Dimension Reduction for Time Series from Intensive Care

Vivian Lanius and Ursula Gather

Department of Statistics, University of Dortmund

Abstract

A modified version of principal component analysis (PCA) for time series is investigated. The approach is in the frequency domain as in Brillinger (1975). Available knowledge on the subject matter can be incorporated via rotational methods. This eases the interpretation of the obtained component series. An application to hemodynamic data from intensive care yields clinically meaningful component series of low dimension. We describe the results from this application and compare them with those obtained from a standard PCA.

1 Introduction

In situations where many variables are simultaneously measured on the same sampling item repeatedly over time, classical monitoring tools from statistical process control must often contend with both high auto– and crosscorrelations in the data. To cope with this flood of observations, we try to reduce firstly the number of monitored time series to a few component series that contain the essential information. The extracted signals can then be processed by an appropriate process control procedure.

Principal component analysis (PCA, Pearson, 1901, Hotelling, 1933) has proven to be a helpful tool for dimension reduction. In statistical process control, this method is often applied in spite of the correlations at various time–lags. Brillinger (1975) suggests a PCA for time series that takes such autocorrelations into account.

In this paper, we are particularly interested in the online monitoring of hemodynamic data acquired from a Clinical Information System (CIS) on the surgical intensive care unit at the Klinikum Dortmund. Because it is the main aim to develop an alarm system there is a strong need for statistical methods which provide the online detection of clinically relevant patterns such as trends, level–shifts and outliers.

Therefore we consider Brillinger's PCA for time series and examine some modifications of that procedure and its application to hemodynamic time series data. The paper is organized as follows. In Section 2 we briefly review the main ideas of standard PCA and introduce the PCA for time series as suggested by Brillinger (1975). Section 3 deals with some complications when performing dynamic PCA. We give a solution that guarantees the extraction of meaningful component series. In Section 4 we apply the method to observations from intensive care and compare the results to those obtained from a classical PCA. We conclude with some suggestions for further research.

2 Dynamic principal component analysis

Principal component analysis is a common statistical technique for analyzing multivariate data. Depending on the application, it can be motivated from different points of view. With the objective of data reduction, a geometrically intuitive explanation of the procedure is the following. Starting from a sample of n independent observations $\boldsymbol{x}_i \in \mathbb{R}^k, i = 1, \ldots, n$, of a random variable $\boldsymbol{X} \in \mathbb{R}^k$, we search for the best approximation of the data points which is located in an r–dimensional hyperplane, $r < k$, in the sense of minimizing the squared Euclidean distances. Looked at the population level, PCA aims at finding the r–dimensional subspace that contains the best linear r–dimensional predictor of $\boldsymbol{X}$. Thus, we have to solve the minimization problem

$$\min_{\boldsymbol{\mu} \in \mathbb{R}^k, \mathbf{A} \in \mathbb{R}^{k \times k}, \mathrm{rk}(\mathbf{A}) = r} \mathsf{E}[(\boldsymbol{X} - \boldsymbol{\mu} - \mathbf{A}\boldsymbol{X})'(\boldsymbol{X} - \boldsymbol{\mu} - \mathbf{A}\boldsymbol{X})], \tag{1}$$

where $r < k$. With $\mathsf{E}(\boldsymbol{X}) = \boldsymbol{\mu}_X$ and $\mathsf{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma}_X$, a solution of (1) is given by $\mathbf{A} = \mathbf{L}\mathbf{L}'$ and $\boldsymbol{\mu} = \boldsymbol{\mu}_X - \mathbf{L}\mathbf{L}'\boldsymbol{\mu}_X$, where $\mathbf{L} = [\boldsymbol{l}_1, \ldots, \boldsymbol{l}_r]$ and $\boldsymbol{l}_1, \ldots, \boldsymbol{l}_k$ are the eigenvectors of $\boldsymbol{\Sigma}_X$ corresponding to the ordered eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k \geq 0$. As a normalization constraint, we will impose $\boldsymbol{l}_j'\boldsymbol{l}_j = 1$, $j = 1, \ldots, k$. The linear combination $\boldsymbol{\xi}_j = \boldsymbol{l}_j'\boldsymbol{X}$, $j = 1, \ldots, k$, is called the j–th principal component (PC) of $\boldsymbol{X}$. It has the appealing property $\mathsf{Var}[\boldsymbol{\xi}_j] = \lambda_j$ and $\mathsf{Cov}[\boldsymbol{\xi}_i, \boldsymbol{\xi}_j] = 0$, $i \neq j$. PCA is often introduced as a method of finding those linear combinations of the observed variables that successively capture a maximal amount of variance under the constraint of uncorrelatedness. Provided that all eigenvalues λ_j, $j = 1, \ldots, k$ are distinct, the principal component transformation is uniquely determined up to multiplications of the eigenvectors by the factor -1. Further details and results can be found in Jolliffe (2002) and Flury (1988).

The standard PCA approach exploits the covariance structure of the k random variables measured on n independent sampling units. However, when dealing with multivariate stochastic processes $\{\boldsymbol{X}(t), t \in \mathbb{Z}\}$, this classical procedure for dimension reduction should be improved because lagged dependencies between components of the random vector, as well as instantaneous ones, can provide important and useful information.

One approach to adapt PCA techniques to the time series situation is singular spectrum analysis (SSA, Danilov, 1997, Goljandina et al., 2001). SSA was originally designed for univariate time series and requires decomposing the so–called trajectory matrix obtained from the initial time series by the method of delays. A couple

of rather specialized approaches from the analysis of multivariate spatio–temporal meteorological data are Multichannel SSA (a multivariate extension to SSA), principal oscillation pattern analysis and Hilbert empirical orthogonal functions (for references compare Jolliffe, 2002). However, the main aim of these methods is not dimension reduction but rather the detection of common (oscillation) patterns.
We consider an earlier approach by Brillinger (1975). It provides a PCA in the frequency domain with the objective of approximating a stationary multivariate time series by a filtered version of itself, where the filter has reduced rank.
For simplicity, let us assume that the k–dimensional stationary time series $\boldsymbol{X}(t) = (X_1(t), \ldots, X_k(t))'$, with $\mathsf{E}[\boldsymbol{X}(t)] = \mathbf{0}$, possesses an absolutely summable autocovariance function $\mathsf{Cov}[\boldsymbol{X}(t), \boldsymbol{X}(t-s)] = \boldsymbol{\Gamma}_X(s)$, $s \in \mathbb{Z}$. While the standard PCA defines PCs as linear combinations of the observations corresponding to the same sampling item, Brillinger suggests to form PC series by a sum of linear combinations

$$\boldsymbol{\xi}(t) = \sum_s \mathbf{b}(t-s)\boldsymbol{X}(s), \quad t = 0, \pm 1, \pm 2, \ldots, \tag{2}$$

such that the approximation error series

$$\boldsymbol{\epsilon}(t) = \boldsymbol{X}(t) - \sum_s \mathbf{c}(t-s)\boldsymbol{\xi}(s), \quad t = 0, \pm 1, \pm 2, \ldots \tag{3}$$

is small. Here, $\{\mathbf{b}(s),\, s \in \mathbb{Z}\}$ and $\{\mathbf{c}(s),\, s \in \mathbb{Z}\}$ are filters with $r \times k$– and $k \times r$–dimensional matrices. Note, that the PC series defined in (2) are sums of linear combinations of past, present and *future* observations of the time series, since $\{\mathbf{b}(s),\, s \in \mathbb{Z}\}$ is not required to be causal. Minimizing the expected approximation error as in (1), affords to solve

$$\min_{\{\mathbf{b}(s)\},\, \{\mathbf{c}(s)\}} \mathsf{E}\left[\left(\boldsymbol{X}(t) - \sum_s \mathbf{c}(t-s)\boldsymbol{\xi}(s)\right)^{*\prime}\left(\boldsymbol{X}(t) - \sum_s \mathbf{c}(t-s)\boldsymbol{\xi}(s)\right)\right]. \tag{4}$$

This yields the filters

$$\begin{aligned} \mathbf{b}(s) &= (2\pi)^{-1}\int_0^{2\pi} \mathbf{B}(\alpha)\exp(is\alpha)\, d\alpha, \\ \mathbf{c}(s) &= (2\pi)^{-1}\int_0^{2\pi} \mathbf{C}(\alpha)\exp(is\alpha)\, d\alpha, \end{aligned} \tag{5}$$

where

$$\mathbf{B}(\alpha) = \begin{bmatrix} \boldsymbol{l}_1(\alpha)^{*\prime} \\ \vdots \\ \boldsymbol{l}_r(\alpha)^{*\prime} \end{bmatrix}, \quad \mathbf{C}(\alpha) = [\boldsymbol{l}_1(\alpha), \ldots, \boldsymbol{l}_r(\alpha)] = \mathbf{B}(\alpha)^{*\prime}.$$

Here $\boldsymbol{l}_j(\alpha)$ denotes the j–th normalized eigenvector of the spectral density matrix $\boldsymbol{f}(\alpha)$, $\alpha \in [0, 2\pi]$, corresponding to the j–th eigenvalue, where $\lambda_1(\alpha) \geq \lambda_2(\alpha) \geq$

$\ldots \geq \lambda_k(\alpha) \geq 0$, and q^* is the complex conjugate of $q \in \mathbb{C}$. Because the spectral decomposition is performed at all frequencies, we also speak of dynamic eigenvectors and dynamic eigenvalues. The minimum of (4) is given by

$$\mathsf{E}[\boldsymbol{\epsilon}(t)'\boldsymbol{\epsilon}(t)] = \int_0^{2\pi} \sum_{j=r+1}^{k} \lambda_j(\alpha)\, d\alpha.$$

Applying the filter from (5) to the time series yields the principal component series $\boldsymbol{\xi}(t)$ with spectral density $\mathrm{diag}(\lambda_1(\alpha), \ldots, \lambda_r(\alpha))$, $\alpha \in [0, 2\pi]$, i.e., the principal components are uncorrelated at all time lags. A frequency domain PCA is equivalent to a classical PCA carried out at each individual frequency.

In practice, the filters $\{\mathbf{b}(s)\}$ and $\{\mathbf{c}(s)\}$ have to be estimated from a stretch of time series observations $\boldsymbol{x}(t)$, $t = 1, \ldots, T$. A consistent estimator of the spectral density matrix can be obtained by smoothing the periodogramm matrices. Eigenvalue–eigenvector decompositions of the estimated spectral density matrices at each Fourier frequency yield estimates of $\lambda(\alpha)$ and $\mathbf{B}(\alpha)$. A more detailed description of the procedure and of its asymptotic properties can be found in Brillinger (1975).

3 Complications and possible remedies

As Brillinger (1975, Section 9.5) points out, some problems may arise when performing a PCA in the frequency domain. First of all, each of the (standardized) *complex* eigenvectors of the spectral density matrices is only defined up to multiplication by the factor $v_j(\alpha)$, where $v_j(\alpha) \in \mathbb{C}$ with modulus $|v_j(\alpha)| = 1$. This fact complicates the situation severely because, at each frequency, each of the eigenvectors can be rotated within the complex space.
PC series are defined as sums of linear combinations of *all* measured variables such that an interpretation can be difficult. However, often the correlation patterns indicate groups of variables, including possibly single variable groups, whose within–group correlations are high while the between–group correlations are low. It is desirable that each of the r component series primarily represents the compressed information provided by solely those variables of one of these groups. Due to numerical difficulties, the algorithm used for the spectral decomposition (the statistical software package R 1.5.1 provides an interface to the EISPACK routines) may produce eigenvectors that are rotated in different, even opposite directions at adjacent Fourier frequencies (despite the corresponding spectral density matrices having nearly identical eigenvectors). We will give an example of this situation in Section 4. As the resulting principal component filters depend on eigenvectors at each frequency, there is consequently a large number of possible filters for the principal component transformation. This does not cause any problem if one is interested

only in the best approximant series $\widetilde{\boldsymbol{X}}(t) = \sum_s \mathbf{c}(t-s)\boldsymbol{\xi}(s)$ or functions thereof. The choice of the filter is however relevant for the construction of the PC series. Problems arise if we wish to extract a few latent components and then interpret them in terms of the subject matter or context.

Moreover, in the matrices $\mathbf{B}(\alpha)$, the position of the eigenvectors with a similar orientation may shift due to a change in size of the corresponding eigenvalues, e.g. for $\alpha_1 \neq \alpha_2$, we may get $\mathbf{B}(\alpha_1) = (\boldsymbol{l}_1(\alpha_1), \boldsymbol{l}_2(\alpha_1), \ldots, \boldsymbol{l}_r(\alpha_r))$ and $\mathbf{B}(\alpha_2) = (\boldsymbol{l}_2(\alpha_1), \boldsymbol{l}_1(\alpha_1), \ldots, \boldsymbol{l}_r(\alpha_r))$. This has an effect on the estimated filters. The PC series from (2) represent a filtered version of the observed series $\boldsymbol{x}(t)$ such that the components successively capture a maximal amount of variance at *each* frequency. If the order of similarly directed eigenvectors changes across different frequency bands, the PC series do not necessarily represent groups of a few closely related variables. In this situation difficulties arise with respect to the interpretation of the resulting PC series. To our knowledge, this aspect has not been discussed in the literature so far. Brillinger (1975)'s example focusses on an analysis of the estimated dynamic eigenvalues and the gain and the phase of the entries in the two first dynamic eigenvectors. From this it is concluded that the first principal component series is essentially proportional to an average of the observed time series, but estimates of the PC series are not given. Shumway and Stoffer (2000) also apply PCA in the frequency domain but concentrate on the dynamic eigenvalues and the analysis of certain frequencies. A generalized dynamic factor model for the analysis of multivariate time series is considered by Forni et al. (2000). Their estimation procedure is based on the concept suggested by Brillinger. However, as their objective is to separate common and idiosyncratic effects of the observed series, they focus on the construction of the approximant series. Our main concern is the construction of a few interpretable PC series, which can subsequently be monitored in order to detect patterns such as level–shifts and trends. The advantage of meaningful PC series in statistical process control is that the source of detected deviations from the process under control can be traced more easily.

In classical factor analysis, interpretable factors are often derived by rotating the first retained factors. A factor analytic model in the time series context is implicitly presumed of the form

$$\boldsymbol{X}(t) = \boldsymbol{\mu} + \sum_s \mathbf{c}(t-s)\boldsymbol{\xi}(s) + \boldsymbol{\epsilon}(t).$$

Such dynamic factor models, with various further assumptions are considered by Geweke (1977), Peña and Box (1987), Stock and Watson (1988), Molenaar et al. (1992), Forni et al. (2000), among others. Here we are simply interested in a linear transformation of the time series based on principal component methods instead of imposing model assumptions. Therefore we will merely borrow the idea of rotating the extracted components in an attempt to achieve a simple structure.

There are several possible criteria according to which a rotation can be performed. We will concentrate on orthogonal transformations such that the rotation reduces to the choice of new coordinate axes of the PC subspace. Common rotation criteria

like varimax or orthomax operate automatically but do not necessarily solve all the problems stated above. As the rotation has to be carried out at each frequency, such "blind" rotation criteria can yield loading matrices with similarly directed but differently ordered column vectors. Even worse, the same criterion can result in different directions at distinct frequencies.
A possible remedy is to exploit background knowledge on the subject matter. Rotations according to the Procrustes criterion can then be employed where a target matrix with an ideal pattern of the loading matrix $\mathbf{L}$ is specified in advance. The eigenvectors are forced as much as possible into default directions that reflect certain dependencies among the variables. This proceeding at each frequency facilitates the interpretation of the resulting PC series.
By the usage of rotational methods, we can gain meaningful PC series on the one hand. On the other hand, we lose one appealing property: the uncorrelatedness of the PC series. As one can easily see by inspecting the power spectrum of $\boldsymbol{\xi}(t)$, the coherency is not necessarily equal to zero anymore for all frequencies. Jolliffe (1995) discusses such drawbacks associated with rotation subject to the normalization constraints.

A further problem is the non–causality of the filters that determine the principal component transformation. There are future observations available in a retrospective analysis, but, in our context of online statistical process control, there is a need for procedures that work in real–time without great time delay. Another approach regarding the non–causality of the filter is pursued by Molenaar (1987) who exploits a result by Robinson and Silvia (1978). They show that within each set of stable filters having the same gain, there exists a unique filter which has minimum phase–lag at all frequencies and which is causal. As this approach may lead to components, which are not easy to interpret, it will not be considered here further.

4 Application to time series from intensive care

In a retrospective analysis 100 multivariate time series representing the hemodynamic system of a patient in intensive care were analyzed by standard PCA and PCA in the frequency domain. Each of these data sets consists of observations of ten variables measured every minute: arterial diastolic pressure (APD), arterial systolic pressure (APS), arterial mean pressure (APM), central venous pressure (CVP), pulmonary arterial diastolic pressure (PAPD), pulmonary arterial systolic pressure (PAPS), pulmonary arterial mean pressure (PAPM), heart rate (HR), pulse (pulse), and blood temperature (TEMP). The length of each of these time series is between 2500 and 7000 observations with possibly a few missing values being replaced by linear interpolation.
The methods as described in Section 2 are not robust against outliers. In order to reduce the effect of measurement errors that can frequently be found in observations

of physiological variables we replaced the outliers as follows. They are detected for each variable individually by means of an online univariate outlier detection procedure (Fried, 2002). According to suggestions by Davies et al. (2002), the repeated median is used to fit a local linear trend to the observations in a moving time window of width 31 first. The robust scale functional $\tilde{\sigma}_{QN}$ suggested by Rousseeuw and Croux (1993) is employed to compute a robust local scale estimate from the residuals of the robust regression. An incoming observation is identified as an outlier and replaced by the extrapolation of the local linear trend if the absolute value of its residual exceeds three times the local scale estimate. This online procedure is well adapted for a robust filtering of noisy physiological time series with trends, level shifts and time–varying variability. A few outliers are not detected in situations where gross errors occur in combination with missing values. Here the linear interpolation produces short trends or level shifts caused by outlying observations. Subsequent to the outlier replacement, the time series were individually standardized in order to accommodate for different scales among the measured variables.
A reason to be cautious when applying the PCA in the frequency domain to physiological time series is their non–stationarity. The stationarity assumption is needed for statistical estimation and inference. We are, however, mainly concerned with an explorative analysis of the intensive care data and therefore restrict ourselves to a purely descriptive examination. A finite sample version of the minimization problem (4) and a geometrical justification serve as a basis for our approach, see the Appendix. In practice we smoothed the periodogram slightly by means of a modified Daniell–window of width less than one percent of the length of the time series in order to reduce some of the variability caused by noise. We took care not to produce a too large bias.

Because a physician is mostly interested in a few meaningful PC series, rotations according to the Procrustes criteria were performed at each frequency. The required target matrix was chosen on the basis of background knowledge gained in preliminary data analyses. Gather, Imhoff, Fried (2002) and Imhoff, Fried, Gather (2002) use graphical models (Lauritzen and Wermuth, 1989) for time series as suggested by Dahlhaus (2000) to analyze the linear dependencies among the hemodynamic variables in the intensive care situation. They find a clear pattern of strong dependencies between the arterial blood pressures (APD, APM, APS), between the intrathoracic pressures (PAPD, PAPM, PAPS, and CVP), and between heart rate and pulse. As these findings well agree with medical experience, we chose a target matrix that forces the eigenvectors to represent the three groups above and in addition one further component essentially representing the blood temperature. This choice automatically fixes the number of retained PC series to four components. Usually for choosing the number of PCs, criteria based on the eigenvalues of the covariance matrix are employed to ensure that a certain amount of variability in the data is explained. Here we draw on previous analyses of the hemodynamic variables that show that retaining four PC series yields satisfactory results (Fried et al., 2002).

As can be seen in Figure 1, this remains true in the frequency domain. We show the

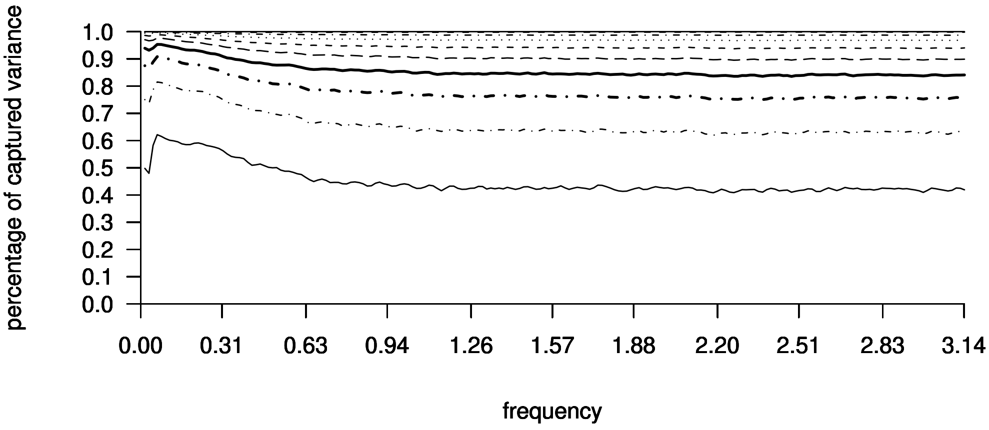

Figure 1: Mean cumulative percentage of variability successively captured by the first r, $r = 1, \ldots, 10$, principal components from an analysis of 100 patients

cumulative percentage of variability accounted for by the first r, $r = 1 \leq r \leq 10$, PC series at each frequency, that is the measure $\sum_{j=1}^{r} \lambda_j(\alpha_i) / \sum_{j=1}^{10} \lambda_j(\alpha_i)$, $\alpha_i = \pi i/200$, $i = 1, \ldots, 200$, averaged over the corresponding curves of the 100 patients. We find that by retaining 4 PC series we account on average for at least 83.6% of the variability at each frequency. As further components do not contribute considerable amounts to the cumulative percentage of variance anymore, we can justify the decision to retain only four components.

If we inspect the dynamic eigenvectors for all patients and all Fourier frequencies we notice that before rotation the first PC series is generally an average of the ten observed variables, with distinct weights of the variables for the individual patients. Generally, the imaginary entries are relatively small compared to the real entries for most frequencies. An example of this is provided in Figure 2, where we depict the progression of the entries of the first dynamic eigenvector of one patient for a period of 100 Fourier frequencies. We find considerable high absolute real entries for the variables HR and pulse as well as the arterial blood pressures. With the exception of two frequencies, the dynamic eigenvectors are strikingly stable across the 100 adjacent Fourier frequencies. However, for two single frequencies the Eispack routines compute the first eigenvector as showing in nearly the opposite direction. This is one of the difficulties pointed out in Section 3. For the second and further dynamic eigenvectors, we find more of such undesirable features along with a great instability of the dynamic eigenvectors. The directions of the eigenvectors can be very

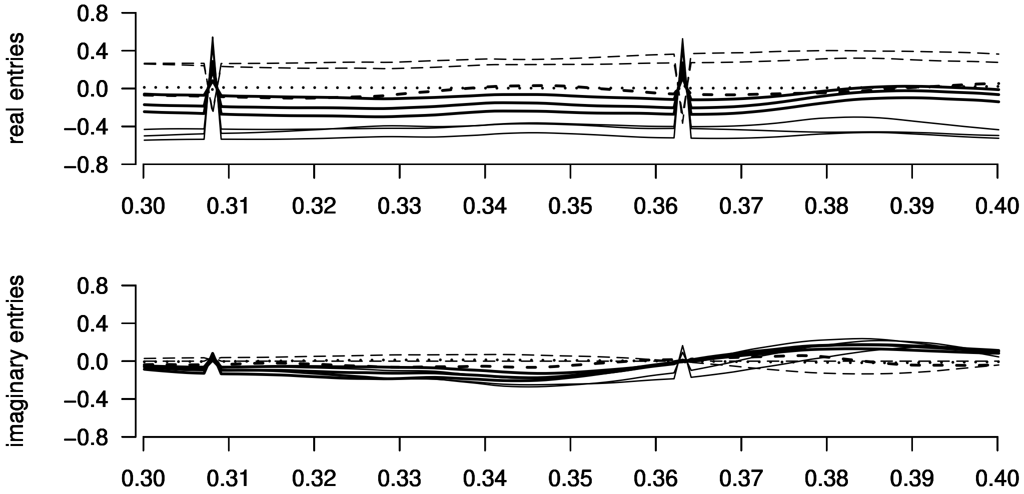

Figure 2: Real and imaginary parts of the entries of the 10–dimensional complex eigenvectors without rotation over a range of 100 Fourier frequencies for one patient, HR and pulse (*dashed*), arterial blood pressures (*solid*), pulmonary artery pressures (*bold solid*), CVP (*bold dashed*), TEMP (*dotted*)

distinct across the various frequencies. This may primarily be due to the very small amount we used for smoothing the periodograms. Although the high variability is still remarkable and we are far from recognizing common dynamic eigenvectors for all Fourier frequencies, the rotations according to the Procrustes criterium clearly improve the situation. In particular, the first three rotated dynamic eigenvectors have relatively high positive entries for the groups of variables found by means of graphical models respectively.

A crucial point in our analysis is to get an idea of the form of the PC filter in the time domain. First, we illustrate the impact of the past, present and future observations on the principal component series $\boldsymbol{\xi}(t)$ at each time point t by plotting the Euclidean norm of the four row vectors in the filter matrices $\{\mathbf{b}(s)\}$ for lags $s = -20, \ldots, 20$. Further lags can be neglected as the filter matrices, and therefore the vector lengths as well, are almost zero. In Figure 3 we show the minimum, quartiles, median and maximum of the vector norms for the PC filters of the 100 patients. It can be seen that the PC series $\boldsymbol{\xi}(t)$ is mainly a linear combination of the observations at time point t. Observations at further lags contribute very little information. As opposed to the first three PC series, the loadings of the fourth PC series at time lag zero have only a rather low impact.
In this regard we are interested in the size of the entries of the filter matrix at lag zero, especially in view of their enhanced interpretability. Figure 4 gives box–plots

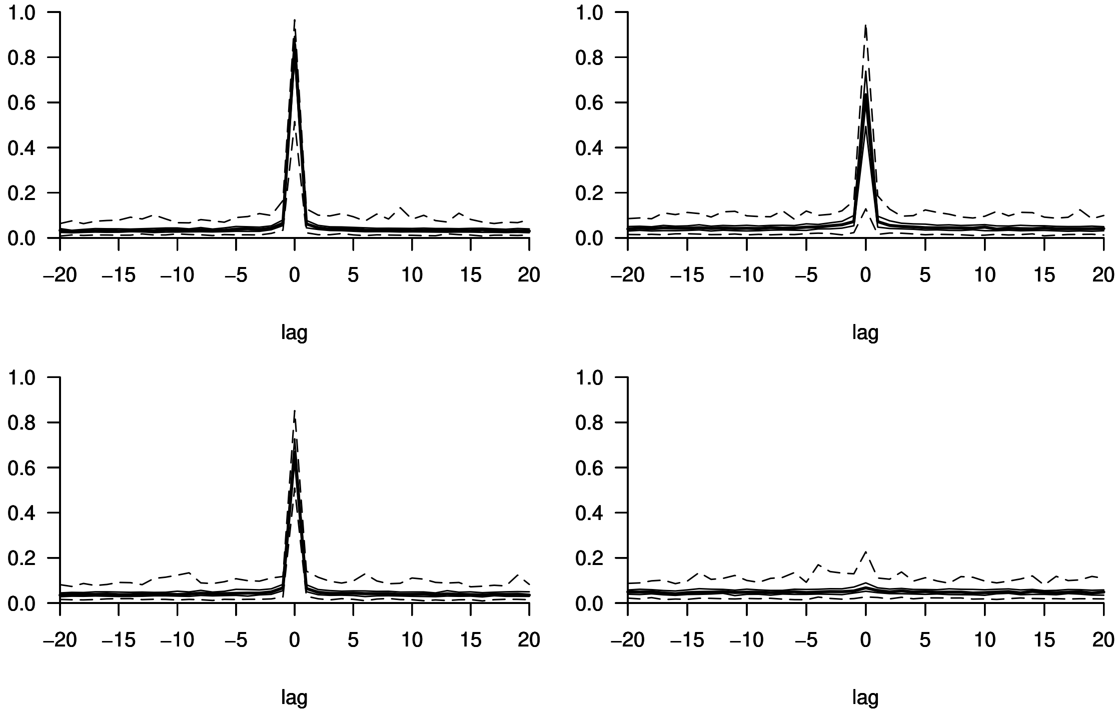

Figure 3: Five–number summaries of the Euclidean norms of the four loading vectors of the principal component filter matrices for lags $-20, \ldots, 0, \ldots, 20$

of the loadings of the four PC series at time lag zero. As anticipated and due to the Procrustes rotations in the frequency domain, inspection of the loadings of the first three PC series shows high positive entries on the groups of variables specified in the target matrix and entries close to zero for the other variables. The loadings of the fourth PC series are rather small, where additionally the entry for the variable temperature has been forced to be positive. An inspection of this loading vector for each of the 100 filter matrices shows that this linear combination can often be interpreted as representing the difference between arterial blood pressures, intrathoracic pressures or heart rate and pulse, instead of representing the temperature depending on the patient.

Similar representations for the loading matrices at lags close to but different from zero still show noticeably higher absolute loadings on the specified groups of variables. However, these loadings can be both positive and negative with no clear pattern being recognizable. Altogether, the filter matrix at lag zero is dominating for the construction of the PC series.

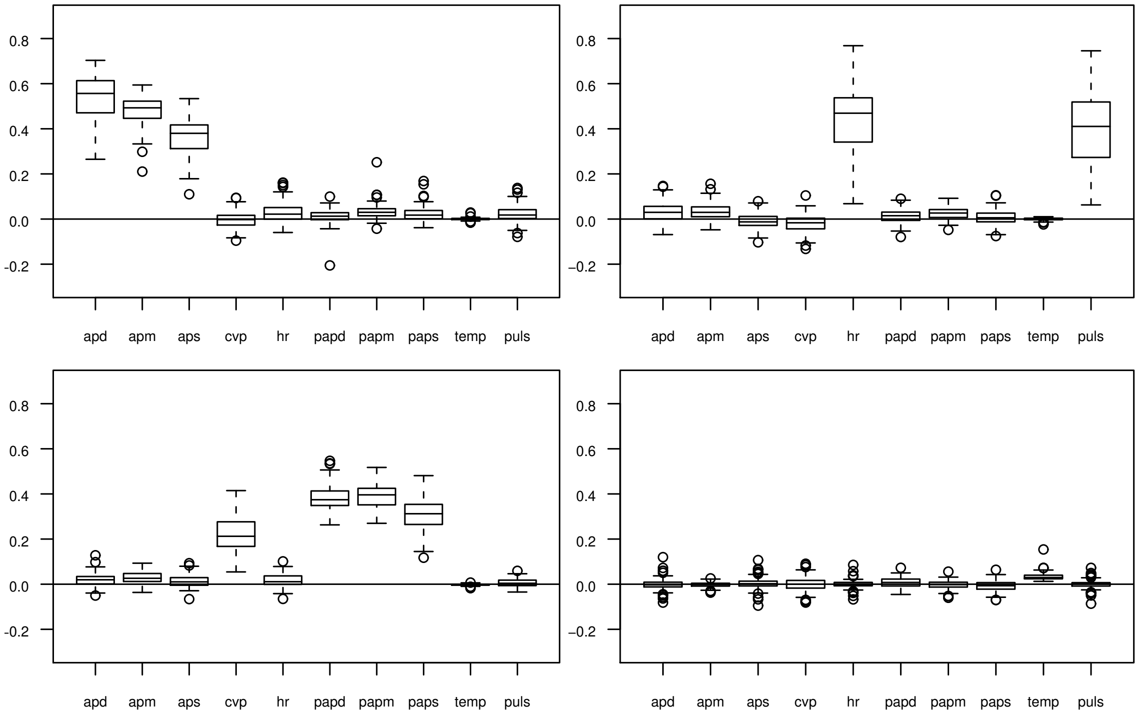

Figure 4: Box–Plots of the real entries in the four loading vectors of the filter matrix at lag zero for 100 patients after Procrustes–rotation at each Fourier frequency

From these findings it can be conjectured that the first three of the resulting PC series after rotation can readily be interpreted in terms of the arterial blood pressures, heart rate and pulse, and the intrathoracic pressures, respectively, while the fourth PC series does not have the same meaning for each of the patients. In Figure 5 we display a sequence of the observed time series and the extracted PC series with and without rotation for one patient. The series have been shifted by constants for purposes of illustration. Obviously the first PC series without rotation captures most of the variability and most of the structure contained in the observed variables. The other PC series are more stable and do not provide that much information. In contrast, the first three PC series after rotation are readily interpretable.

With regard to the fact that the loadings at lags different from zero are very small, it is of interest to compare the performance of the PCA for time series and the classical PCA with respect to the approximation error. Therefore we computed the sample total error variances, i.e., the trace of the sample error covariance matrices $\mathrm{tr}\,(\sum_{t=1}^{T} \hat{\boldsymbol{\epsilon}}(t)\,\hat{\boldsymbol{\epsilon}}(t)^{*\prime})$, of the approximation error obtained from the standard and the dynamic approach for each patient.

As the minimization problem (4) includes that of (1) as a special case, the total error variance of the PCA according to Brillinger is of course at least as small as

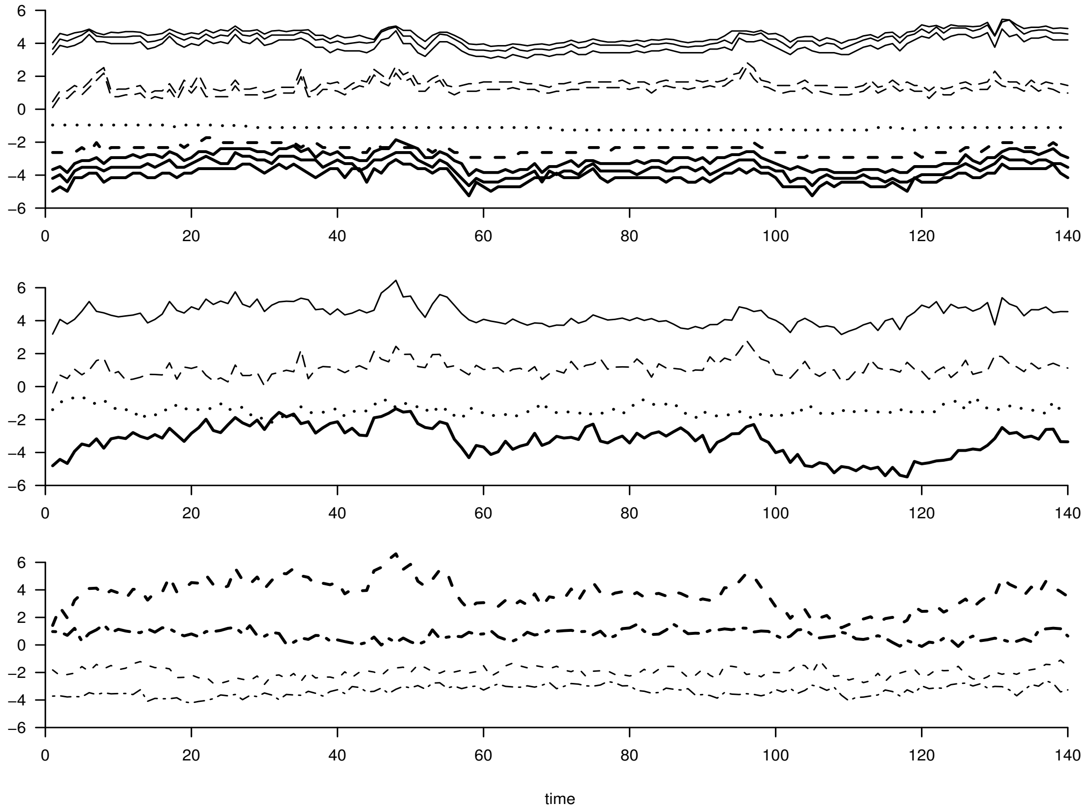

Figure 5: Top: Standardized time series of ten hemodynamic vital signs, arterial blood pressures (*solid*), HR and pulse (*dashed*), TEMP (*dotted*), pulmonary artery pressures (*bold solid*), CVP (*bold dashed*). Center: Extracted dynamic PC series after Procrustes–rotation representing arterial blood pressures (*solid*), HR and pulse (*dashed*), intrathoracic pressures (*bold solid*), linear combination of various variables (*dotted*). Bottom: Extracted dynamic PC series without rotation, ordered from first PC series(*top*) to fourth PC series (*bottom*)

that of the standard PCA. Figure 6 gives an impression of the improved approximation that can be achieved by taking into account the time series structure of the data. On average, the total error variance from the PCA in the frequency domain is approximately 64% of the one resulting from a classical PCA.

Regarding the considerably low filter loadings at lags different from zero, we are also interested in an examination of truncated versions of the PC filters, where future and possibly also past observations beyond some time lags s_p^* and s_f^* are not considered. Neglecting future observations is an intuitive way to "remove" the non–causality of the PC filter. In comparison to a classical PCA, this approach still allows for some

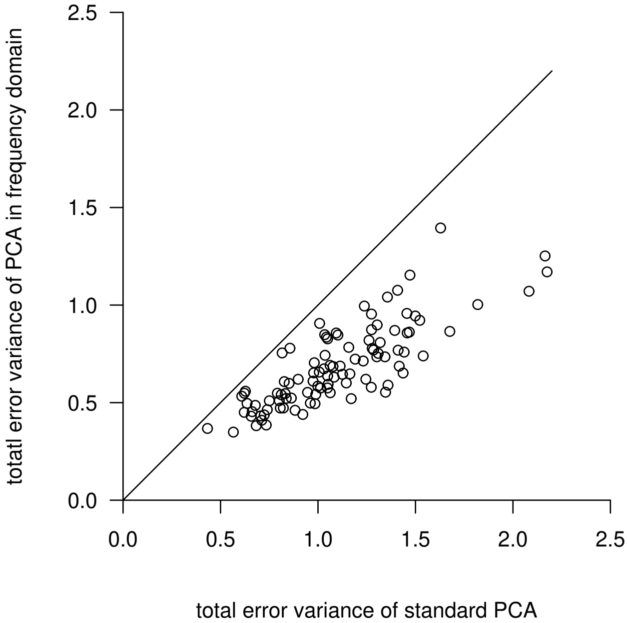

Figure 6: Scatterplot of the total error variances after performing a classical PCA and a PCA in the frequency domain for 100 patients

of the time–lagged dependencies in the data. On the other hand, relatively short filters meet the demands of an application in the online–monitoring context.
We therefore examine how a truncation of the PC filter affects the approximation error (3), knowing that the total error variance is minimized for the filter of full length.

In Figure 7 we show the total variance of the approximation error for one patient subject to the truncation lag $s^* = s_f^* = s_p^*$, where $\{\mathbf{b}(s), -s^* \leq s \leq s^*\}$, $\{\mathbf{c}(s), -s^* \leq s \leq s^*\}$ and $s^* = 1, \ldots, 900$. For cut off lags with $900 < s^* \leq 3743 = T$ the approximation error is almost constant, slightly varying between 0.47 and 0.53. For truncation lags $s^* \leq 60$, the approximation is even worse than that resulting from a standard PCA on the same time series (total error variance approximately 0.81). This behaviour of the total error variance subject to the truncation lag is typical for all of the analyzed time series. Hence, such a brute force truncation at "small" time lags is not advisable if one's interest is the minimization of the approximation error (3) without any constraints on the form of the filter. In our example, truncation lags beyond some cut off value, say $s^* = 420$, are acceptable in terms of the size of the total error variance. Yet, such PC filters are not desirable in an online application as they exceed the number of (future) observations that can appropriately be considered. One–sided truncations that completely neglect future observations seem to be even worse. Therefore, a mere truncation of the resulting PC filters is not a solution if we are interested in both a consideration of the time–

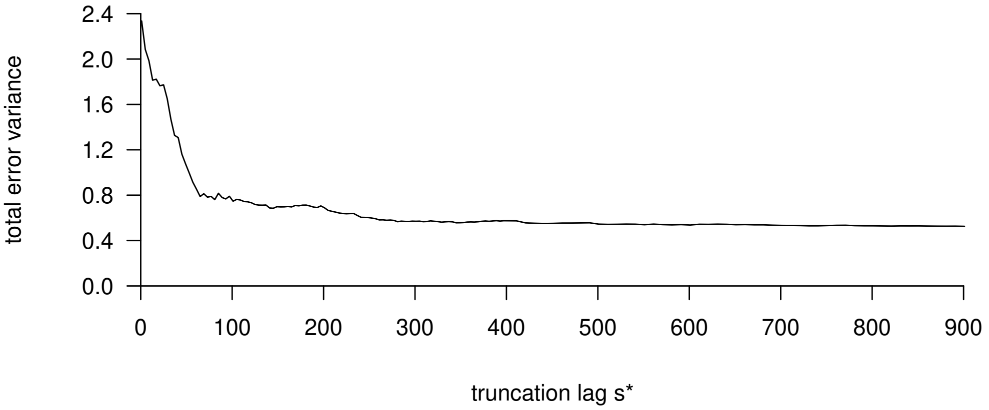

Figure 7: Total variance of the approximation error subject to the PC filter truncation lag $s^* \leq 900$ for one patient with a total of $T = 3743$ observations

lagged dependency structure and short time filters. To this end, one could modify the minimization problem (4) and impose suitable constraints on the PC filter.

5 Discussion and Conclusion

The aim of our investigation was to examine the performance of PCA for time series in the frequency domain (compared to standard PCA) with the intention of reducing multivariate time series of hemodynamic variables to a small number of meaningful component series.

We find that the non–uniqueness of the dynamic eigenvectors heavily complicates the computation of the PC filter. Nevertheless, Procrustes rotation at all Fourier frequencies improves the situation and yields interpretable PC series. Compared to standard PCA, the consideration of the dependencies between the time series observations yields a better approximation in terms of the total error variances. However, this advantage should be balanced by the fact that the PCA in the frequency domain needs long sequences of observations, yields non–causal filters and requires more computation time, which is undesirable in the online–monitoring situation.

An investigation of the obtained PC filters in the time domain reveals a high impact of the observations at lag zero on the PC series but considerably low loadings at further lags. On the other hand, brute force truncations of the PC filters at small or moderate time lags offer no acceptable solution in terms of a small approximation error. Here further investigations of the effect of a mere filter truncation and modifications or refined approaches seem to be necessary.

One further difficulty is that due to the variability across patients we need patient–specific filters. These have to be estimated from a sequence of the individual time series observations. The obtained filter could then be used in the online–monitoring context, but this approach does not allow for changes in the dynamic dependency structure that might become noticeable through the estimated PC filter.
In view of these complications it remains questionable whether the time series approach is to be preferred to a classical PCA for an online dimension reduction of physiological time series. Further examinations of the obtained filters are required.

6 Appendix: Geometrical justification

For simplicity we set $\mathsf{E}(\boldsymbol{X}(t)) = \mathbf{0}$ and denote by $\boldsymbol{x}(t)$ the centered observations of the observed k–dimensional time series with mean 0. The objective is to find filters $\{\mathbf{b}(s),\ s \in \mathbb{Z}\}$ and $\{\mathbf{c}(s),\ s \in \mathbb{Z}\}$, with $r \times k$– and $k \times r$–dimensional matrices as entries, $r < k$, that minimize

$$\sum_{t=1}^{T} \left[\left(\boldsymbol{x}(t) - \sum_{s} \mathbf{c}(t-s)\boldsymbol{\xi}(s) \right)^{*\prime} \left(\boldsymbol{x}(t) - \sum_{s} \mathbf{c}(t-s)\boldsymbol{\xi}(s) \right) \right], \tag{6}$$

where

$$\boldsymbol{\xi}(t) = \sum_{s} \mathbf{b}(t-s)\boldsymbol{x}(s), \quad t = 0, \pm 1, \pm 2, \ldots .$$

Note, that given a sequence $(\boldsymbol{y}(s)) \in \mathbb{C}^k$, $s = 1, 2, \ldots, T$ we can consider the expression

$$d_{\boldsymbol{y}}^{(T)}(\alpha) = \sum_{s=1}^{T} \boldsymbol{y}(s)\ \exp(-i\alpha s), \quad -\infty < \alpha < \infty$$

that is defined as the finite Fourier transform of $(\boldsymbol{y}(s))$.
The periodogram of $\boldsymbol{y}(s)$ at frequency $\alpha_j = 2\pi j/T$, $j \in \mathbb{Z}$ is defined in terms of the discrete Fourier transform $d_{\boldsymbol{y}}^{(T)}$ of $\boldsymbol{y}(s)$ by

$$\boldsymbol{I}_{\boldsymbol{y}\boldsymbol{y}}^{(T)}(\alpha_j) = (2\pi T)^{-1}\ d_{\boldsymbol{y}}^{(T)}(\alpha_j)\ d_{\boldsymbol{y}}^{(T)}(\alpha_j)^{\prime *}. \tag{7}$$

These transformations are not restricted to realizations of stationary processes but are defined for any set of time series observations.

Replacing the error series $\boldsymbol{\epsilon}(t) = \boldsymbol{x}(t) - \sum_s \mathbf{c}(t-s)\boldsymbol{\xi}(s)$ in (6) we have

$$\begin{aligned}
\sum_{t=1}^{T} \boldsymbol{\epsilon}(t)^{*\prime}\, \boldsymbol{\epsilon}(t) &= \sum_{t=1}^{T} \mathrm{tr}\ (\boldsymbol{\epsilon}(t)^{*\prime}\, \boldsymbol{\epsilon}(t)) = \mathrm{tr}\ \left(\sum_{t=1}^{T} \boldsymbol{\epsilon}(t)\, \boldsymbol{\epsilon}(t)^{*\prime} \right) \\
&= \mathrm{tr}\ \left(T^{-1} \sum_{j=1}^{T} d_{\boldsymbol{\epsilon}}^{(T)}(\alpha_j)\, d_{\boldsymbol{\epsilon}}^{(T)}(\alpha_j)^{*\prime} \right) \\
&= \mathrm{tr}\ \left(T^{-1} \sum_{j=1}^{T} (\mathbf{I} - \mathbf{A}(\alpha_j))\, d_{\boldsymbol{x}}^{(T)}(\alpha_j)\, d_{\boldsymbol{x}}^{(T)}(\alpha_j)^{*\prime}\, (\mathbf{I} - \mathbf{A}(\alpha_j))^{*\prime} \right),
\end{aligned}$$

where $\mathbf{A}(\alpha) = \mathbf{C}(\alpha)\mathbf{B}(\alpha)$, $\mathbf{B}(\alpha) = \sum\limits_s \mathbf{b}(s)\ \exp(-i\alpha s)$, and $\mathbf{C}(\alpha)$ analogous to $\mathbf{B}(\alpha)$. Inserting (7) in the result above, we get the minimization problem

$$\min_{\{\mathbf{b}(s)\},\{\mathbf{c}(s)\}}\ \mathrm{tr}\ \left(2\pi \sum_{j=1}^{T} (\mathbf{I} - \mathbf{A}(\alpha_j))\, \mathbf{I}_{\boldsymbol{x}\boldsymbol{x}}^{(T)}(\alpha_j)\, (\mathbf{I} - \mathbf{A}(\alpha_j))^{*\prime} \right),$$

which can be solved by minimizing

$$\mathrm{tr}\ \left((\mathbf{I} - \mathbf{A}(\alpha_j))\, \mathbf{I}_{\boldsymbol{x}\boldsymbol{x}}^{(T)}(\alpha_j)\, (\mathbf{I} - \mathbf{A}(\alpha_j))^{*\prime} \right)$$

separately for each α_j with $\mathrm{rk}\,(\mathbf{A}(\alpha)) \leq r$. Similar to the proof of Theorem 9.3.2 in Brillinger (1975), we see that a minimum is attained at

$$\mathbf{A}(\alpha_j) = \sum_{i=1}^{r} \boldsymbol{l}_i(\alpha_j)\boldsymbol{l}_i(\alpha_j)^{\prime *},$$

where $\boldsymbol{l}_i(\alpha_j)$ is the eigenvector corresponding to the i–th eigenvalue of $\mathbf{I}_{\boldsymbol{x}\boldsymbol{x}}^{(T)}(\alpha_j)$, and thus

$$\mathbf{B}(\alpha_j) = \begin{bmatrix} \boldsymbol{l}_1(\alpha_j)^{*\prime} \\ \vdots \\ \boldsymbol{l}_r(\alpha_j)^{*\prime} \end{bmatrix}, \quad \mathbf{C}(\alpha_j) = [\boldsymbol{l}_1(\alpha_j), \ldots, \boldsymbol{l}_r(\alpha_j)] = \mathbf{B}(\alpha_j)^{*\prime}.$$

The above minimization problem yields a purely geometrical justification if it is the objective to find a best lower dimensional approximation of an observed time series $\boldsymbol{x}(t)$. It can be understood as a finite sample version of the population result (Theorem 9.3.2) in Brillinger (1975). We did not make use of stochastic properties of the periodogram. Therefore stationarity of the underlying stochastic process is not required.

Acknowledgement

We would like to thank Roland Fried for helpful comments, Dan Nordman for careful reading, and Britta Pouwels for providing a first version of the program in R. Financial support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures") is gratefully acknowledged.

References

Brillinger, D.R. (1975), *Time Series: Data Analysis and Theory*, first edition, Holt, Rinehart and Winston, Inc., New York, (expanded edition: (1981), Holden Day, Inc., San Francisco, classics edition: (2001), SIAM, Philadelphia).

Dahlhaus, R. (2000), Graphical interaction models for multivariate time series, *Metrika* **51**, 157–172.

Davies, P.L., Fried, R., Gather, U. (2002), Robust signal extraction for on–line monitoring data, *Journal for Statistical Planning and Inference*, provisionally accepted.

Danilov, D.L. (1997), Principal components in time series forecast, *Journal of computational and graphical statistics* **6**, 112–121.

Flury, B. (1988), *Common principal components and related multivariate models*, Wiley, New York.

Forni, M., Hallin, M., Lippi, M., Reichlin, L. (2000), The generalized dynamic factor model: identification and estimation, *Review of Economics and Statistics* **82**, 540–554.

Fried, R. (2002), Robust filtering of time series with trends, Preprint, Department of Statistics, University of Dortmund.

Fried, R., Gather, U., Imhoff, M., Keller, M., Lanius, V. (2000), Combining graphical models and principal component analysis for statistical process control, in: *COMPSTAT 2002 – Proceedings in Computational Statistics,* Härdle, W., Rönz, B. (eds.), Springer, Heidelberg, 237 – 242.

Gather, U., Imhoff, M., Fried, R. (2002), Graphical models for multivariate time series from intensive care monitoring, *Statistics in medicine* **21**, 2685–2701.

Geweke, J. (1977), The dynamic factor analysis of economic time series, in: *Latent Variables in Socio–Economic Models,* Aigner, D., Goldberger, A.S. (eds.), Amsterdam, North–Holland, 365–383.

Goljandina N.E., Nekrutkin V.V., Zhigljavsky A.A. (2001), *Analysis of time series structure: SSA and related technique*, Chapman & Hall / CRS, Boca Raton.

Hotelling, H. (1933), Analysis of a complex of statistical variables into principal components, *Journal of Educational Psychology* **24**, 417–441, 498–520.

Imhoff, M., Fried, R., Gather, U. (2002), Detecting relationships between physiological variables using graphical modelling, *Journal of the American Medical Informatics Association,* in print.

Jolliffe, I.T. (1995), Rotation of principal components: Choice of normalization constraints, *Journal of Applied Statistics* **22**, 29–35.

Jolliffe, I.T. (2000), *Principal component analysis*, Springer, New York.

Lauritzen, S.L., Wermuth, N. (1989), Graphical models for associations between variables, some of which are qualitative and some quantitative, *Annals of Statistics* **17**, 31–57.

Molenaar, P.C.M. (1987), Dynamic factor analysis in the frequency domian: causal modeling of multivariate psychophysiological time series, *Multivariate Behavioral Research* **22**, 329–353.

Molenaar, P.C.M., De Gooijer, J.G., Schmitz, B. (1992), Dynamic factor analysis of nonstationary multivariate time series, *Psychometrika* **57**, 333–349.

Pearson, K. (1901), On lines and planes of closest fit to systems of points in space, *Phil. Mag.* **2**, 559–572.

Peña, D., Box, G.E.P. (1987), Identifying a simplifying structure in time series, *Journal of the American Statistical Association* **82**, 836–843.

Robinson, E.A., Silvia, M.T. (1978), *Digital signal processing and time series analysis*, Holden–Day, San Francisco.

Shumway, R. H., Stoffer, D. S. (2000), *Time Series Analysis and Its Applications*, Springer, New York, Berlin.

Stock, J.H., Watson, M.W. (1988), Testing for common trends, *Journal of the American statistical Association* **83**, 1097–1107.