RESEARCH REPORT SERIES
(Statistics #2008-3)

A Review of Some Modern Approaches to the Problem of Trend Extraction

Theodore Alexandrov[1]
Silvia Bianconcini[2]
Estela Bee Dagum[2]
Peter Maass[1]
Tucker McElroy

Center of Industrial Mathematics, University of Bremen[1]

Department of Statistics, University of Bologna[2]

Statistical Research Division
U. S. Census Bureau
Washington, D.C. 20233

Report Issued: March 28, 2008

A review of some modern approaches to the problem of trend extraction

Theodore Alexandrov[1], Silvia Bianconcini[2], Estela Bee Dagum[2], Peter Maass[1], Tucker S. McElroy[3]

[1] Center of Industrial Mathematics, University of Bremen, Postfach 33 04 40, 28334 Bremen

[2] Department of Statistics, University of Bologna, Via Belle Arti, 41, 40126 Bologna

[3] Statistical Research Division, U.S. Census Bureau, 4700 Silver Hill Road, Washington, D.C. 20233-9100

Abstract

Trend extraction is one of the major tasks of time series analysis. The trend of a time series is considered as a smooth additive component that contains information about global change. This paper presents a review of some modern approaches to trend extraction for one-dimensional time series. We do not aim to review all the novel approaches, but rather to observe the problem from different viewpoints and from different areas of expertise. The paper contributes to understanding the concept of a trend and the problem of its extraction. We present an overview of advantages and disadvantages of the approaches under consideration, which are: the Model-Based Approach, nonparametric linear filtering, Singular Spectrum Analysis, and wavelets. The Model-Based Approach assumes the specification of a stochastic time series model for the trend, which is usually either an ARIMA model or a state space model. The nonparametric filtering methods (i.e., the Henderson, LOESS, and Hodrick-Prescott filters) do not require specification of a model; they are quite easy to apply and are used in all applied areas of time series analysis. For these well-known methods we show how their properties can be improved by exploiting Reproducing Kernel Hilbert Space methodology. In addition to these extremely popular approaches, we consider Singular Spectrum Analysis (SSA) and wavelet-based methods. Singular Spectrum Analysis is widespread in the geosciences; its algorithm is similar to that of Principal Components Analysis, but SSA is applied to time series. Wavelet-based methods are currently a de facto standard for denoising in many fields. We summarize how the powerful wavelets approach can be used for trend extraction.

Key words: Model-based approach; Nonparametric linear filtering; Singular spectrum analysis; Time series; Trend; Wavelets.

Disclaimer: This report is released to inform interested parties of research and to encourage discussion. Any views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau or the Universities of Bremen and Bologna.

1 Introduction

Mathematical approaches to the trend extraction problem have a long history. At the beginning this problem was formulated as fitting of a simple deterministic function (usually a linear one) to data. With rapid growth of the theory of stationary time series, the trend was considered as a deterministic component that needed to be subtracted out in order to obtain a stationary time series, which afterwards can be successfully modeled – see Wold (1938). At the present time, stochastic approaches to the definition of a trend are widely used, especially in econometrics. In this paper, we present the contemporary research on the trend extraction problem for one-dimensional time series.

Let us define a *time series* of length N as $X = (x_0, \ldots, x_{N-1})$, $x_n \in \mathbb{R}$. There are a variety of definitions of trend but all of them imply the following additive model:

$$x_n = t_n + r_n, \quad \text{or} \quad X = T + R, \tag{1}$$

where $T = (t_0, \ldots, t_{N-1})$ denotes a trend and $R = (r_0, \ldots, r_{N-1})$ is referred as *residual.* The latter can have both deterministic and stochastic parts, d_n and s_n. Hence, we come to the following expansion

$$x_n = t_n + d_n + s_n. \tag{2}$$

Model (2) with a periodic d_n and a zero mean s_n is referred to as the *classical decomposition model* (Brockwell & Davis, 2003).

In his remarkable book, Chatfield (1996) defines trend as "a long-term change in the mean level" and expresses a typical viewpoint on trend by considering it as a smooth additive component which contains the information about time series global change. The problem of extraction of such a component occurs in many applied sciences and attracts scientists in different areas which apply their knowledge for its solution. Therefore, at the present time there exist many methods for trend extraction. These methods differ in their complexity and interpretability, as well as the mathematical tools that they use, and hence produce different results. One can hardly determine the best method, because each method is superior to the others when applied within its intended context. On the other hand, the estimation of the quality of processing of real-life time series is subject to the preferences of experts in the area of application.

In this paper, we evaluate the following prominent approaches to trend extraction: Model-Based Approach, nonparametric filtering, singular spectrum analysis, and wavelets. *Model-Based*

Approach (MBA) unites methods by assuming specification of a stochastic time series model which is usually either an autoregressive model or a state space model. The developments of such methods were heavily influenced by engineering and econometric problems. Currently, they are the most popular in econometrics and used in many other areas. The *nonparametric filtering* methods do not require specification of a model; they are quite easy in application and are used in all applied areas of time series analysis. In addition to these extremely popular approaches, we consider *Singular Spectrum Analysis* (SSA) and wavelets-based methods. Algorithm and principles of SSA are very similar to those of Principal Components Analysis, but SSA is applied to time series. It mainly originated in dynamical systems and at the present time is widespread in geosciences. *Wavelets-based methods* are currently a *de facto* standard for denoising in many applications. We provide an overview of how they can be used for trend extraction.

Certainly, there are many other methods for trend extraction. Nevertheless, the majority of methods for this problem follow one of the considered approaches. This paper is organized as follows. In section 2, we discuss the problem of trend extraction. Section 3 reviews the Model-Based Approach. In section 4, we describe the nonparametric linear filtering methods, based on the Henderson, the LOESS and the Hodrick-Prescott filters. Then we show how the "kernel trick" of operating in a Reproducing Kernel Hilbert Space can improve their properties. In section 5, we consider Singular Spectrum Analysis. Wavelet methods for trend extraction are reviewed in section 6. Finally, in section 7, an example of application of the methods to a real-life time series is examined. Even though one cannot derive general conclusions from this single example, nevertheless it demonstrates some special aspects of the methods.

2 The trend extraction problem

For a time series satisfying model (1), the *problem of trend extraction* is defined as estimation of an unknown T having only X. Note that in the literature devoted to state space methods (see section 3) one refers to the *smoothing problem* as estimating the whole vector T from X (Durbin & Koopman, 2001). However, if we only desire t_{N-1}, i.e., the most current value of the trend, then this is called the *filtering problem.*

2.1 Deterministic and stochastic trend definition

Following a *deterministic approach*, the trend is defined as a deterministic function of time which belongs to some class. The function can be defined explicitly, specifying a parametric model, or implicitly, exploiting some long-changing (or smoothness) condition. A widespread example of the parametric trend is a polynomial trend and its elementary version, that is a linear trend. Its simplicity still makes it of use as a deterministic portion in stochastic time series modeling (Hamilton, 1994). The smoothness condition is usually formulated in terms of derivatives or Fourier coefficients. The advanced approaches to the definition of smoothness imply the use of Sobolev and Besov spaces (Triebel, 1992).

There exists a *stochastic* viewpoint on the definition of trend, namely when the trend is defined as a realization of a stochastic process. This trend is supposed to be smooth where the smoothness is expressed in terms of a variance or autocorrelation function (Froeb & Koyak, 1994). An elementary example of a stochastic trend is a random walk with a drift. Modeling of this trend is used in economic applications (Pierce, 1978; Stock & Watson, 1988). In the stochastic approaches, it is typical to assume orthogonality between the trend and the residual (the latter is generally supposed to be stochastic) – see section 3.

Certainly, the property of smoothness (or slow-changing) is related to the length of the time series. Economists and statisticians are often interested in the "short" term trend of socioeconomic time series. The short-term trend generally includes cyclical fluctuations and is referred to as the *trend-cycle*.

2.2 Difference from denoising problem

If s_n corresponds to noise then the problem of trend extraction is similar to the problem of *denoising*. The difference lies in two facts. Firstly, having nonzero d_n in trend extraction, one should take care on separating a trend from d_n. For example, when a time series has some periodic component, a trend extraction procedure should omit it. Secondly, in denoising s_n is usually assumed to follow one of the typical noise models (i.e., white or red noise), whereas in trend extraction s_n may follow much more general models (i.e., it may be non-stationary).

2.3 Updating trend with new data

In signal processing, the filter which uses only past and present values is called a *causal filter*. In econometrics, such a filter is named a *concurrent filter*.

Given a time series where more data will arrive in the future, one has two options for presenting trends: the online approach and the window approach. The *online approach* generally uses concurrent filters (or trend extraction methods that do not require future data), and each time a new data point arrives, one generates the next concurrent estimate. This approach is favored by many users of data who are not statisticians, from day-traders to bankers – see Wildi (2005) for a discussion for the motivations behind online estimates of trend. Exponential smoothing is really online filtering (Durbin & Koopman, 2001).

On the other hand, the *window approach* produces trend estimates at every time point in the sample. When new data arrives, trend estimates in the middle of the sample must be updated, or revised. For references on revisions in the context of seasonal adjustment, see Pierce (1980); Findley et al. (1998); Maravall & Caporello (2004); McElroy & Gagnon (2006).

Generally statisticians prefer the window method, whereas public users who are non-experts dislike it because revisions are problematic.

2.4 Trend detection

Sometimes, one needs to determine whether there exists a trend in a given time series. This problem is usually referred to as *trend detection* and solved by means of statistical tests, which require specification of a trend model.

The detection of monotonic trends has been the most extensively studied. One of the widespread tests for this problem is the non-parametric *Mann-Kendall test* and its modifications for handling seasonality (Hirsch & Slack, 1984) and autocorrelated data (Hamed & Rao, 1998). Among others are: the parametric *t-test* and the nonparametric *Mann-Whitney* and *Spearman tests*. Berryman et al. (1988) describes many methods for monotonic trend detection and provides a selection algorithm for them. For more recent developments see the review of Esterby (1996), which is focused on hydrological applications; also see the comparative study Yue & Pilon (2004), which describes some bootstrap-based tests.

3 Model-based approach

3.1 Preamble

The Model-Based Approach (MBA) to trend estimation refers to a family of methods, which have in common the reliance upon time series models for the observed, trend, and residual processes. The history of this approach is briefly discussed below.

In the discussion of trend estimation in this article, we naturally focus upon finite samples, since this is the only data available in practice. The early literature on MBA signal extraction developed the theory for bi-infinite (Wiener, 1949; Kolmogorov, 1939, 1941) or semi-infinite samples (Whittle, 1963), and exclusively focused on stationary processes. We also note that this early theory encompassed continuous-time processes as well, since early "filters" were essentially given by the operation of analog-type hardware; however we do not pursue continuous-time trend estimation here; see Koopmans (1974) for a discussion.

The MBA literature on trend extraction began to be generalized in two directions: dealing with boundary effects (i.e., the finite sample) and handling nonstationarity (generally speaking, homogeneous nonstationarity exemplified by ARIMA processes). The engineering community focused on the former, the pivotal discovery being the so-called Kalman filter (Kalman, 1960). Rauch (1963) extended the Kalman filter to a smoother that could handle boundary effects; these algorithms rely on a State Space Formulation (SSF) of trend extraction. Additional discussion of state space methods from an engineering perspective can be found in Anderson & Moore (1979); Young (1984).

However, since engineers are primarily concerned with stationary data, the SSF approach did not handle nonstationary data until econometricians became involved later on. Books that discuss SSF from an econometrics/statistical perspective include Harvey (1989); West & Harrison (1997); Kitagawa & Gersch (1996); Durbin & Koopman (2001). Generally speaking, trends are nonstationary processes, so the basic stationary approach of the older engineering literature is not adequate.

In our description of MBA we focus upon methods following the window approach (see section 2.3) although the discussion is easily adapted to online-trend extraction as well, since concurrent filters are included within a linear smoother.

General questions The MBA method of trend estimation generally requires a specification of the dynamics of signal and noise (trend and residual). These are typically considered to be stochastic, but with possible deterministic portions as well. The deterministic portions of the trend refer to linear or quadratic polynomial functions that are related to the initial values of the stochastic process, and underlie the stochastic trend. Given this fundamental notion, the MBA method requires some thought about the following issues, in more or less this order: (i) How are the trend and residual processes related to the observed process? (ii) How are trend and residual related to one another? (iii) How are trend estimates to be generated? (iv) What types of models are being considered, and how are they estimated?

3.2 Trend in MBA

One approach is to view the trend as the output of a known linear filter applied to the data. This assumption is implicitly understood in many of the trend estimation approaches in the engineering literature. One example is the Direct Filter Approach of Wildi (2005). However, it is more common for statisticians to specify a target trend T that is not a direct function of the data X, but whose dynamics are specified to a greater or lesser extent by the practitioner, typically by specifying a model.

Relation between trend and residual The two most popular assumptions which regulate the relations between trend and residual are the orthogonal and *Beveridge-Nelson* (Beveridge & Nelson, 1981), or BN. The BN assumes that both trend and residual can be written as linear filters of the data innovation process, and thus are "fully" correlated. The *orthogonal approach* assumes that "differenced" trend and residual (i.e., the components after nonstationary effects have been removed by differencing) are uncorrelated with one another. This supposition is more consistent with economic data, since diverse components are thought to originate from diverse aspects of the economy, and thus should not be correlated. Naturally, the orthogonal decomposition and BN decomposition represent the opposite ends of the spectrum; some work by Proietti (2006) deals with the case that the components are less than fully correlated.

3.3 Construction of the trend model

For MBA trend extraction, we require a model for the trend and residual – note that this residual may contain seasonal effects, and can therefore be nonstationary. There are several popular approaches for obtaining these models from the data: Decomposition, Structural, and BN.

The *Decomposition approach* (Hillmer & Tiao, 1982; Burman, 1980) begins by attempting to fit an optimal model to the observed data, where optimality is often equated with maximum likelihood estimation of model parameters after different model specifications have been compared via information criteria (e.g., Akaike Information Criterion or other goodness-of-fit diagnostic tests; see Findley et al. (1998) for a discussion). Then the models for trend and residual are determined via partial fraction decomposition techniques applied to the autocovariance generating function of the model for the data (this assumes that ARIMA or Seasonal ARIMA models are being used). Some amount of user-specification is required, since all differencing operators and autoregressive operators that appeared in the data model must be allocated (subjectively) to the various components. For more discussion of this, see Bell & Hillmer (1984) and Chapter 8 of Peña et al. (2001). Typically there is indeterminancy of the derived component models; maximizing the variance of the irregular component results in the *Canonical Decomposition*, which results in signals that are as stable as possible.

The *Structural approach* (Harvey, 1989) on the other hand also uses maximum likelihood estimation, but the form of the likelihood is dictated by a pre-specified model form for the components. This is also referred to as an Unobserved Components (UC) approach. While sometimes a canonical decomposition does not exist (mathematically this is possible and not uncommon), the structural approach is always viable. However, the implied model for the data process need not be the best among all contenders, as it is in the decomposition approach. Also, more *a priori* information about the trend and residual dynamics are needed from the user, such as specifying the differencing order for the trend ahead of time. See Durbin & Koopman (2001) for more discussion.

Note that the term *Structural Model* refers to a particularly simple class of component models promoted by Gersch & Kitagawa (1983), which are essentially parameter-restricted ARIMA models. Here we distinguish between the Structural Approach to estimating component models (which can be general ARIMA models) and the more specific Structural Models utilized in STAMP and SsfPack (Koopman et al., 1999).

The *BN approach* is much like the Structural, though now the component models are fully correlated. Naturally this dictates a different form for the likelihood of the data in terms of the component models, since they are no longer orthogonal. However, we can still utilize maximum likelihood estimation to get the component models, and the corresponding trend filters are then easy to obtain; see Morley et al. (2003) and Proietti (2006).

3.4 Penalty function

The next issue is: what sort of penalty function is used to determine optimal signal extraction? Mean Squared Error (MSE) is very popular among statisticians, and is the original penalty function used in Wiener (1949) and Kolmogorov (1939, 1941). The conditional expectation of the trend T onto the data F is the estimate that minimizes MSE; when the data has a Gaussian distribution, this will be a linear function of the data, i.e., a smoother. This explains the central role of linear smoothers in the MBA literature. However, when alternative distributions are present (e.g., log-normals), other penalty functions such as Relative MSE may be more appropriate; see Thomson & Ozaki (2002) and McElroy (2006).

If we are interested in linear smoothers, how do we find the best one? Bell (1984) discusses how these are computed from the autocovariance generating functions of trend and residual, assuming certain conditions on the initial values of the data; also see Cleveland & Tiao (1976). Adaptations of this theory to finite samples can be found in Bell & Hilmer (1988); Bell (2004); McElroy & Sutcliffe (2006) and McElroy (2008). The following references discuss the finite-sample theory from an SSF viewpoint: Bell & Hilmer (1991); De Jong (1991); Koopman (1997) and Durbin & Koopman (2001). Since these MSE-optimal linear smoothers produce estimates with dynamics that differ from that of the target, some work has been done in producing estimates whose dynamics exactly match those of the target (Wecker (1979); also see the discussion in Findley & Martin (2006)). Another type of smoothing is given in the Square-Wave work of Pollock (2000). An excellent discussion of smoothers, contrasting the SSF approach with deterministic methods can be found in Young & Pedregal (1999).

Given that most statistical readers will be interested in linear MSE-optimal smoothers, we focus on the SSF approach and the matrix approach, which are equivalent. The widely-used SSF smoother requires *Assumption A* of Bell (1984) for its estimates to be MSE-optimal, as is discussed in Bell & Hilmer (1988). Assumption A states that the d initial values of the process x_n are independent of differenced trend and differenced residual. Note that the SSF smoother is a linear operation on the data F that produces a vector of trend estimates T; the linear matrix that accomplishes this is derived in McElroy (2008). For some purposes, it is convenient to have this matrix, e.g., the full error covariance matrix is easily obtained using this approach.

3.5 Model classes

Finally, the MBA requires a choice of model classes. As mentioned above, the Decomposition approach relies on seasonal ARIMA models for the components. The ARIMA and Structural models (Harvey, 1989) are the most popular models in econometric MBA trend estimation. In theory, one only needs the autocovariance generating function for trend and residual in order to proceed. For example, another class of models are time-varying coefficient models, where the parameters evolve according to a random walk or other such process; a discussion of using such models can be found in Young et al. (1999).

3.6 Software

Several of the main software packages for MBA trend estimation include X-12-ARIMA (Findley et al., 1998), TRAMO-SEATS (Maravall & Caporello, 2004), STAMP (Koopman et al., 2000), and microCAPTAIN (Young & Benner, 1991). X-12-ARIMA mixes nonparametric linear filtering with model-based forecast and backcast extension, so it can be viewed as a partial MBA. SEATS is fully MBA, and utilizes the *Canonical Decomposition* approach. On the other hand, STAMP utilizes a *Structural Approach* as well as *Structural Models*. The program microCAPTAIN also uses a Structural Approach, but with time-varying coefficient models that are estimated using the frequency domain method of Dynamic Harmonic Regression (DHR), as opposed to the Maximum Likelihood Estimation method of the other software (Young et al., 1999). These are some of the core MBA software products for trend estimation, of which some other products (DEMETRA of EuroStat, SAS implementations of *Structural Models*) are derivatives. We also mention the RegComponent software (Bell, 2004), which uses a *Structural Approach* with ARIMA models (although extensions now allow for a *Decomposition Approach* as well); this was one of the first programs to simultaneously handle smoothing by SSF methods and also estimate fixed regression effects.

4 Nonparametric trend predictors

4.1 Preamble

In the nonparametric trend filtering approach the model (1) is usually considered, where t_n is referred as a signal and r_n is assumed to be either a white noise, $NID(0, \sigma^2)$, or, more generally, to follow a stationary and invertible Autoregressive Moving Average process. Assuming that the input series X is seasonally adjusted or without seasonality, the signal represents the trend and cyclical components, usually referred to as *trend-cycle* for they are estimated jointly. Expecting that the signal t_n is smooth, it can be *locally approximated* by a polynomial of degree d on the time distance j between x_n and its neighbors x_{n+j}

$$t_{n+j} = a_0 + a_1 j + ... + a_d j^d + \varepsilon_{n+j},$$

where $a_k \in \mathbb{R}$ and ε_n is assumed to follow a white noise process mutually uncorrelated with r_n. The coefficients $\{a_k\}$ can be estimated by ordinary or weighted least squares or by summation formulae.

The solution for $\hat{a}_0$ provides the trend-cycle estimate $\hat{T} = \{\hat{t}_n\}$, which equivalently consists in a moving weighted average

$$\hat{t}_n = \sum_{j=-m}^{m} b_j x_{n-j}.$$

The applied weights $\{b_j\}$ depend on: (i) the degree of the fitted polynomial d, (ii) the amplitude of the neighborhood $2m+1$, and (iii) the shape of the function used to average the observations in each neighborhood.

The local polynomial regression predictor developed by Henderson (1916) and LOESS due to Cleveland (1979) are the most widely applied nonparametric local filtering methods to estimate the short-term trend of seasonally adjusted economic indicators. In this section we also consider the Hodrick & Prescott (1997) filter which is widely used for economic and financial applications.

4.2 Henderson filter

The Henderson filters are derived from the graduation theory, known to minimize smoothing with respect to a third degree polynomial within the span of the filter. The minimization problem

$$\min_{a_k, 0 \leqslant k \leqslant 3} \left\{ \sum_{j=-m}^{m} w_j \left[x_{t+j} - a_0 - a_1 j - a_2 j^2 - a_3 j^3 \right]^2 \right\}, \tag{3}$$

is considered, where the symmetric weights w_j are chosen to minimize the sum of squares of their third differences (*smoothing criterion*). This filter has the property that fitted to exact cubic functions will reproduce their values, and fitted to stochastic cubic polynomials it will give smoother results than those obtained by OLS – see Macauley (1931). Henderson (1916) proved that two alternative smoothing criteria give the same formula, as shown explicitly by Kenny & Durbin (1982) and Gray & Thomson (1996): (i) minimization of the variance of the third differences of the series $\hat{t}_n$ defined by the application of the moving average; (ii) minimization of the sum of squares of the third differences of the coefficients b_j of the moving average formula. Moreover, Henderson (1916) showed that the nth element of the trend estimation, $\hat{t}_n$, is given by

$$\hat{t}_n = \sum_{j=-m}^{m} \phi(j) w_j x_{n-j}$$

where $\phi(j)$ is a cubic polynomial whose coefficients have the property that the smoother reproduces the data if they follow a cubic. Henderson also proved the converse: if the coefficients b_j of a cubic-reproducing summation formula do not change their sign more than three times within the filter span, then the formula can be represented as a local cubic smoother with weights $w_j > 0$ and a cubic polynomial $\phi(j)$ such that $\phi(j) w_j = b_j$. To obtain w_j from b_j one simply divides b_j by a cubic polynomial whose roots match those of b_j.

The asymmetric filters commonly used in association with the Henderson smoother were developed by Musgrave (1964) on the basis of minimizing the mean squared revision between final and preliminary estimates. Although the basic assumption is that of fitting a linear trend within the span of the filter, the asymmetric weights can only reproduce a constant for the only imposed constraint is that the weights add to one, see Doherty (1992). Important studies related to these kind of trend-cycle estimators have been made, among many others, by Pierce (1975), Burman (1980), Cleveland & Tiao (1976), Kenny & Durbin (1982), and Dagum & Bianconcini (2007).

4.3 LOESS filter

The LOESS estimator is based on nearest neighbor weights and is applied in an iterative manner for robustification. This filter consists of locally fitting polynomials of degree d by means of weighted least squares on a neighborhood of q observations around the estimated point. As q increases, the estimated trend $\hat{T}$ becomes smoother.

In general, LOESS is defined for not equally spaced observations, but for time series $X = \{x_n\}$ where each x_n is taken in time point χ_n. Moreover, the LOESS estimator exists not only at points $\{\chi_n\}$ but everywhere. This feature allows one to fill in the missing values if necessary. Let $\lambda_q(\chi)$ be the distance from χ to the qth outermost time point and introduce a *weight function* $W(x)$. Then for estimation of the trend at the point χ the regression weights for any χ_k are given by

$$w_k(\chi) = W\left(|\chi_k - \chi|\lambda_q^{-1}(\chi)\right).$$

The estimate is defined as $\hat{t}_n = \sum_{k=0}^{d} \hat{a}_k \chi_n^k$, where

$$\{\hat{a}_k\} = \arg\min_{a_k, 0\leqslant k\leqslant d} \left\{ \sum_{j=-m}^{m} w_j(\chi_n) \left[x_{n+j} - \sum_{k=0}^{d} a_k \chi_{n+j}^k \right]^2 \right\}.$$

The LOESS estimator is quite similar to the Henderson one, but the former (i) allows one to fit polynomials of degree d, (ii) is defined everywhere and (iii) does not impose the Henderson smoothing criterion for the weights.

The degree d of the fitting polynomial, the shape of the weight function $W(x)$, and the value of the smoothing parameter q are the three crucial choices to be made in LOESS.

Polynomials of degree $d = 1$ or $d = 2$ are generally suitable choices. The highest degree is more appropriate when the plot of observations against the target points presents many points of maximum and minimum. For this reason, in general, the flexibility of a quadratic fitting is preferred to the computational easiness of a linear one. As weighting function one can use the tricube proposed by Cleveland (1979) and defined by

$$W(t) = \left(1 - |t|^3\right)^3 I_{[0,1]}(t),$$

where $I_{[0,1]}(t)$ is the indicator function.

The ratio between the amplitude of the neighborhood q and the full span of the series N defines the *smoothing parameter*. It is sensible to choose an odd value for q in order to allow symmetric neighborhoods for central observations. A low smoothing parameter gives unbiased but highly variable estimates, while increasing its value reduces the variance but augments the bias. Therefore, in choosing the smoothing parameter, the aim is therefore to take a large q in order to minimize the variability in the smoothed points but without distorting the underlying trend.

The asymmetric weights of the filters are derived following the same technique by weighting the data belonging to an asymmetric neighborhood which contains the same number of data points of the symmetric one, as described by Gray & Thomson (1990). However, these authors showed that this implies a heavier than expected smoothing at the ends of the series with respect to the body, and represents a drawback, particularly for economic time series where turning points are important to identify.

4.4 Hodrick-Prescott filter

The Hodrick & Prescott (1997) filter follows the cubic smoothing spline approach. The framework used in Hodrick & Prescott (1997) is that a given time series X is the sum of a growth component T and a cyclical component C: $X = T + C$. The *measure of the smoothness* of the trend T is the sum of the squares of its second difference. The C are deviations from T and the conceptual framework is that over long time periods, their average is near zero. These considerations lead to the following programming problem of estimation of the trend T

$$\min_{\{t_n\}_{n=0}^{N-1}} \left\{ \sum_{n=0}^{N-1} (x_n - t_n)^2 + \lambda \sum_{n=0}^{N-1} \left[(t_n - t_{n-1}) - (t_{n-1} - t_{n-2})\right]^2 \right\}. \qquad (4)$$

The parameter λ is a positive number which penalizes variability in the growth component series. The larger the value of λ, the smoother is the solution series. For a sufficiently large λ, at the optimum all the $t_{n+1} - t_n$ must be arbitrarily near some constant β and therefore for t_n arbitrarily near $t_0 + \beta n$. This implies that the limit of the solution to (4) as λ approaches infinity is the least squares fit of a linear time trend model.

The Hodrick-Prescott (HP) filter was not developed to be appropriate, much less optimal, for specific time series generating processes. Rather, apart from the choice of λ, the same filter is intended to be applied to all series. Nevertheless, the smoother that results from the solution of eq. (4) can be viewed in terms of optimal signal extraction literature pioneered by Wiener (1949) and Whittle (1963) and extended by Bell (1984) to incorporate integrated time series generating processes. King & Rebelo (1993) and Ehglen (1998) analyzed the HP filter in this framework, motivating it as a generalization of the exponential smoothing filter. On the other hand, Kaiser & Maravall (2001) showed that under certain restriction the HP filter can be well approximated by a Integrated Moving Average model of order 2, whereas Harvey & Jaeger (1993) interpreted the HP filter in terms of structural time series models (see section 3.5). Several authors have analyzed shortcomings and drawbacks of the filter, concentrating on the stochastic properties of the estimated components induced by the filter. We refer to Ravn & Uhlig (1997) for a detailed summary.

4.5 Filters in Reproducing Kernel Hilbert Space

A different characterization of the nonparametric estimators previously introduced in this section can be provided using the Reproducing Kernel Hilbert Space (RKHS) methodology.

The main theory and systematic development of reproducing kernels and associated Hilbert spaces was laid out by Aronszajn (1950), who showed that the properties of RKHS are intimately bounded up with properties of nonnegative definite functions. A *RKHS is a Hilbert space characterized by a kernel that reproduces, via an inner product, every function of the space or, equivalently, by the fact that every point evaluation functional is bounded.* Loéve (1948) proved that there is an isometric isomorphism between the closed linear span of a second order stationary random process and the RKHS determined by its covariance function. Parzen (1959) was the first to apply this fundamental result to time series problems by means of a strictly parametric approach. Recently, reproducing kernel methods have been prominent as a framework for penalized spline methodology (see Wahba (1990)) and in the support vector machine literature, as described in Wahba (1999), Evgeniou et al. (2000), and Pearce & Wand (2006).

Dagum & Bianconcini (2006, 2007) have found reproducing kernels in Hilbert spaces of the Henderson and LOESS local polynomial regression predictors with particular emphasis on the asymmetric filters applied to the most recent observations. These authors show that the asymmetric filters can be derived coherently with the corresponding symmetric weights, or from a lower or higher order kernel within a hierarchy, if preferred. In the particular case of the currently applied asymmetric Henderson and LOESS filters, those obtained by means of the RKHS are shown to have superior properties relative to the classical ones from the view point of signal passing, noise suppression and revisions.

An important consequence of the RKHS theory is that nonparametric linear smoothers can be grouped into hierarchies with the following property: each hierarchy is identified by a density f_0 and contains estimators of order 2, 3, 4, ... which are products of orthonormal polynomials with f_0. The density function f_0 represents the second order kernel within the hierarchy, and provides the "initial weighting shape" from which the higher order kernels inherit their properties. Therefore, if f_0 *is optimal in a certain sense, each kernel of the hierarchy inherits the optimality property at its own order.* Estimators based on different assumptions of smoothing building can be compared by considering smoothers of different order within the same hierarchy as well as kernels of the same order, but belonging to different hierarchies. Filters of any length, including the infinite ones, can be derived in the RKHS framework. Therefore, for every estimator the density function f_0 and the corresponding reproducing kernel are derived. In this framework, the LOESS kernel hierarchy is based on the tricube density function f_{0T}. Higher order kernels are obtained via multiplication of f_{0T} by a linear combination of its corresponding orthonormal polynomials up to order two. These latter are derived using a determinantal expression based on the moments of the tricube density function f_{0T}, as shown in Dagum & Bianconcini (2006). On the other hand, Dagum & Bianconcini (2007) showed that the weight diagram of the Henderson smoother can be well-reproduced by two different density functions and corresponding orthonormal polynomials. These functions are the exact density derived by the penalty weight w_j given in eq. (3), and the biweight density function f_{0B}. The two density functions are very close to one another, hence the former can be well-approximated by the latter. One of the main advantages of this approximation is that the biweight density, and also the corresponding hierarchy, does not need to be calculated any time that the length of the filter changes, as happens for the exact probability function. Furthermore, f_{0B} belongs to the well-known Beta distribution family, and the corresponding orthonormal polynomials are the Jacobi ones, for which explicit expressions for computation are available and their properties have been widely studied in the literature.

The hierarchies, here considered, reproduce and describe several temporal dynamics by estimating polynomial trends of different degrees that solve several minimization problems, and we refer to Dagum & Bianconcini (2006, 2007) for a theoretical study of their properties by means of Fourier analysis.

4.6 Software

The Henderson filter is available in nonparametric seasonal adjustment software such as the X11 method developed by the U.S. Census Bureau (Shiskin et al., 1967), and its variants X-11-ARIMA (Dagum, 1980) and X-12-ARIMA (Findley et al., 1998). The LOESS filter is implemented in STL (Cleveland et al., 1990). Software implementing RKHS variants of Henderson and LOESS filters is available upon request. The Hodrick-Prescott filter can be found in the most widely used statistical packages, such as Eviews, Stata, S-plus, R, Matlab, SAS.

5 Singular Spectrum Analysis

5.1 Preamble

In this section, we consider the use of Singular Spectrum Analysis (SSA) for trend extraction. This approach is based on building some matrix from a time series and on operating with the Singular Value Decomposition of this matrix. Based on the information provided by singular vectors, a matrix approximation is obtained and then it is converted into an additive component of the time series. Apart from the transformation of a time series to a matrix and vice-versa, the algorithm of SSA coincides with the procedure of Principal Component Analysis (Danilov, 1997). Sometimes SSA is referred to as the Karhunen-Loéve decomposition of time series (Basilevsky & Hum, 1979).

SSA originated between the late 70s and early 80s, mainly in the area of dynamical systems as the result of Bertero & Pike (1982); Broomhead & King (1986); Fraedrich (1986). The name Singular Spectrum Analysis was introduced by Vautard & Ghil (1989), but this approach is also referred to as the *Caterpillar approach*. For historical surveys see Golyandina et al. (2001) and Ghil et al. (2002). The ideas of SSA appeared in other areas, such as digital signal processing (Kumaresan & Tufts, 1980) or oceanology (Colebrook, 1978). The present literature on SSA includes two monographs (Elsner & Tsonis, 1996; Golyandina et al., 2001), three book chapters (Schreiber, 1998; Vautard, 1999; Alonso et al., 2004), and over a hundred papers.

Singular Spectrum Analysis can be used in a wide range of issues: trend or periodical component extraction, denoising, forecasting, and change-point detection. At the present time, SSA is a proven technique in the geosciences (Ghil & Vautard, 1991; Ghil et al., 2002), and it is starting to be applied in other areas, e.g., biology (Alonso et al., 2004), tomography (Pereira et al., 2004), material processing (Salgado & Alonso, 2006) and nuclear science (Verdu & Ginestar, 2001).

5.2 Basic algorithm and general questions

The basic algorithm of SSA consists of two parts: *decomposition* of a time series and *reconstruction* of a desired additive component (e.g., trend). At the stage of decomposition, we choose only the *window length*, denoted by L, and we construct a trajectory matrix $\mathbf{X} \in \mathbb{R}^{L \times K}$, $K = N - L + 1$, with stepwise portions of the time series X taken as columns:

$$X = (x_0, \ldots, x_{N-1}) \quad \rightarrow \quad \mathbf{X} = [X_1 : \ldots : X_K], \quad X_j = (x_{j-1}, \ldots, x_{j+L-2})^{\mathrm{T}}.$$

Then we perform the Singular Value Decomposition (SVD) of $\mathbf{X}$ where the jth component of SVD is described by an eigenvalue λ_j and a real-valued eigenvector U_j of $\mathbf{X}\mathbf{X}^{\mathrm{T}}$:

$$\mathbf{X} = \sum_{j=1}^{L} \sqrt{\lambda_j} U_j V_j^{\mathrm{T}}, \quad V_j = \mathbf{X}^{\mathrm{T}} U_j / \sqrt{\lambda_j}.$$

The SVD components are numbered in the decreasing order of their eigenvalues. The reconstruction stage combines (i) selection of a group $\mathcal{J}$ of several SVD components and (ii) reconstruction of a trend by Hankelization (averaging through anti-diagonals) of the matrix formed from the selected part $\mathcal{J}$ of the SVD:

$$\sum_{j\in\mathcal{J}} \sqrt{\lambda_j} U_j {V_j}^{\mathrm{T}} \quad \rightarrow \quad \text{estimation of } \ T.$$

For the complete description of the algorithm see Golyandina et al. (2001).

The problem of trend extraction in SSA is reduced to (i) the choice of window length L and (ii) the selection of the group $\mathcal{J}$ of SVD components. The former problem had no reasonable solution before Nekrutkin (1996) showed how the quality of SSA decomposition depends on L. His separability theory provides instructions for choosing L according to the properties of assumed components of a time series, such as trend, periodical components, and noise (Golyandina et al., 2001, Chapter 6). The selection of SVD components is the major task in SSA and at the present time there exist several methods realizing trend extraction.

5.3 Trend in SSA

The SSA approach is essentially a nonparametric approach and does not need a priori specification of a model for a time series or for a trend, neither deterministic nor stochastic one. The classes of trends and residual which can be successfully separated by SSA are characterized as follows.

Firstly, since we extract a trend by selecting a subgroup of all L SVD components, it should generate only d ($d < L$) of them. For infinite time series, the class of such trends coincides with the class of time series governed by finite difference equations (Golyandina et al., 2001). This class can be described explicitly as linear combinations of products of polynomials, exponentials and sines (Buchstaber, 1995). An element of this class approximates well a smooth time series and is not suitable for the approximation of a swiftly changing time series.

Secondly, a residual should belong to the class of time series which can be separated from a trend. The separability theory allows to define this class and postulates that (i) every deterministic function can be asymptotically separated from any ergodic stochastic noise (Nekrutkin, 1996; Golyandina et al., 2001) as the time series length and the window length tends to infinity; (ii) under some conditions a trend can be separated from a quasi-periodic component. These properties of SSA allow one to apply this approach to extraction of trend in the presence of noise and quasi-periodic components.

Notice that SSA takes into account the information about the whole time series, for it considers the SVD of the trajectory matrix built from all parts of the time series. Therefore SSA is not a local method – in contrast to the linear filtering or wavelet methods. On the other hand, this property makes SSA robust to outliers.

5.4 Methods of trend extraction in SSA

The naïve idea of SVD components selection for trend extraction is to take several of the first SVD components. This simple approach works in many real-life cases, given the optimal properties of SVD (Golyandina et al., 2001, chapter 4). An eigenvalue represents the contribution of the corresponding SVD component into the form of the trajectory matrix and of the original time series, respectively. Since a trend usually characterizes the time series, its eigenvalues are larger than the other ones, which implies small order numbers for the trend SVD components. However, the selection procedure fails when the values of a trend are small as compared with the residual (Golyandina et al., 2001, Section 1.6). Note that this approach takes into account only eigenvalues but not eigenvectors.

A more clever way of selection of trend SVD components is to choose the components with smooth Empirical Orthogonal Functions (EOFs), where the nth EOF is defined as the sequence of elements of the nth eigenvector. This approach was presented in Golyandina et al. (2001), where the cases of polynomial and exponential trends are thoroughly examined. Using the concept of trajectory vector space which is spanned by the columns of the trajectory matrix and has the eigenvectors as an orthonormal basis, one can prove that the smoothness of a trend controls the smoothness of its EOFs on the assumption of separability of the trend and the residual.

The methods following this approach are presented in Vautard et al. (1992); Golyandina et al. (2001); Salgado & Alonso (2006) and Alexandrov (2006). Golyandina et al. (2001) proposed to select trend SVD components by visual examination of EOFs. Alexandrov (2006) presented a parametric method which follows the frequency approach to smoothness and exploits properties of the Fourier decomposition of EOFs. Earlier, in Vautard et al. (1992), another parametric method was described which is based on the Kendall correlation coefficient, but for the properties of this coefficient this method is aimed at extraction of monotonous trends. An original modification of SSA for producing smooth trends was proposed in Solow & Patwardhan (1996). Instead of calculating the eigensystem of $\mathbf{X}\mathbf{X}^{\mathrm{T}}$, the authors considered some special matrix depending on the first differences of a time series.

5.5 Pros and cons

SSA is a model-free approach that provides good results for short time series (Vautard et al., 1992), and allows one to extract trends from a wide class of time series. An essential disadvantage of SSA is its computational complexity in the calculation of the SVD. This cost can be reduced by using parallel computing (Jessup & Sorensen, 1994). For updating the SVD in the case of receiving new data points (trend revision), a computationally attractive algorithm of Gu & Eisenstat (1993) can be used. Moreover, Drmač & Veselić (2005) recently proposed a new method for SVD calculation which is as fast as QR-factorization and as stable as the conventional Jacobi method.

5.6 Software

The main software packages for trend extraction implementing SSA include CaterpillarSSA (Golyandina et al., 2001), SSA-MTM Toolkit (Vautard et al., 1992), AutoSSA (Alexandrov, 2006) and kSpectra Toolkit, which allows one to apply SSA on many computing platforms. CaterpillarSSA provides an interactive framework for time series processing and can be used for trend extraction and forecast. SSA-MTM Toolkit implements only the Kendall method, and kSpectra Toolkit represents the commercial version of SSA-MTM Toolkit. AutoSSA for Windows realizes three parametric methods (in particular, the Low Frequencies (LF) and Kendall methods); AutoSSA for Matlab has only the LF method but with adaptive selection of the parameters. Moreover, SSA is available in a collection of scripts and packages; for details see the website SSAwiki (*http://www.math.uni-bremen.de/∼theodore/ssawiki*).

6 Wavelets

6.1 Preamble

The name *wavelet* first appeared in the early 1980's in the context of seismic data analysis. Earlier, the term was introduced to the general scientific community in a pioneering paper (Grossmann & Morlet, 1984) jointly written by a geophysicist, a theoretical physicist and a mathematician. This combination of different points of view has demonstrated the practical importance of the theoretical findings since the beginning and has sparked the rapid development of wavelet analysis in the subsequent years.

However, a closer look at mathematical history (Meyer, 1993) reveals that several almost identical approaches – or at least similar concepts – have been around since the 1930's. A first wavelet construction can be found in several investigations of suitable model spaces for signals and functions (Littlewood-Payley theory). In addition, the Calderon's identity – or more recently the so-called pyramidal algorithms (Burt & Adelson, 1983) – share some features with wavelet methods.

Nevertheless, the early input from diverse neighboring scientific fields created a rich theoretical framework, which has led both to algorithms that are mathematically justified as well as to theoretical generalizations. These generalized concepts exceed the previous approaches by far, particularly in terms of their potential for different applications in signal and image processing.

The term *wavelet analysis* is currently used for the somewhat larger field of multiscale analysis, with both its theoretical, mathematical foundations and its resulting algorithms in signal and image processing. Wavelet analysis in general rests on a formal framework for decomposing a signal or function in its different components on different "scales". The scales can be distinguished either by different levels of resolution or different sizes/scales of detail. In this sense it generates what is commonly called a phase space decomposition, where the phase space is defined by two parameters (scale, time/location). It is a counterpart to classical Fourier or Gabor decompositions, which generate a phase space decomposition defined via a frequency and time/location parametrization.

On a purely discrete level, every wavelet algorithm is defined by a dual pair (low pass, band pass) of Finite Impulse Response (FIR) filters, which allows a decomposition and reconstruction by symmetric convolution operators. Its efficient implementations (Daubechies, 1992; Sweldens, 1997; Mallat, 2001) as well as its flexibility have led to several outstanding applications, including the design of efficient image and video compression standards (JPEG2000, MPEG) and advanced audio technology (Kowalski & Torresani, 2006). In addition, the analysis of wavelet methods in a statistical framework has led to some powerful methods, e.g., for denoising signals and images (Donoho & Johnstone, 1994; Donoho, 1995). Wavelet methods are now a generally accepted alternative to more classical statistical approaches or filtering techniques in a wide range of applications – see Chui (1992); Vedam & Venkatasubramanian (1997); Maass et al. (2003); Partal & Kuecuek (2006) and the citations therein. The most recent applied developments are regularly reported in several series of specialized conferences, e.g., SPIE, IEEE conferences.

6.2 Basic algorithms and general considerations

The basic wavelet algorithm computes a decomposition (wavelet transform) of a time series with a pair of FIR filters in several steps. Let us introduce some additional notation and call the original time series $c^0 = \{c^0_k \mid k = 0, 1, \ldots, N-1\}$. In a first step, the data c^0 is convolved with a low pass $\{h_k\}$ and a band pass filter $\{g_k\}$. The results of both filters are then subsampled by factor of 2 in order to reduce complexity:

$$c^1_k = Hc^0 = \sum_{\ell=-\infty}^{\infty} h_\ell c^0_{2k-\ell} \ , \quad d^1_k = Gc^0 = \sum_{\ell=-\infty}^{\infty} g_\ell c^0_{2k-\ell} \ .$$

This first step separates the components on a finest level of resolution. The result of the band pass filter d^1 is stored as the fine scale component of the original time series. The low pass filter generates a smoothed version c^1 of the original time series, which is further processed iteratively. In a second step, all computations of the first step are repeated on a low pass filtered version c^1. The outcome of the subsampled band pass filtering in the second step is again stored, it contains the details of the original time series on scale two. The application of the subsampled low pass filter gives an even smoother version of the original time series.

Repeating this process for a fixed number of s steps produces a family of sequences, which represent details on different scales $\{d^j, j = 1, \ldots, s\}$ as well as a final version c^s of the original sequence c^0 that is very smooth:

$$c^{j+1}_k = Hc^j = \sum_{\ell=-\infty}^{\infty} h_{2k-\ell} c^j_\ell \ , \quad d^1_k = Gc^j = \sum_{\ell=-\infty}^{\infty} g_{2k-\ell} c^j_\ell \ , \quad j = 0, 1, \ldots, s-1.$$

Notice that due to the subsampling the overall number of coefficients to be stored (it is the sequences d^j, $j = 1, \ldots, s$, and c^s) is equivalent to the length of the original time series.

The success of wavelet methods relies on the existence of a dual pair of low and band pass filters for reconstruction. By a symmetric algorithm, these dual filters allow an equally efficient reconstruction of the original time series from its multi-scale decomposition:

$$c_k^j = \tilde{H}c^{j+1} + \tilde{G}c^{j+1} = \sum_{\ell=-\infty}^{\infty} h_{k-2\ell} c_\ell^{j+1} + \sum_{\ell=-\infty}^{\infty} g_{k-2\ell} d_\ell^{j+1} \ , \quad j = s-1, s-2, \ldots, 0.$$

It results in an additive decomposition of the original data by

$$c^0 = \tilde{H}^s c^s + \sum_{j=1}^{s} H^{j-1} G d^j \ .$$

The filters need to satisfy some stability criteria in order to control reconstruction errors. The choice of an appropriate filter bank for decomposition and reconstruction is crucial for the success of wavelet methods. There exists an extensive library of wavelet filters that are appropriate for all kinds of applications.

In principle, a wavelet algorithm is fully defined by the choice of the wavelet filters and the number of decomposition steps s. However, we want to emphasize that before using a wavelet algorithm one should answer the two basic questions: (i) why should we use a wavelet method? (ii) how do we want to analyze the wavelet decomposition?

The first question can be answered positively whenever the time series under consideration has a multi-scale (as opposed to a multi-frequency) structure. This in particular includes the analysis of non-stationary effects, e.g., defects in signals for monitoring technical processes or applications of change-of-trend detection.

The second question is extremely important. The wavelet transform only generates an alternative representations of the time series: no information is lost, no information is added by the wavelet transform. For certain applications, however, we might expect that any sought-after information can be more easily detected in the transformed data. Hence, at the beginning we need to consider how we want to extract this information after having computed the wavelet transform.

6.3 Trend extraction with wavelet methods

Following the general considerations described in the previous subsection, we first need to determine why wavelet methods should be useful for trend analysis. In this section we follow the classical decomposition model (see section 1)

$$X = T + P + R$$

with trend component T , seasonal component P and noise-component R. This model is well-suited for a wavelet decomposition, which results in an additive multi-scale decomposition $\{c^s, d^j, j = 1, .., s\}$ of the time series X. An increasing "scale of detail" is assigned to every component of the decomposition, i.e., we interpret the different components as being the sum of all details in the time series which live on a prescribed scale (resolution or size of detail).

Therefore, a typical wavelet decomposition concentrates the noise component on the first fine scales, then the seasonal components are detected on the subsequent intermediate scales, and finally the trend component is given on the coarse scale components. However, notice that the notion of seasonal component is more general in a wavelet setting, since it does not only refer to purely periodic components, but also models variations and non-stationary seasonality.

Hence, the basic trend extraction procedure with wavelet methods proceeds by (i) choosing an appropriate wavelet filter bank, (ii) computing a wavelet decomposition up to scale s, (iii) deleting all fine scales (scales of noise and seasonal components), and (iv) reconstructing the remaining additive component:

$$T \sim \tilde{H}^s c^s.$$

The seasonal component is determined by the components on the intermediate scales, $d^2, .., d^s$, and the noise is approximated by the difference coefficients d_1 on the finest scale.

$$P \sim \sum_{j=2}^{s} \tilde{H}^{j-1} \tilde{G} d^j \ , \quad R = \tilde{G} d^1.$$

Some of the most prominent applications for trend extraction by wavelet methods include process monitoring of technical processes (Vedam & Venkatasubramanian, 1997; Bakhtazad et al., 2000), analysis of environmental data (Tona et al., 2005; Partal & Kuecuek, 2006) or applications to financial data (Maass et al., 2003). This list is neither complete nor representative, but rather serves to demonstrate some of basic examples presented in the vast literature on wavelet trend analysis.

This typically results in a good "visualization" of the underlying non-stationary trend. In this sense, wavelet methods use a semi-parametric trend model: the choice of the wavelet filters determines the trend model, since it determines whether we capture piecewise constant, linear, polynomial or exponential trends – see Bakshi & Stephanopoulos (1994) or the general references on polynomial reproduction by wavelet basis in Louis et al. (1997); Mallat (2001). On the other hand, the choice of the wavelet is only important for the intermediate computations. After reconstruction we obtain a trend model in the physical space given by the measurement data. The trend can be subsequently analyzed without any underlying model.

We want to emphasize that applying a wavelet method usually constitutes just one step in a more complex scenario for trend analysis. Typical tasks, such as change point detection, require one to analyze the extracted trend and to give precise estimates for the time instances of change-of-trend. There exist refined wavelet methods based on shrinkage operations for this kind of analysis. Change-of-trend features exist on all scales of resolution, hence these methods use the full wavelet decomposition and rely on adaptive thresholding procedures on all scales (Vedam & Venkatasubramanian, 1997; Mallat, 2001; Partal & Kuecuek, 2006).

A different scenario for the application might require the determination of a physical model for the underlying process. In this case, the trend extraction and the determination of different time intervals with a stationary behavior are only a first step. Hence, the characterization of the model on each time interval with a stable trend is then left to other methods of signal analysis (e.g., dynamical systems or the methods following MBA, see section 3).

To conclude, the potential of wavelet methods for trend extraction is based on its non-stationary, quasi-local properties. Wavelet methods are well-suited for the determination of change-of-trend points, as well as the decomposition of the time axis in different time intervals with stable trend behavior.

6.4 Advantages and disadvantages

The application of a wavelet algorithm starts by choosing an appropriate wavelet basis or wavelet filter bank. This offers some flexibility for optimization, since there exist highly specialized wavelet filters for a large variety of complex situations. Its main feature is a non-stationary multi-scale decomposition, which is particularly suited for analyzing localized effects. This flexibility, which allows one to finely tune the wavelet algorithms to different specific tasks, is also one of its major disadvantages; some experience with wavelet methods is required in order to fully exploit its power.

In addition, the treatment of boundary effects is crucial for some applications. Wavelet algorithms are based on a repeated application of linear filters, which require an adjustment at the end of the given data series. This could be done by adapting the filter coefficients, even if the standard approach is to add a sufficient amount of data points. Several algorithms generally use either a zero-padding or a periodic continuation. Both approaches yield acceptable results for most applications of trend detection, in particular if one uses a short wavelet filter. However, optimal results are achieved by fine-tuning the treatment of boundary effects depending on whether one wants to determine a polynomial or exponential trend. A simple least square fit of a polynomial to the last given data points allows one to determine a more suitable continuation.

Also, the visualization of wavelet transforms has not yet been fully standardized. Again, some experience is required in order to "understand" the results of the wavelet transform. Typically, the best way to analyze a wavelet transform is to reconstruct the manipulated wavelet decomposition and to display the result in the physical domain of the original signal.

The major advantage is the efficiency of the fast wavelet decompositions. In its basic form the wavelet transform is an $\mathcal{O}(n)$-algorithm. Efficient implementations are by now included in any software toolbox for signal or image analysis, e.g., MATLAB or S+.

7 Example

Let us consider an application of the described methods to observations of the electric power use by industry in the U.S. for the period from 1972/1 to 2005/10 provided by the Federal Reserve Board (FRB). The data is monthly of length 406, and is available on the FRB webpage: *http://www.federalreserve.gov/releases/g17/ipdisk/kwh_nsa.txt*.

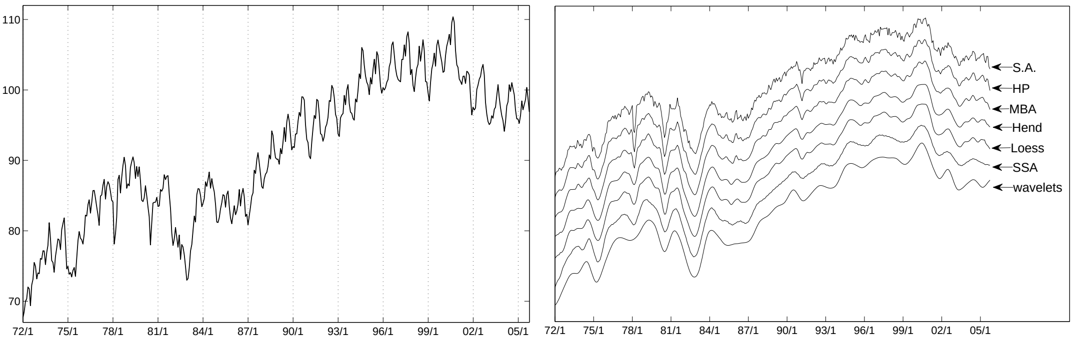

Figure 1: Left panel: electric power use by industry in US, monthly data; Right panel: seasonally adjusted time series (S.A.) and the trends (each shifted from the previous one).

We selected this time series because: (i) it contains a clear and complex trend, (ii) we can test trend extraction in the presence of a sizeable seasonal component, (iii) the length of several hundred points is usual for many applications, and (iv) noise is significant enough to demonstrate the smoothing properties of the various methods.

The time series has clear seasonality. MBA, SSA and wavelets are able to extract trend from data of such kind, while nonparametric filtering methods (Henderson, LOESS, Hodrick-Prescott) require seasonally adjusted data. Fortunately, the Federal Reserve Board also provides seasonally adjusted time series, and we have applied the nonparametric methods to these data. The resulting trends are shown in Figure 1.

MBA The regARIMA modeling of X-12-ARIMA software was used to identify regression parameters and significance, and determine the best model using AICC. This was a (1 1 0)(0 1 1) SARIMA model, and the canonical decomposition into trend, seasonal, and irregular exists. The minimum MSE trend extraction filter for finite sample (McElroy, 2008) was determined and applied.

SSA We exploited parametric methods implemented in AutoSSA software for Matlab, available at the webpage: *http://www.pdmi.ras.ru/∼theo/autossa*. A trend can be extracted directly from the original time series, but we first performed seasonal adjustment and then extracted the trend, both by AutoSSA. Such an iterative application of AutoSSA is reasonable because its methods are quite restrictive and repeated application can refine the results. For seasonal adjustment we used the Fourier method for extraction of periodical components (Alexandrov, 2006). Window length was selected close to $N/2$ and divisible by seasonal period (12), $L = 12\lfloor N/24 \rfloor = 192$. The trend was extracted by the Low Frequencies method, where the low frequencies boundary was selected smaller but close to 1/12, equal to 0.07.

Wavelets For trend extraction by means of wavelets, the Coifman wavelet of order 4 (*coif4*) was selected given its symmetry and good smoothing properties. After wavelet transformation, a trend was reconstructed by all wavelet coefficients excepting only detail coefficients on levels 1 and 2. As a multiscale approach, wavelet transformation allows one to extract trends of different resolution (or scale). The extracted trend seems to contain some insignificant portions of the seasonal component opposite to a trend reconstructed without details on levels 1, 2 and 3. Nevertheless, we selected the former one because it better represents our point of view on the sought-for trend-cycle.

Nonparametric Notice that for nonparametric filtering we used seasonally adjusted time series. The length of Henderson and LOESS filters was selected according to the signal-to-noise ratio (provided by X-11-ARIMA software), which is equal to 1.14. Hence, a 13-term filter is appropriate for the estimation of the trend-cycle. The trend estimates were obtained based on the following RKHS filters: (i) 13-term 3rd order Henderson kernel within the biweight hierarchy; (ii) 13-term 3rd order LOESS kernel within the tricube hierarchy. For the Hodrick-Prescott trend we applied the *pspline* package of R, with the smoothing parameter selected by means of generalized cross-validation.

7.1 Comparative analysis of the trends

As the quality of extraction of a prior unknown trend is hard to evaluate, we do not search for the best method but rather seek to evaluate some of their features. Firstly, the LOESS estimator is quite close to the Henderson trend almost everywhere, even if at the ends it represents the shifted Henderson trend (plus $\approxeq 0.3$); hence, hereafter we do not consider the LOESS trend.

Capturing the details With respect to captured details the trends can be organized as follows: Hodrick-Prescott (detailed), MBA, Henderson, SSA and wavelets (coarse). From our point of view, the Hodrick-Prescott trend is not smooth enough, which contradicts the definition of the trend as a smooth component – see Figure 2. These resulting trends are not unique for the considered methods. In general, by specifying different parameters, one can change the resolution. For example, in wavelets/AutoSSA, the degree of selected scale levels/value of low-frequencies-boundary controls the resolution. But in this example including more scales in wavelets leads to inclusion of a portion of the seasonal component into the trend. On the other hand, selecting a larger low-frequencies-boundary in AutoSSA reduces the smoothness of the trend considerably.

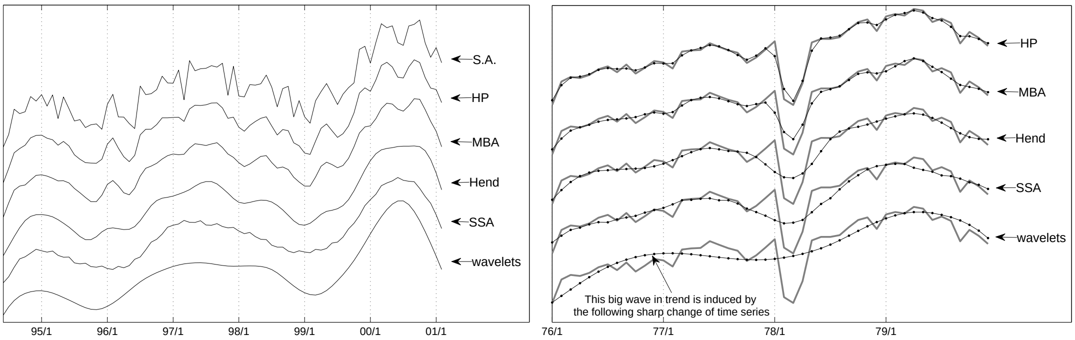

Figure 2: Left panel: The seasonally adjusted time series (S.A.) and the trends (each shifted from the previous one), points 270-350; Right panel: the trends (each shifted from the previous one) (dotted line) with the seasonally adjusted time series in background (gray bold line), points 49-96.

Boundary effect Notice that in this example the wavelet trend goes up close to the end, which demonstrates the boundary effect discussed in section 6.4.

Influence of sharp changes Finally, we examine how sharp changes of time series influence behavior of the trend. Usually, it is desirable to have trend robust to such changes. We consider a interval of points [49-96], including such a change in the beginning of 1978 – see Figure 2. Hodrick-Prescott and MBA trends track the dip and one can argue about whether a trend should contain it. Henderson and SSA trends smooth out the dip. The wavelet trend is distorted around the sharp changes, because their impact spreads over all levels of wavelet coefficients and affect the trend coefficients.

Acknowledgments

It is a pleasure to thank Nina Golyandina for stimulating discussions on SSA and helpful remarks.

References

Alexandrov, T. (2006). Batch extraction of additive components of time series by means of the “Caterpillar”-SSA method. *Vestnik St. Petersburg Univ.: Math.*, **39**, 112–114.

Alonso, F. J., María Del Castillo, J., & Pintado, P. (2004). An automatic filtering procedure for processing biomechanical kinematic signals. *Biol. and Med. Data Analysis*, pp. 281–291. Springer.

Anderson, B. & Moore, J. (1979). *Optimal Filtering.* Englewood Cliffs: Prentice-Hall.

Aronszajn, N. (1950). Theory of reproducing kernels. *Transaction of the AMS*, **68**, 337–404.

Bakhtazad, A., Palazoglu, A., & Romagnoli, J. (2000). Process trend analysis using wavelet-based de-noising. *Control Eng. Pract.*, **8**(6), 657–663.

Bakshi, B. & Stephanopoulos, G. (1994). Representation of process trendsiii. multiscale extraction of trends from process data. *Comput. Chem. Eng.*, **18**(4), 267–302.

Basilevsky, A. & Hum, D. P. J. (1979). Karhunen-loeve analysis of historical time series with an application to plantation births in Jamaica. *J. Am. Stat. Assoc.*, **74**, 284–290.

Bell, W. (1984). Signal extraction for nonstationary time series. *Ann. Stat.*, **12**, 646–664.

Bell, W. (2004). On RegComponent time series models and their applications. *State Space and Unobserved Component Models: Theory and Applications*, Eds. A. C. Harvey, S. J. Koopman, & N. Shephard. Cambridge, UK: Cambridge University Press.

Bell, W. & Hillmer, S. (1984). Issues involved with the seasonal adjustment of economic time series. *J. Bus. Econ. Stat.*, **2**, 291–320.

Bell, W. & Hilmer, S. (1988). *A Matrix Approach to Likelihood Evaluation and Signal Extraction for ARIMA Component Time Series Models.* Tech. report, $RR-88/22$. U.S. Census Bureau.

Bell, W. & Hilmer, S. (1991). Initializing the Kalman filter for nonstationary time series models. *J. Time Ser. Anal.*, **12**, 283–300.

Berryman, D., Bobee, B., Cluis, D., & Haemmerli, J. (1988). Nonparametric tests for trend detection in water quality time series. *Water Resour. Bull.*, **24**(3), 545–556.

Bertero, M. & Pike, E. R. (1982). Resolution in diffraction-limited imaging, a singular value analysis I. the case of coherent illumination. *Optica Acta.*, **29**, 727–746.

Beveridge, S. & Nelson, C. (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention ro measurement of the business cycle. *J. Monetary Econ.*, **7**, 151–174.

Brockwell, P. J. & Davis, R. A. (2003). *Introduction to time series and forecasting.* Springer.

Broomhead, D. S. & King, G. P. (1986). Extracting qualitative dynamics from experimental data. *Physica D*, **20**, 217–236.

Buchstaber, V. M. (1995). Time series analysis and grassmannians. *Amer. Math. Soc. Trans*, volume **162**, pp. 1–17. AMS.

Burman, J. (1980). Seasonal adjustment by signal extraction. *J. R. Stat. Soc. Ser. A-G*, **143**, 321–337.

Burt, P. & Adelson, E. (1983). The Laplacian pyramid as a compact image code. *IEEE Trans. Commun.*, **31**(4), 532–540.

Chatfield, C. (1996). *The analysis of time series: An introduction.* Chapman & Hall/CRC.

Chui, C., Ed. (1992). *Wavelets: A tutorial in theory and applicaitons.* NY: Academic Press.

Cleveland, R., Cleveland, W., McRae, J., & Terpenning, I. (1990). STL: A seasonal trend decomposition procedure based on LOESS. *Journal of Official Statistics*, **6**(1), 3–33.

Cleveland, W. (1979). Robust locally regression and smoothing scatterplots. *J. Am. Stat. Assoc.*, **74**, 829–836.

Cleveland, W. & Tiao, G. (1976). Decomposition of seasonal time series: A model for the Census X-11 program. *J. Am. Stat. Assoc.*, **71**, 581–587.

Colebrook, J. M. (1978). Continuous plankton records — zooplankton and evironment, northeast Atlantic and North Sea, 1948–1975. *Oceanol. Acta.*, **1**, 9–23.

Dagum, E. (1980). *The X11ARIMA seasonal adjustment method.* Ottawa, Statistics Canada Publication, Catalogue No. 12-564E.

Dagum, E. & Bianconcini, S. (2006). Local polynomial trend-cycle predictors in rkhs for current economic analysis. *Anales de Economia Aplicada*, , pp. 1–22.

Dagum, E. & Bianconcini, S. (2007). The Henderson smoother in reproducing kernel hilbert space. *Journal of Business and Economic Statistics.* (forthcoming).

Danilov, D. L. (1997). Principal components in time series forecast. *J. Comp. Graph. Stat.*, **6**, 112.

Daubechies, I. (1992). *Ten lectures on wavelets.* SIAM, Philadelphia.

De Jong, P. (1991). The diffuse Kalman filter. *Ann. Stat.*, **19**, 1073–1083.

Doherty, M. (1992). *The Surrogate Henderson Filters in X-11.* Tech. report, Stat. New Zealand.

Donoho, D. (1995). Denoising by soft-thresholding. *IEEE Trans. Inf.Theory*, **41**(3), 613–627.

Donoho, D. & Johnstone, I. (1994). Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, **81**(3), 425–455.

Drmač, Z. & Veselić, K. (2005). *New fast and accurate Jacobi SVD algorithm: I, II.* Tech. Report LAPACK Working Note 169, Dep. of Mathematics, University of Zagreb, Croatia.

Durbin, J. & Koopman, S. J. (2001). *Time series analysis by state space methods*, volume **24** of *Oxford Statistical Science Series.* Oxford: Oxford University Press.

Ehglen, J. (1998). Distortionary effects of the optimal Hodrick-Prescott filter. *Econ. Lett.*, **61**, 345–349.

Elsner, J. B. & Tsonis, A. A. (1996). *Singular Spectrum Analysis: A New Tool in Time Series Analysis.* Plenum.

Esterby, S. R. (1996). Review of methods for the detection and estimation of trends with emphasis on water quality applications. *Hydrol. Process.*, **10**, 127–149.

Evgeniou, T., Pontil, M., & Poggio, T. (2000). Regularization networks and support vector machines. *Advanced in Computational Mathematics*, **13**, 1–50.

Findley, D. & Martin, D. (2006). Frequency domain analyses of SEATS and X12ARIMA seasonal adjustment filters for short and moderate-length time series. *J.Off.Stat.*, **22**, 1–34.

Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C., & Chen, B. C. (1998). New capabilities and methods of the X-12-ARIMA seasonal adjustment program. *J. Bus. Econ. Stat.*, **16**(2), 127–177. (with discussion).

Fraedrich, K. (1986). Estimating dimensions of weather and climate attractors. *J. Atmos. Sci.*, **43**, 419–432.

Froeb, L. & Koyak, R. (1994). Measuring and comparing smoothness in time series: the production smoothing hypothesis. *J. Econometrics*, **64**(1-2), 97–122.

Gersch, W. & Kitagawa, G. (1983). The prediction of time series with trends and seasonalities. *J. Bus. Econ. Stat.*, **1**(3), 253–264.

Ghil, M., Allen, R. M., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., Robertson, A., Saunders, A., Tian, Y., Varadi, F., & P., Y. (2002). Advanced spectral methods for climatic time series. *Rev. Geophys.*, **40**(1), 1–41.

Ghil, M. & Vautard, R. (1991). Interdecadal oscillations and the warming trend in global temperature time series. *Nature*, **350**, 324–327.

Golyandina, N. E., Nekrutkin, V. V., & Zhigljavsky, A. A. (2001). *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman&Hall/CRC.

Gray, A. & Thomson, P. (1990). Comments on STL: A seasonal trend decomposition procedure based on LOESS. *Journal of Official Statistics*, **6**, 47–55.

Gray, A. & Thomson, P. (1996). Design of moving-average trend filters using fidelit and smoothness criteria. *Time series analysis in memory of E.J. Hannan*, , pp. 205–219.

Grossmann, A. & Morlet, J. (1984). Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM J. of Math. Anal.*, **15**(4), 723–736.

Gu, M. & Eisenstat, S. C. (1993). *A Stable and Fast Algorithm for Updating the Singular Value Decomposition*. Tech. Report YALEU/DCS/RR-966, Dep. of Computer Science, Yale University.

Hamed, K. & Rao, A. (1998). A modified Mann-Kendall trend test for autocorrelated data. *J. Hydrol.*, **204**(1), 182–196.

Hamilton, J. D. (1994). *Time series analysis*. Princeton.

Harvey, A. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey, A. & Jaeger, A. (1993). Detrending, stylized facts and the business cycle. *Journal of applied econometrics*, **8**, 231–247.

Henderson, R. (1916). Note on graduation by adjusted average. *Trans. Actuar. Soc. Amer.*, **17**, 43–48.

Hillmer, S. & Tiao, G. (1982). An ARIMA-model-based approach to seasonal adjustment. *J. Am. Stat. Assoc.*, **77**(377), 63–70.

Hirsch, R. M. & Slack, J. R. (1984). Nonparametric trend test for seasonal data with serial dependence. *Water Resources Research*, **20**(6), 727–732.

Hodrick, R. & Prescott, E. (1997). Postwar u.s. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, **29**(1), 1–16.

Jessup, E. R. & Sorensen, D. C. (1994). A parallel algorithm for computing the singular value decomposition of a matrix. *SIAM J. Matrix Anal. Appl.*, **15**(2), 530–548.

Kaiser, R. & Maravall, A. (2001). *Measuring Cycles in Economic Statistics*. Lecture Notes in Statistics, NY: Springer-Verlag.

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *J. Basic Eng-t ASME D*, **82**, 35–45.

Kenny, P. & Durbin, J. (1982). Local trend estimation and seasonal adjustment of economic and social time series. *Journal of the Royal Statistical Sociaety A*, **145**, 1–41.

King, R. & Rebelo, S. (1993). Low frequency filtering and real business cycles. *Journal of Economic Dynamics and Control*, **17**, 207–233.

Kitagawa, G. & Gersch, W. (1996). *Smoothness Priors Analysis of Time Series*. Springer.

Kolmogorov, A. N. (1939). Sur l'interpretation et extrapolation des suites stationnaires. *C.R. Acad. Sci. Paris*, **208**, 2043–2045.

Kolmogorov, A. N. (1941). Interpolation and extrapolation von stationaeren zufaelligen Folgen. *Bull. Acad. Sci. U.R.S.S. Ser. Math.*, **5**, 3–14.

Koopman, S. (1997). Exact initial Kalman filtering and smoothing for nonstationary time series models. *J. Am. Stat. Assoc.*, **92**, 1630–1638.

Koopman, S., Harvey, A., Doornik, J., & Shepherd, N. (2000). *Stamp 6.0: Structural Time Series Analyser, Modeller, and Predictor*. London: Timberlake Consultants.

Koopman, S., Shepherd, N., & Doornik, J. (1999). Statistical algorithms for models in state space using SsfPack 2.2. *Economet. J.*, **2**, 113–166.

Koopmans, L. (1974). *The Spectral Analysis of Time Series*. NY: Academic Press.

Kowalski, M. & Torresani, B. (2006). A family of random waveform models for audio coding. *Proc. ICASSP'06, Toulouse*, pp. III–57–60.

Kumaresan, R. & Tufts, D. W. (1980). Data-adaptive principal component signal processing. *Proc. of IEEE Conference On Decision and Control*, pp. 949–954.: Albuquerque.

Loéve, M. (1948). *Fonctions aléatories du second ordre.* Appendix to Levy, P., Stochastic Processes and Brownian Motion, ed. Gauthier-Villars, Paris.

Louis, A., Maass, P., & Rieder, A. (1997). *Wavelets - Theory and Applications.* Wiley.

Maass, P., Köhler, T., Costa, R., Parlitz, U., Kalden, J., Wichard, J., & Merkwirth, C. (2003). *Mathematical methods for forecasting bank transaction data.* Tech. report, Zentrum für Technomathematik. DFG SPP1114, Pre. 24.

Macauley, F. (1931). *The Smoothing of Time Series.* NY: Nat. Bureau of Econ. Research.

Mallat, S. (2001). *A wavelet tour of signal processing.* Academic Press, 2nd ed.

Maravall, A. & Caporello, G. (2004). *Working Paper 2004. Program TSW: Revised Reference Manual.* Tech. report, Research Dep., Bank of Spain. http://www.bde.es.

McElroy, T. (2006). Model-based formulas for growth rates and their standard errors. *2006 Proc. Am. Stat. Assoc.*: [CD-ROM]: Alexandria, VA.

McElroy, T. (2008). Matrix formulas for nonstationary signal extraction. *Econometric Theory*, **24**, 1–22.

McElroy, T. & Gagnon, R. (2006). *Finite Sample Revision Variances for ARIMA Model-Based Signal Extraction.* Tech. report, $RRS2006-05$. U.S. Census Bureau.

McElroy, T. & Sutcliffe, A. (2006). An iterated parametric approach to nonstationary signal extraction. *Comp. Stat. Data Anal.*, **50**, 2206–2231.

Meyer, Y. (1993). *Wavelets: Algorithms and Applications.* SIAM, Philadelphia.

Morley, J., Nelson, C., & Zivot, E. (2003). Why are beveridge-nelson and unobserved component decompositions of gdp so different? *Review of Economics and Statistics*, **85**, 235–243.

Musgrave, J. (1964). *A set of end weights to end all end weights.* Tech. report, US Census Bureau.

Nekrutkin, V. (1996). Theoretical properties of the "Caterpillar" method of time series analysis. *Proc. 8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, pp. 395–397.: IEEE Computer Society.

Partal, T. & Kuecuek, M. (2006). Long-term trend analysis using discrete wavelet compomnents of annual precipitation measurements in the marmara region. *Phys. Chem. of the Earth*, **31**(18), 1189–1200.

Parzen, E. (1959). *Statistical Inference on Time Series by Hilbert Space Methods.* Technical Report No. 53, Statistics Department, Stanford University, Stanford, CA.

Peña, D., Tiao, G., & Tsay, R. (2001). *A Course in Time Series Analysis.* New York: John Wiley & Sons.

Pearce, N. & Wand, M. (2006). Penalized splines and reproducing kernel methods. *Amer. Statistician*, **60**(3).

Pereira, W. C. A., Bridal, S. L., Coron, A., & Laugier, P. (2004). Singular spectrum analysis applied to backscattered ultrasound signals from in vitro human cancellous bone specimens. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, **51**(3), 302–312.

Pierce, D. (1975). On trend and autocorrelation. *Commun. Statist.*, **4**, 163–175.

Pierce, D. (1978). Seasonal adjustment when both deterministic and stochastic seasonality are present. *Seas. Anal. of Econ. Time Series*, Ed. A. Zellner, pp. 242–269.: US Dep. of Commerce.

Pierce, D. (1980). Data revisions with moving average seasonal adjustment procedures. *J. Econometrics*, **14**, 95–114.

Pollock, D. (2000). Trend estimation and de-trending via rational square-wave filters. *J. Econometrics*, **99**, 317–334.

Proietti, T. (2006). Trend-cycle decompositions with correlated components. *Economet. Rev.*, **25**, 61–84.

Rauch, H. (1963). Solution to the linear smoothing problem. *IEEE Transactions on Automatic Control*, **8**, 371–372.

Ravn, M. & Uhlig, H. (1997). *On adjusting the HP-Filter for the frequency of observations.* Tech. report, Tilburg University.

Salgado, D. R. & Alonso, F. J. (2006). Tool wear detection in turning operations using singular spectrum analysis. *J. of Mat. Proc. Tech.*, **171**(3), 451–458.

Schreiber, T. (1998). Processing of physiological data. *Nonlinear analysis of physiological data*, Eds. H. Kantz, J. Kurths, & G. Mayer-Kress, pp. 7–22. Springer.

Shiskin, J., Young, A., & Musgrave, J. (1967). The X11 variant of the Census method II seasonal adjustment program. *Technical Paper 15, US Department of Commerce, Bureau of the Census, Washington.*

Solow, A. R. & Patwardhan, A. (1996). Extracting a smooth trend from a time series: A modification of singular spectrum analysis. *J. Climate*, **9**, 2163–2166.

Stock, J. H. & Watson, M. W. (1988). Variable trends in economic time series. *J. Econ. Perspect.*, **2**(3), 147–174.

Sweldens, W. (1997). The lifting scheme: a construction of second generation wavelets. *SIAM J. Math. Anal.*, **29**(2), 511–546.

Thomson, P. & Ozaki, T. (2002). Transformation and trend-seasonal decomposition. *Proc. of the 3rd Int. Symp. on Frontiers of Time Series Modeling*: Tokyo.

Tona, R., Benqlilou, C., Espuna, A., & Puigjaner, L. (2005). Dynamic data reconciliation based on wavelet trend analysis. *Ind. Eng. Chem. Res.*, **44**(12), 4323–4335.

Triebel, H. (1992). *Theory of function spaces. II.* Birkhäuser.

Vautard, M. & Ghil, M. (1989). Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D*, **35**, 395–424.

Vautard, R. (1999). Patterns in time: SSA and MSSA. *Analysis of Climate Variability: Applications of Statistical Techniques*, pp. 265–286. Springer, 2nd ed.

Vautard, R., Yiou, P., & Ghil, M. (1992). Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. *Physica D*, **58**, 95–126.

Vedam, H. & Venkatasubramanian, V. (1997). A wavelet theory-based adaptive trend analysis system for process monitoring and diagnosis. *Proc. Amer. Cont. Conf.*, pp. 309–313.

Verdu, G. & Ginestar, D. (2001). Neutronic signal conditioning using a singular system analysis. *Ann. Nucl. Energy*, **28**(6), 565–583.

Wahba, G. (1990). *Spline Models for Observational Data*. Philadelphia: SIAM.

Wahba, G. (1999). *Advances in Kernel Methods: Support Vector Learning*, chapter Support Vector Machine, Reproducing Kernel Hilbert Spaces, and Randomized GACV, , pp. 69–88. MIT press.

Wecker, W. (1979). A new approach to seasonal adjustment. *Proc. Am. Stat. Assoc., Bus. Econ. Stat. Section*, pp. 322–323.

West, M. & Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. NY: Springer-Verlag.

Whittle, P. (1963). *Prediction and Regulation*. London: English Universities Press.

Wiener, N. (1949). *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series With Engineering Applications*. NY: Wiley.

Wildi, M. (2005). *Signal extraction*. Berlin: Springer-Verlag.

Wold, H. (1938). *A Study in the Analysis of Stationary Time Series*. Almquist & Wiksells.

Young, P. (1984). *Recursive Estimation and Time Series Analysis*. NY: Springer-Verlag.

Young, P. & Benner, S. (1991). *microCAPTAIN Handbook: Version 2.0*. Center for Research on Environmental Systems and Statistics, Lancaster University.

Young, P. & Pedregal, D. (1999). Recursive and en-block approaches to signal extraction. *J. Appl. Stat.*, **26**, 103–128.

Young, P., Pedregal, D., & Tych, W. (1999). Dynamic harmonic regression. *J. Forecasting*, **18**, 369–394.

Yue, S. & Pilon, P. (2004). A comparison of the power of the t-test, Mann-Kendall and bootstrap tests for trend detection. *Hydrol. Sci. J.*, **49**(1), 21–37.