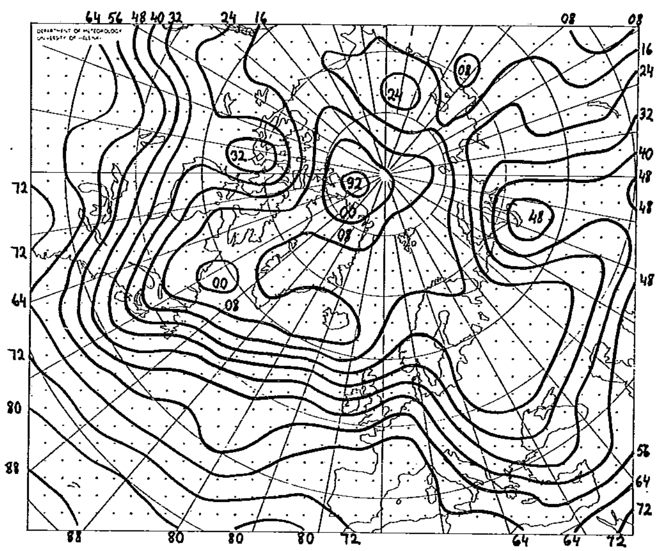
IV.



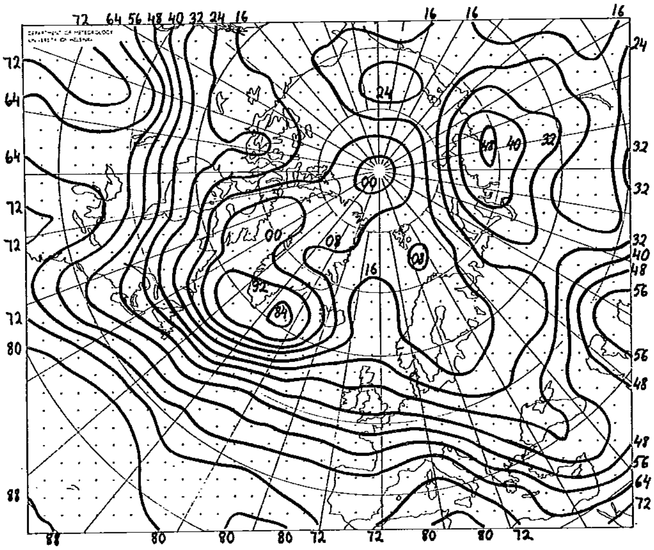
V.



VI.



VII.



VIII.

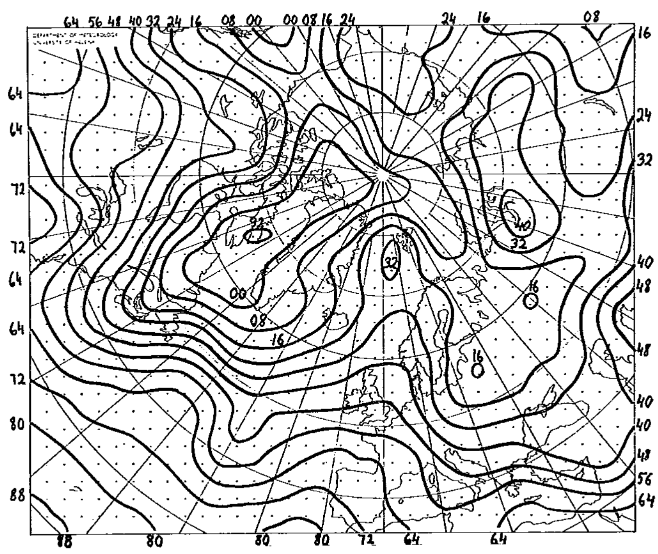


Fig. 4. Eight probability variants of the theme in Fig. 3c.

nounced wave is found in the variants I, V, VIII whereas the variants IV and VI show a smaller wave. Variant II and REPRO have no essential oscillations in the vicinity of Scandinavia. In other words, though the forecast and even the reproduced forecast show no essential marks of a wave in the vicinity of Scandinavia, the meteorologist may conclude that the probability of the existence of such a wave is very large, about 70%.

Is the trough axis to be located west or east of Helsinki? The location is eastward in REPRO and in the variants V, VI, VIII, whereas the variant IV shows a westward location. Hence the probability of the eastward location is rather large whereas the westward case is less probable.

In this connection it is interesting to note that in the 48 hour forecast, corresponding to that in Fig. 3, issued by SMHI (Swedish Meteorological and Hydrological Institute) with a three-parameter model, only a weakly developed ridge over Scandinavia could be recognized whereas a trough was located far to the west of Ireland.

4. The significance of the results

In many respects the sample sizes in this study have been too limited. For instance, the observed difference between the error variances of the barotropic and the short-circuited barotropic forecasts is perhaps not significant because the comparison has been based only on 24 forecasts. The drawing of final conclusions can in some cases be difficult due to the existing systematic tendencies, like the increasing values of »total var.» and the decreasing efficiency in the variance reduction, presented in Table 3.

The efficiency of the reproducing method requires a more detailed analysis.

Theoretically it is clear that the method works. One may expect that the error in forecasts may be minimized by some statistical procedure. As the method of the orthogonal functions in a certain sense is an optimal one, it should decrease the error variance.

It has been shown empirically in this study that the method actually decreases the error. For instance, the elimination of the error associated with the yearly oscillation (Fig. 1, $\nu = 1$) results in a significant change. Similarly, the reduction of the error variance in the case of the 120 h forecasts has obviously been significant.

In Sec. 3 it was pointed out that the statistical parameters required by the method, *i.e.* the rules of choice, may be determined from a rather small sample. This increases the significance of the results.

In connection with the 24 hour forecasts, it was concluded in Part I that the basic sample has been too small to produce eigenfunctions describing all the 500 mb phenomena to be predicted. This difficulty has been avoided in the present work by studying forecasts where most of the lacking 500 mb patterns no longer were predictable and therefore were to be eliminated.

Hence at least some of the observed error reduction has been real and it is possible, with the aid of orthogonal functions determined from small sample, to reduce the error of forecasts adjacent to the basic sample.

There will still remain some question. How much would the error be reduced in the absence of the boundary error? How should a complete set of orthogonal functions work? On what basis a sample can be termed as an »adjacent sample»? Is the achieved error reduction of any practical value?

5. *Conclusions*

The empirical orthogonal functions determined from a *small* sample of 500 mb height analyses, reducing the variance of this basic sample by $\sim$ 99%, have been in some adjacent samples capable of reducing the sample variance by $\sim$ 90%. Even the high-indexed components have been of importance. The deviations in the sample means played a minor role.

The method of reproducing of forecasts derived in Part I has been extended. A reproduced forecast will now include also information from the corresponding persistence forecast in addition to that of the conventional one. This new information seemed not to be of noticeable importance. The method, which could theoretically be expected to be a working one, has been shown empirically to be capable of reducing the forecasting error in the case of a longer forecasting period, in spite of the smallness of the basic sample.

Forecasts of a new kind, probability variants were prepared from the reproduced forecasts. In the case studied, the conventional and the reproduced forecast showed no trace of a wave found in the verification analysis. However, it would have been possible to conclude from the simultaneous probability forecasts that the probability of the presence of such a wave had been very large, about 70%.

Acknowledgements: This study has been supported by the Sohlberg Foundation of the Finnish Society of Sciences.

Appendix

The determination of the approximate wavelengths of the functions $f_\nu(i,j)$.

For a trigonometric wave we have

$$\nabla_a^2 a \sin\frac{2\pi x}{L_x}\sin\frac{2\pi y}{L_y} \approx -\left[\left(\frac{2\pi}{L_x}\right)^2 \delta^2 x + \left(\frac{2\pi}{L_y}\right)^2 \delta^2 y\right] a \sin\frac{2\pi x}{L_x}\sin\frac{2\pi y}{L_y},$$

where the numerical Laplace operator is defined by

$$\nabla^2 h(x,y) = h(x+\delta x, y) + h(x-\delta x, y) + h(x, y+\delta y) + h(x, y-\delta y) - 4h(x,y).$$

Define the mean value

$$f_{m\nu} = \sum_{i,j} f_\nu(i,j)\, dA(i\,j).$$

Then assuming that $f_\nu(i,j) - f_{m\nu}$ resembles a trigonometric wave, it follows that

$$\nabla^2 (f_\nu(i,j) - f_{m\nu}) \approx -\left[\left(\frac{2\pi}{L_x}\right)^2 \delta^2 x + \left(\frac{2\pi}{L_y}\right)^2 \delta^2 y\right] (f_\nu(i,j) - f_{m\nu}).$$

The scale of the functions $f_\nu(i,j)$ seems to be rather uniform so that $L_x \sim L_y$ (*cf.* Part I). It follows that, for the present grid interval $\delta x(i,j)$,

$$\nabla^2 (f_\nu(i,j) - f_{m\nu}) \approx -2\left(\frac{2\pi}{L}\right)^2 (f_\nu(i,j) - f_{m\nu})\, \delta^2 x(i,j)$$

The area elements have been scaled by $\sum_{i,j} dA(i,j) = 1$ (Part I). Hence

$$\delta^2 x(i,j) = A \cdot dA(i,j),$$

where A is the grid area. Further

$$\sum_{i,j}\{[f_\nu(i,j)-f_{m\nu}]\,\nabla^2[f_\nu(i,j)-f_{m\nu}]\}\approx-2\left(\frac{2\pi}{L}\right)^2\cdot A\sum_{i,j}[f_\nu(i,j)-$$
$$-f_{m\nu}]^2\,dA(i,j)^*\,.$$

By computing the sums, the wavelength L may be solved. The »wave number» at 60°N will simply be determined from »wave number» $= 20.000$ km $/ L$.

* *Cf.* footnote in Ch. 3.

REFERENCES

1. Rinne, Juhani, 1970: Investigation of the forecasting error of a simple barotropic model with the aid of empirical orthogonal functions. Part I. This issue of *Geophysica*, 185—213.