# Controlling a large data sample with the aid of EOF's

by

Juhani Rinne

University of Helsinki, Department of Meteorology.

## Controlling a large data sample with the aid of EOF's

Juhani Rinne

When index $\nu$ is small, the coefficients $C_\nu(t)$ of the series

$$z(t,i) = \sum_\nu C_\nu(t)\, f_\nu(i)$$

are slowly varying. If a sudden change can be observed in the time series, then the corresponding analysis most probably is wrong. This fact will in the following be used to identify erroneous analyses.

Unexpected values of $C_\nu(t)$ are easily found by the eye when $\nu$ is small. In the case of a larger $\nu$, the following method is used. The value of $C_\nu(t)$ is forecasted with a linear predictor

$$(1) \qquad \vec{C}_\nu(t) = \sum_{\tau=1}^{n} a_\tau\, C_\nu(t-\tau).$$

In addition to that, a backward forecast is made

$$\overleftarrow{C}_\nu(t) = \sum_{\tau=1}^{n} b_\tau\, C_\nu(t+\tau).$$

The mean value $1/2\ (\vec{C}_\nu + \overleftarrow{C}_\nu)$ is compared to the actual value $C_\nu$. If there is a difference (large enough), a warning will be reported.

The data consists of 6401 (FNWC) analyses (500 mb, height) from years 1962-70. Every analysis has originally been labelled with a number indicating the possible quality of the analysis. There are four classes of this old classification: good, suspicious, obviously bad, inconsistent. In the present work, a new classification will be made: good, suspicious, bad, very bad.

"Very bad" analyses are most often typical products of computing errors, the human eye could never accept such analyses.

Many of the values $C_\nu(t)$, $\nu = 1,2,\ldots$, give strong warnings.

"Bad" analyses include, for instance, cases where a clear analysis error is due to the lack of data.

"Suspicious" analyses include cases where the listed warnings are based on small differences between the forecasted and observed values of $C_\nu(t)$'s. The analysis error may be clear but small or local, there may exist differences between the analysis and the corresponding German (DWD) analysis, or there seem to exist errors but it is difficult to say where and in which one of the successive analyses.

"Good" analyses include cases where no or only one weak warning was reported. Because of practical reasons, single weak warnings could not be synoptically checked though they may indicate errors in the analysis.

| The present classification / The old classification | Good | Suspicious | Bad | Very bad | |
|---|---|---|---|---|---|
| Good | 6237 | 19 | 5 | 2 | 6263 |
| Suspicious | 69 | - | - | 1 | 70 |
| Obvious bad | 4 | - | - | - | 4 |
| Inconsistent | 3 | - | - | 3 | 6 |
| | 6313 | 19 | 5 | 6 | 6343 |

Table 1. The new and old classification of the analyses 1962-70 (Jan, 1962 excluded). The numbers are still subject to changes.

The two groups of warnings do not coincide. From originally inconsistent analysis (6) only 3 were identified, the others

(3) were classified as "good". On the other hand, from the originally "good" analyses, two caused severe warnings and 5 were classified as "bad". Because of this disagreement, all of the extreme cases were studied synoptically. The analyses were compared to the preceding and following ones, also use of German (DWD) analyses was made. The new classification was always reasonable. For instance, in the cases of "Inconsistent (old) / Good (new)" no analysis errors could be found. Thus these analyses probably were not inconsistent.

The material seems to be rather good, because the number of bad analyses is small. However, bad analyses may cause harm in some applications like spectrum analysis or testing of forecasting models (for instance, the model given by eq. (1) ). In this sense the present status of the material is good, because we now know nearly all (if not all) of the bad cases. Note the gigantic size of the material. Checking it manually would be hopeless. Even the few synoptic checks, which were performed, required a large amount of work and time.

Some examples of the bad analyses will be shown.

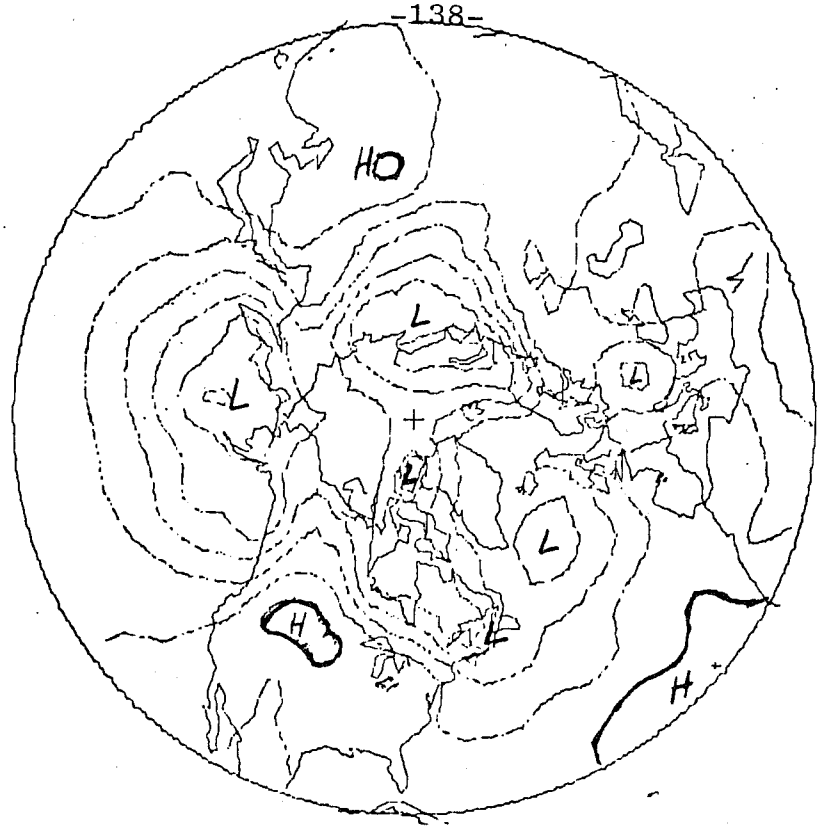


Fig. 1. A very bad case from 1965-06-11:12 (500 mb). Spacing is 80 m, the thick line indicates the height of 5800 m. The analysis is bad except over a dense network of stations. Note that nearly everywhere the height is less than 5800 m though this is a summer case. Perhaps the preliminary field used in the analysis system has been wrong.

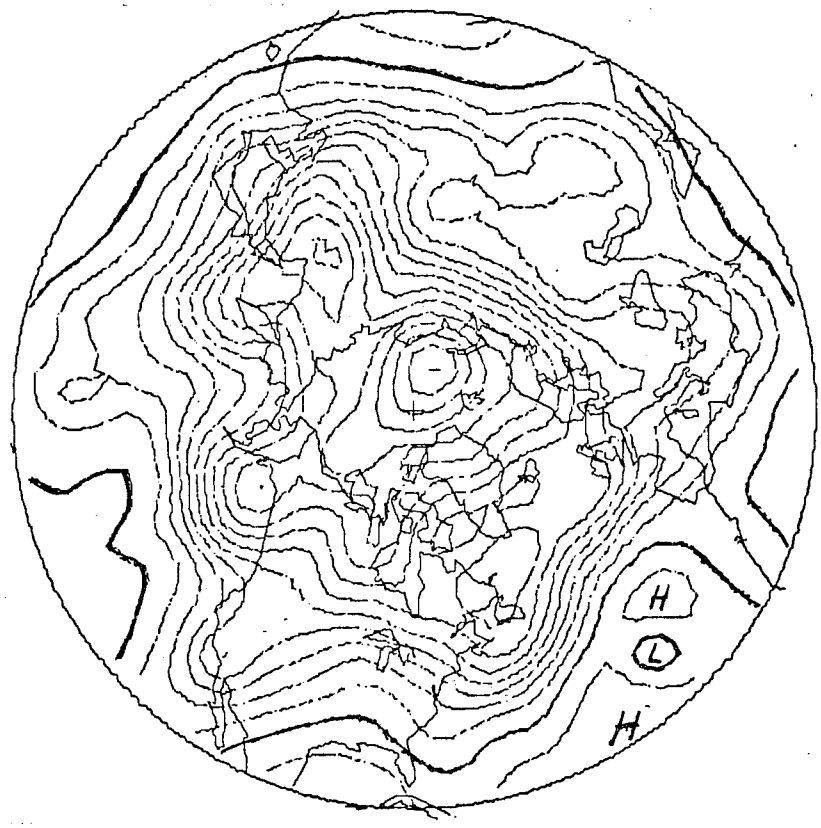


Fig. 2. A suspicious case from 1965-12-23:00. The small low in the middle of the Azores high is absent in the preceding analysis (1965-12-22:12) and in the following analysis (1965-12-23:12). A deeper synoptic study is required in order to decide which one of the analyses is wrong.

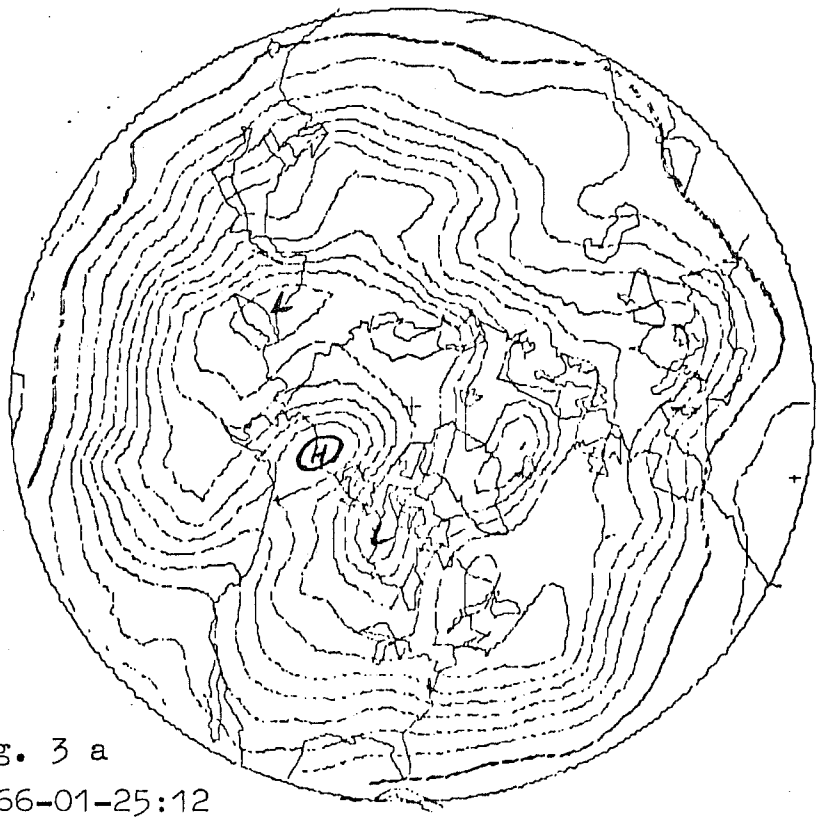


Fig. 3 a
1966-01-25:12

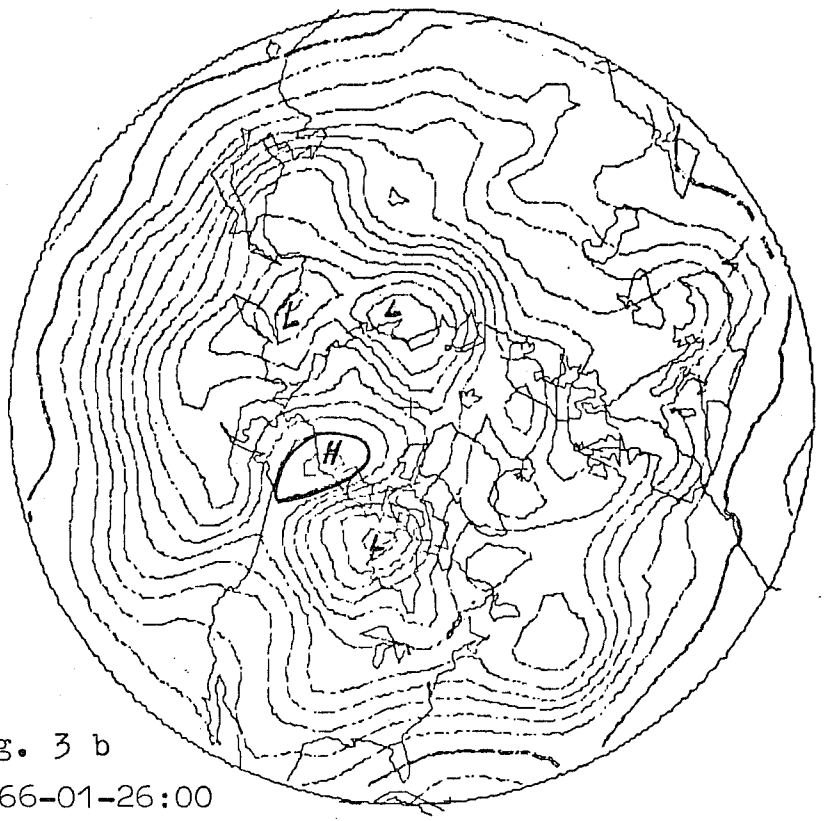


Fig. 3 b
1966-01-26:00

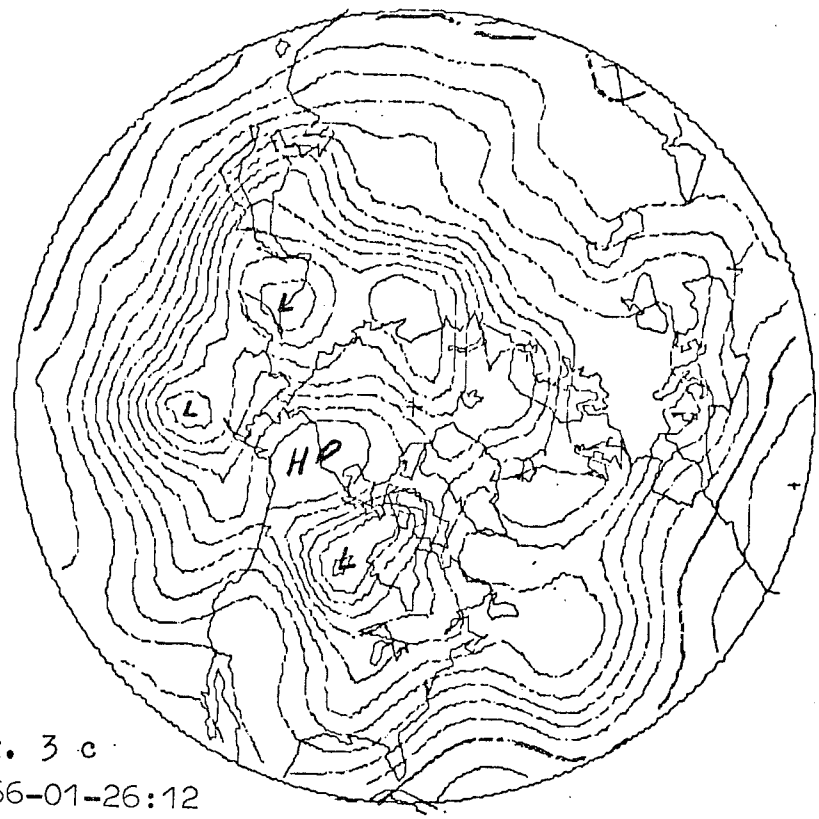


Fig. 3 c
1966-01-26:12

Figs. 3 a,b,c. A bad case from Jan. 1966. The isoline of 5400 m around the high over Alaska is indicated. In Fig. 3 b this high is too strong and the trough over the Pacific too weak. It would be very difficult to find this kind of error in a manual check!