Filtering of multivariate time series in the principal component space

Ludwik Liszka
Swedish Institute of Space Physics
Sörfors 634
S-905 88 Umeå, Sweden

Introduction

Observing a non-stationary physical phenomenon, a time series of a variable is measured. Usually, there are several variables, which must be measured in order to get complete information about the process. These variables are, if the experiment is properly designed, causaly connected. Some variables may be caused by other variables. The measured variables may also be effects (indicators) of an unmeasured variable (or variables). The question is whether the conventional method of handling time series (each one separately), reveals all the properties of the process.
An alternative method to approach the problem is to treat all the observed time series as a single multivariate time series.
There is another aspect of the problem. When looking on multivariate wave phenomena, often submerged in noise, filtering of one variable at a time would lead to information losses. If all variables are processed, for example filtered, at the same time, the intrinsic relations between variables will help to improve the result without distorting it.
The simplest way to process all variables at the same time is to use a suitable transformation technique, for example, transforming all variables into a smaller number of principal components (PC). It corresponds to a linear projection of the multivariate time series onto a specific direction in the multispace. Each principal component is a direction of the maximum variance in the multispace selected in such a way that each two principal components are orthogonal to each other.

Description of the method

A multivariate time series consisting of N variables measured at M equidistant instants forms a matrix **B**.
The next step of the analysis is to perform the principal component analysis (PCA) for the matrix **B**. The results of PCA are:

The vector of eigen values of the matrix, telling how much of the total variance in the matrix may be explained by the consecutive principal components.

The matrix of component score coefficients **a**, a transformation matrix between the old system of N variables and the principal components (the new coordinate system).

The matrix of component scores **S**, with one column for each principal component, being a projection of old N variables upon the new coordinate axis (directions of principal components).

The matrix of component scores **S** is thus the new multivariate time series in the principal component space. It has been found by the present author that, in all cases when filtering was performed in the principal component space, a considerable improvement of signal-to-noise ratio has been obtained without distorting the signal. Each column of **S** is low-pass filtered using a simple filter of moving average type. The result of filtering is matrix $\mathbf{S}_f$.
After filtering an inverse transform:

$$\mathbf{B}_f = \mathbf{S}_f \cdot \mathbf{a}^{-1}$$

is performed resulting in a new version of the matrix **B**.
It is possible to combine the filtering procedure with a decomposition procedure (Liszka, 1997).
If one wants to know what the variations of the N-component vector would be with only one mechanism (or cause), corresponding to the principal component *l* active, it is possible to mask all other columns in $\mathbf{S}_f$, except of column *l* and to perform a calculation of a new matrix $\mathbf{B}_{lf}$:

$$\mathbf{B}_{lf} = \mathbf{S}_f \cdot \mathbf{a}^{-1}$$

The operation may be repeated for each interesting component *l*.
The present technique will be illustrated by two examples.

Processing the multisensor data

An example, which may visualise the problem of processing of multisensor data is a temperature measurement with a number of termistors located on a mast at different heights above the ground. A conventional approach would be to treat the temperature T as a function of time (t) and height (h): T(t, h). However, when looking at non-linear phenomena, as gravity waves breaking at the ground surface, it may be more convenient to study a multivariate time series $T_i(t)$, where T_i is the temperature time series measured at the termistor *i*. It will be shown in the present report that the filtering in the principal component space will significantly increase the ability to detect wave phenomena in the data.

The data analysis technique described in the previous chapter will be applied to temperature measurements using 5 termistors located on a vertical mast, at 0, 1, 2, 3 and 4 meters height. The measurements are made for comparison with infrasonic data collected at a 3-microphone array, in order to study the influence of the acoustic gravity waves on the propagation of infrasound with frequencies 0.5 - 5 Hz. As the temperature data are rather noisy, the present technique helps to find events when acoustic gravity waves are observed.
The measurements are made at a rate of about two measurements per minute. An example of temperature recording during November 30 - December 2, 1995 is shown in Fig. 1. A wave event may be seen around the sample # 2800. Unfiltered projections of the temperature time series onto the principal component axis for the same sequence are shown in Fig. 2.

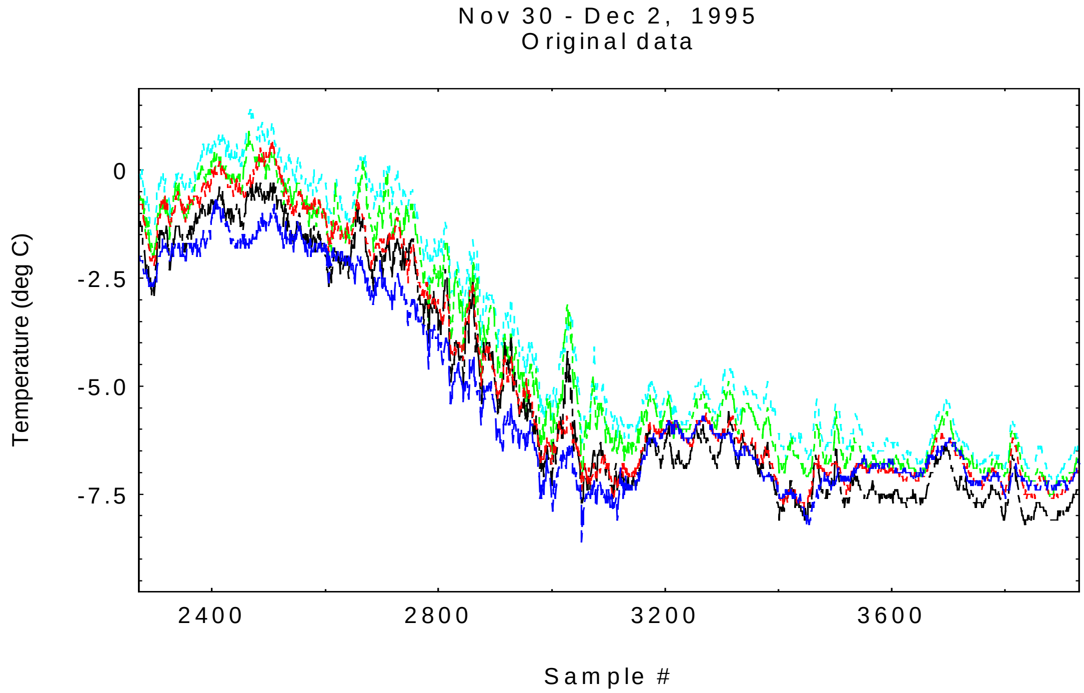

Fig. 1. A part of temperature measurements during November 30 - December 2, 1995 by 5 termistors at 0, 1, 2, 3 and 4 meters height.

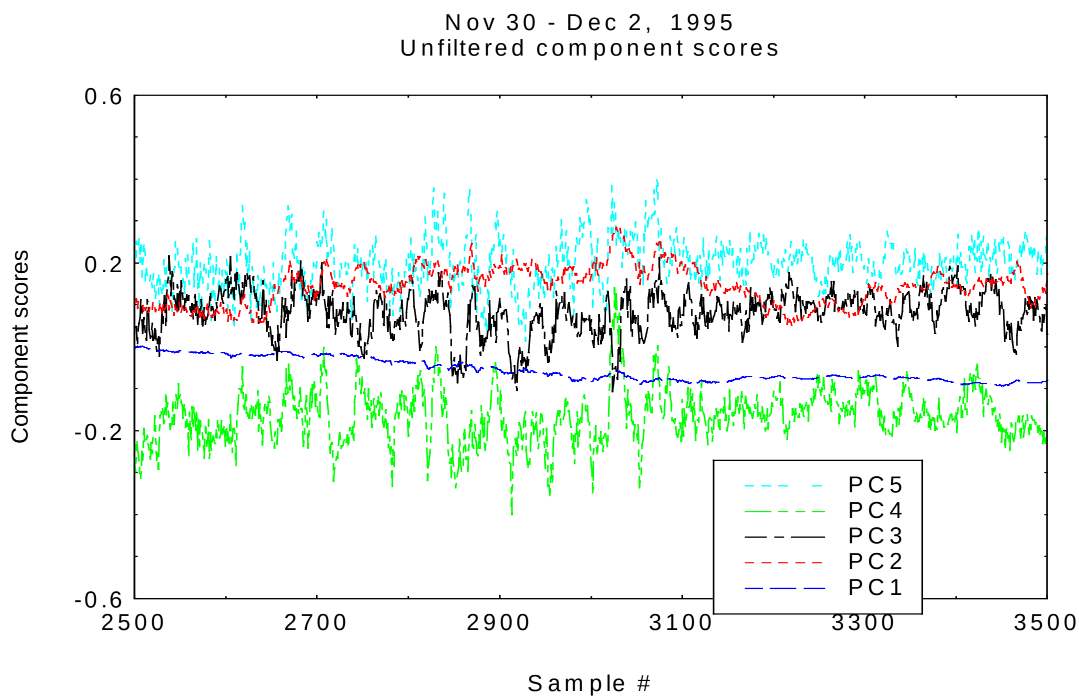

Fig. 2. Unfiltered component scores for the same sequence of temperature measurements as shown in Fig. 1.

Contours of constant temperature for original data are shown in Fig. 3. The noise level is so high, that the wave event is hardly seen. The event is clearly visible in Fig. 4, which is constructed using the temperature data of Fig. 5 computed from the filtered component scores.

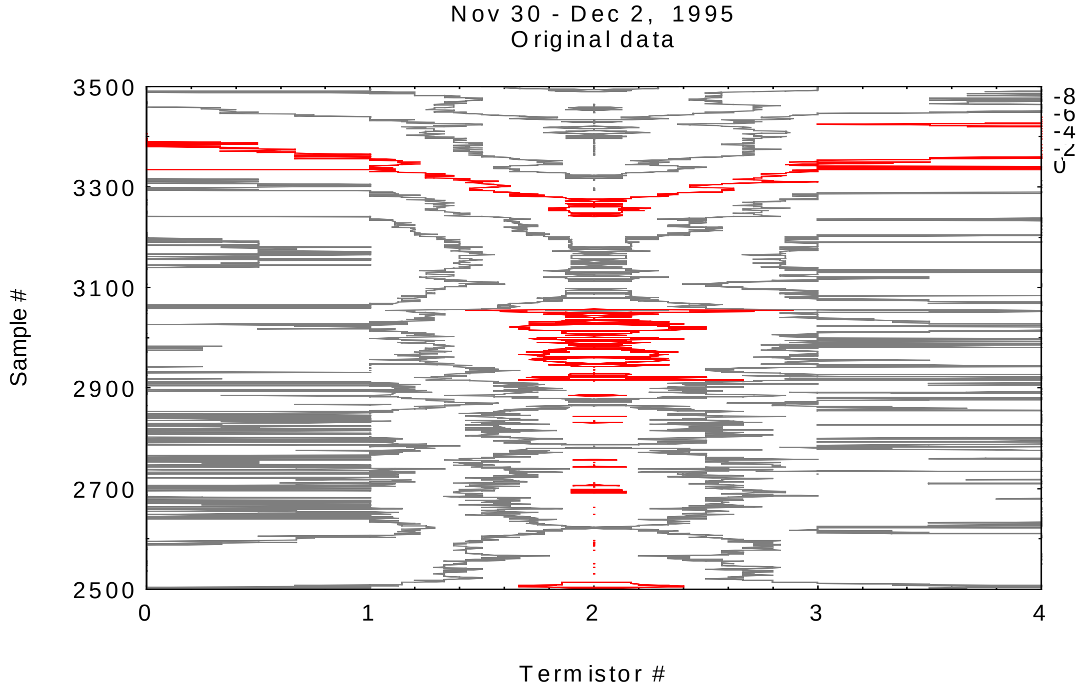

Fig. 3. Contours of constant temperature constructed from original data for the same sequence as that shown in Fig. 1.

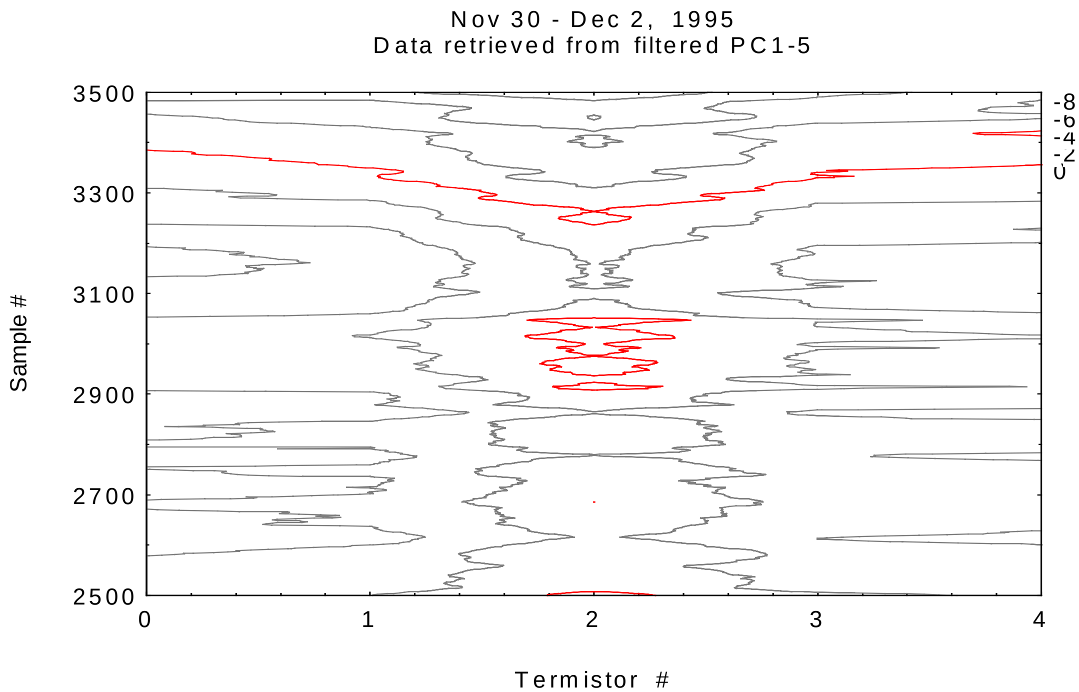

Fig. 4. Contours of constant temperature constructed from temperatures computed from filtered component scores for the same sequence as that shown in Fig. 1

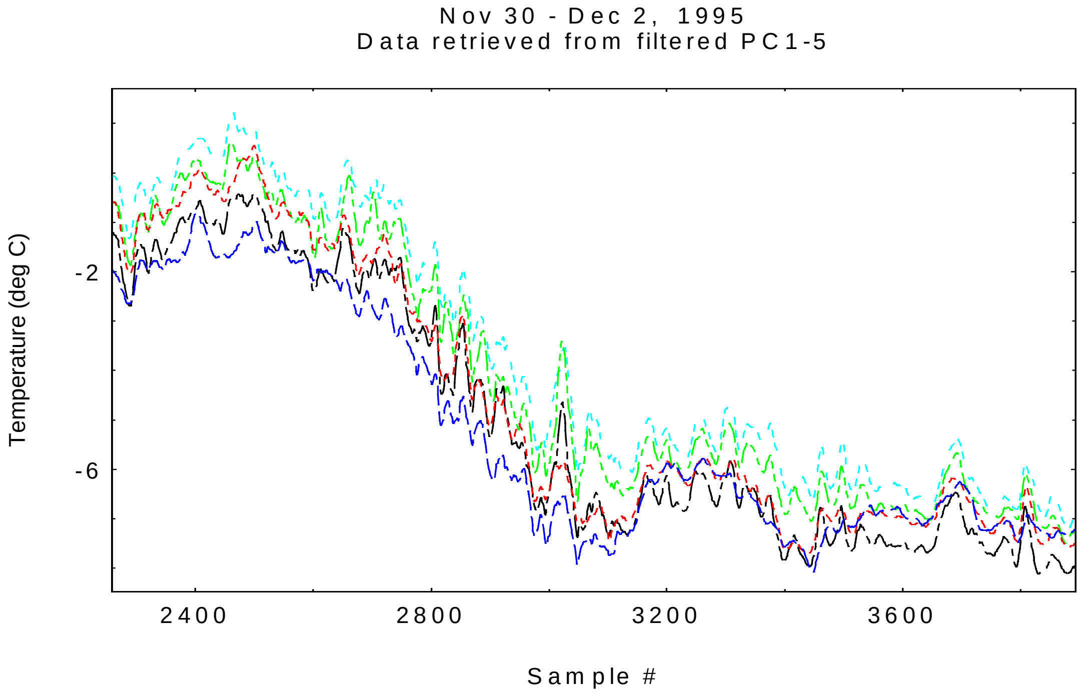

Fig. 5. Temperatures reconstructed from filtered component scores for the same sequence as that shown in Fig. 1.

Filtering of frequency spectra

The present method may be used for example for processing long sequences of frequency spectra. Filtering in the principal component space will enhance spectral components showing a consistent temporal pattern of variance. That application is shown for a frequency spectrum of variations in the geomagnetic field (X-component) sampled at 10 seconds intervals during a minor magnetic disturbance. The time series in Fig. 6 starts on March 5, 1996 at 1630UT.

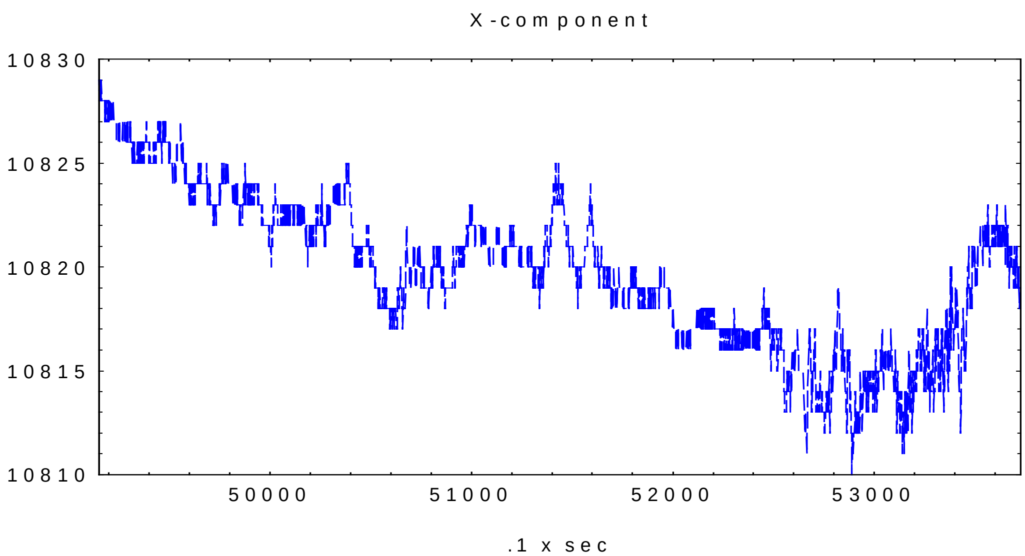

Fig. 6. Time series of the geomagnetic X-component, length 14 hours, starting on March 5, 1996 at 1630UT.

In order to study the temporal structure of the time series in Fig. 6 the short time Fourier transform (STFT) has been used. A window of 512 samples (5120 seconds) was moved in 8 samples (80 seconds) steps. The result of analysis is shown in Fig. 7.

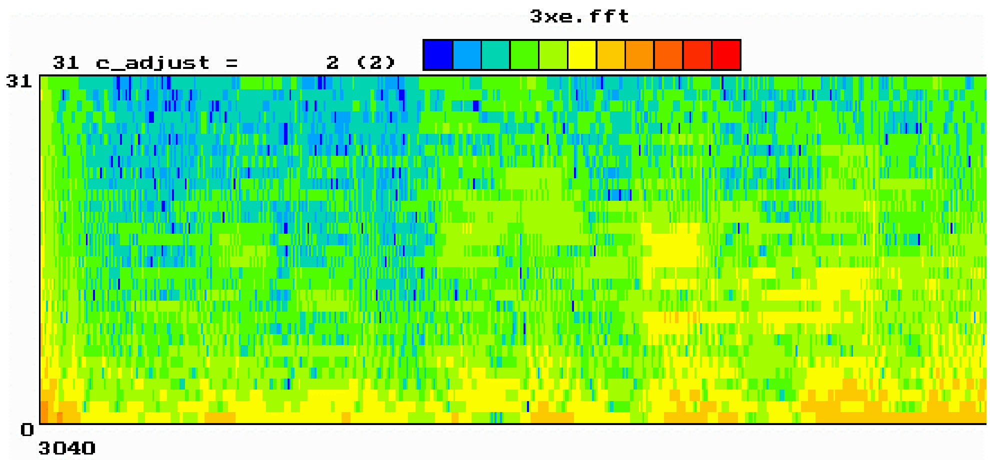

Fig. 7. Frequency spectrum (STFT) obtained from time series of Fig 6. Maximum frequency is 6.25×10^{-4} Hz.

The regular structures visible in the low frequency part of the spectrum are due to the limited measuring accuracy (round-off effects). It corresponds to the step-by-step amplitude variations visible in Fig. 6.
The spectral matrix corresponding to Fig. 7 (32 columns x 600 rows) was further processed using the principal component method. The spectrum is decomposed into:
1.A part computed from PC1 explaining 60% of the total variance in the spectral matrix (shown in Fig.8).
2. A part computed simultaneously from PC2 - PC9 explaining altogether 20% of the variance (shown in Fig. 9).

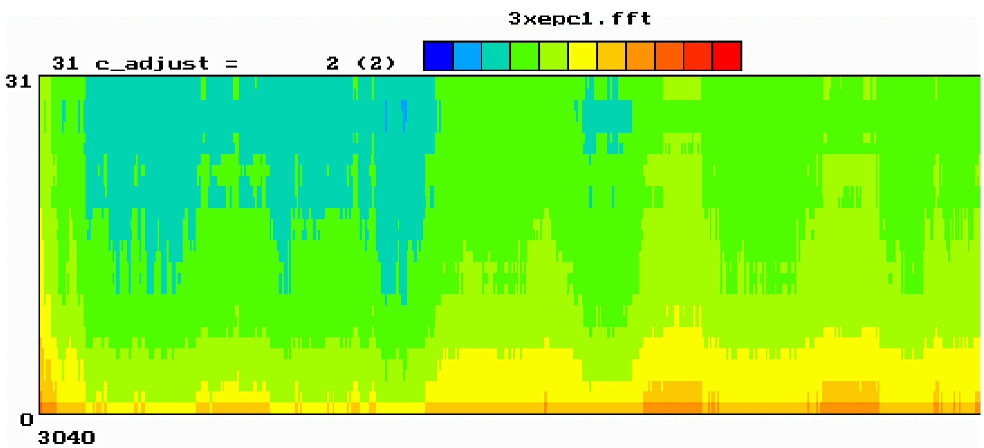

Fig. 8. A component of frequency spectrum computed from PC1. Maximum frequency is 6.25×10^{-4} Hz.

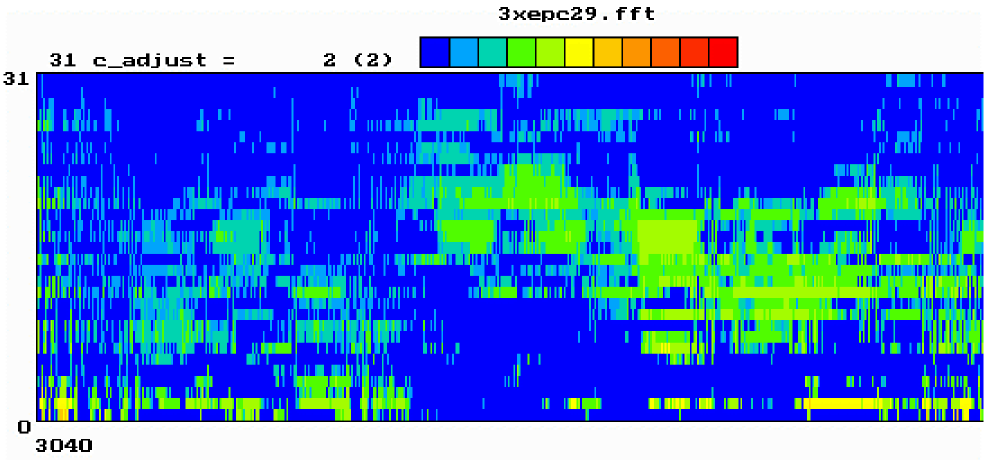

Fig. 9. A component of frequency spectrum computed from PC2 - PC9 corresponding to higher order components of the spectrum. Maximum frequency is 6.25 x 10^{-4} Hz.

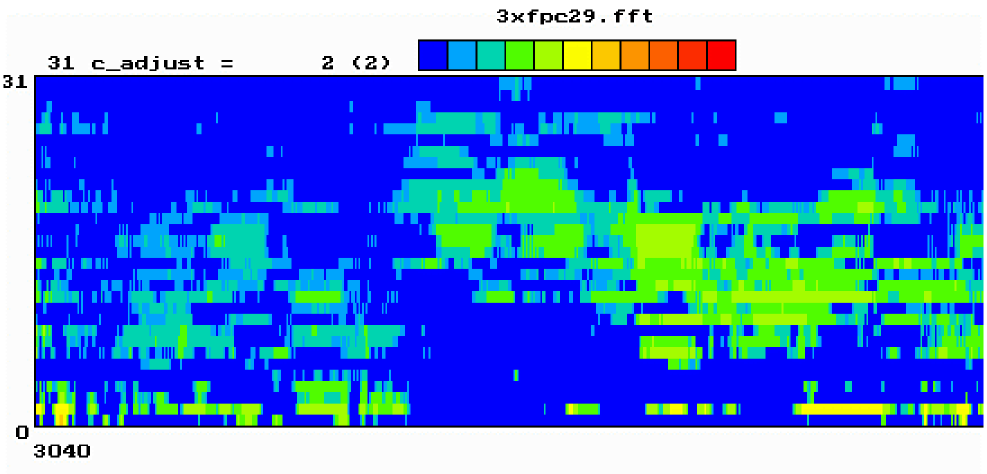

Fig. 10. A component of frequency spectrum computed from PC2 - PC9 using low-pass filtered component scores. Maximum frequency is 6.25 x 10^{-4} Hz.

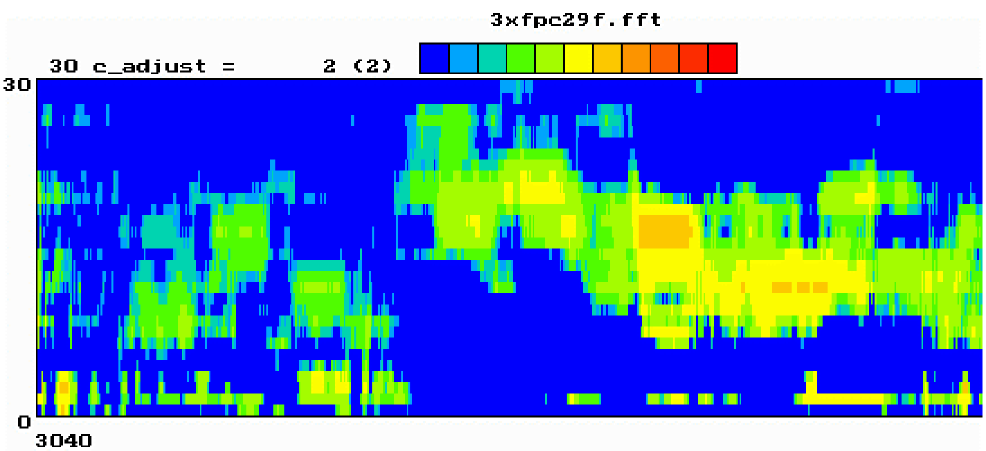

Fig. 11. The same high order spectral components as shown in Fig. 10. The spectrum has been enhanced using a conventional image enhancement technique. Maximum frequency is 6.25 x 10^{-4} Hz.

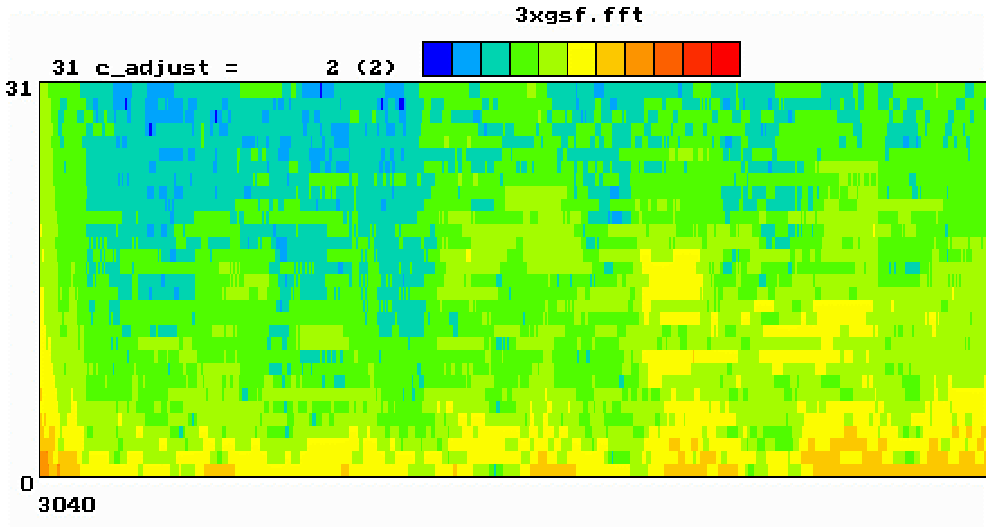

Fig. 12. The original frequency spectrum where all 32 component scores were low-pass filtered and then inverse-transformed back to the frequency domain. Maximum frequency is 6.25×10^{-4} Hz.

When higher order components of the spectrum, shown in Fig. 9, are calculated using the low-pass filtered component scores, the result shown in Fig. 10 is obtained. It is obvious that the morphology of the higher order components is easier seen on the spectrum derived from filtered component scores. Using any of commercial image processing software's, the image of Fig. 10 may be enhanced. The result is shown in Fig. 11.

Low-pass filtering of all 32 component scores corresponding to all principal gives, after an inverse transform, the spectrum shown in Fig. 12. It may be seen that the structures corresponding to "round-off" errors are now absent in the spectrum.

High order components are not visible in the original spectrum. The low-pass filtering in the principal component space improve possibilities to study those components. In the case of the analysed frequency spectrum of geomagnetic variations the high order components carry information about the substorm morphology. The filtering and the subsequent decomposition technique may thus be an useful tool to study complex geophysical processes.

Reference

Liszka, L.: Decomposition of particle energy spectra using the principal component method. Scientific Report No. , Swedish Institute of Space Physics, 1997.