Editorial

2nd Special Issue on Statistical Signal Extraction and Filtering

The two special issues of *Computational Statistics and Data Analysis* that have been devoted to the topic of *Statistical Signal Extraction and Filtering* represent the beginning of an ongoing enterprise. The accumulated papers seem set to become a major resource for scientists working in a wide range of disciplines. The contents of both issues are displayed in the bibliography of this editorial introduction; and the list serves to represent the scope and the extent of the enterprise.

The first special issue contained a total of 18 papers, including the introduction of Pollock (2006b), which gave an account of the history and of the recent developments in the areas of Wiener–Kolmogorov and Kalman filtering. The largest group of papers was in these two areas. They included the papers of Broto and Ruiz (2006), Pascual et al. (2006), Maravall (2006), McElroy and Sutcliffe (2006), Godolphin and Triantafyllopoulos (2006), Strickland et al. (2006), Mazzocchi (2006) and Pollock (2006a). A second group of papers was concerned with principal components analysis. They were by Ombao and Ho (2006), Coakley et al. (2006) and Aminghafari et al. (2006). The third major topic, which concerned wavelets analysis, was represented by the papers of Whitcher (2006), of Amato et al. (2006) and of Stoev et al. (2006). The remaining papers fell outside the clusters but well within the ambit of the special issue. Zeng and Garcia-Frias (2006) used hidden Markov models in the analysis of gene expression. Fried et al. (2006) considered a class of nonlinear filtering methods and Lehtinen (2006) applied the methods of signal extraction to simulated games.

In the second special issue, an emphasis has been given to multivariate methods, since it was perceived that there have been rapid developments in this direction. These developments have been stimulated both by advances in statistical computation and by the increasing availability of data in many areas of research, including Medicine, Engineering, Economics and Climatology, to name but a few. The second special issue features 15 articles. Many of these contributions are addressing problems of multivariate signal extraction; and they are doing so in a variety of ways. Some papers have adopted a parametric approach, whereas others are primarily nonparametric.

Bakker and Heskes (2007) present a dynamic hierarchical model for a large number of parallel time series. Each time series follows a dynamic linear model; and the individual series are linked together by a common or top-level unobserved component, which drives the evolution of the individual latent variables. In spite of the simple structure of the model, inference becomes unfeasible when the dimension of the cross section is large.

Shin et al. (2007) discuss the problem of constructing the optimal linear combination of local estimates of the states of a dynamic system arising in multisensor filtering problems. The optimal linear combination depends on the cross-covariances between the state estimates delivered by the Kalman filter. The paper presents explicit equations for their computation in the context of linear continuous-time filtering, where several parallel sensors provide information for reconstructing the state variable.

An area in which signal extraction has undergone considerable recent development is the processing of two-dimensional signals such as picture images. Klauzenberg and Lagona (2007) present an interesting application of hidden Markov random field models for estimating the age of a human being from the count of the number of tooth rings that appear in the tooth cement band. The observed image of a tooth slice is modelled by a hidden Markov random field, where pixel intensities are assumed to be independently and normally distributed, conditional on a Markov random field of hidden binary labels. The latter is based on the convolution of independent Bernoulli random variables with a parametric Gabor filter, so as to incorporate the main characteristics of the image, such as the long-range dependence among intensities and the quasi-periodicity in the placement of tooth rings.
Howlett et al. (2007) deal with the optimal design of a causal filter (i.e. one that uses current and past observations) with a finite impulse response. They propose a multilinear polynomial estimator and an accompanying computational procedure, which they illustrate using aerial photographic data.

The discrete wavelet transform maps a data sequence into a set of amplitude coefficients associated with a hierarchy of wave packets of which the frequency content and the temporal duration vary in an inverse relationship in the manner of Heisenberg. A wavelet analysis can coexist with the more traditional methods of multivariate analysis, such as cluster analysis, discriminant analysis and principal components analysis.

Maharaj and Alonso (2007) combine the techniques of wavelet analysis with those of discriminant analysis in order to determine the current status of a process that can switch between several alternative stationary regimes.

In the econometric analysis of a multivariate time series, the co-movements amongst economic time series are an important issue. Three papers deal with this topic. The paper of Cubadda (2007) extends the notion of a common feature in a multivariate time series, and it presents statistical tests based on canonical correlations that can be used to discriminate among different forms of common cyclical features.

Mastromarco and Woitek (2007) investigate the co-movements and the synchronisation of the regional business cycle in Italy using spectral methods, which attribute the overall variance of a series to sinusoidal motions within a range of frequencies.

Some of the papers in this issue describe linear methods of signal extraction. A linear signal-extraction method is one that represents features of the data by linear combinations of the available observations. The combinations may vary with time. Such methods are well understood and they are readily amenable to computation. They are a key ingredient for signal extraction in more elaborate nonlinear and non-Gaussian models. Several of the papers deploy such methods.

Proietti (2007) reviews the principal methods of inference that are applied to linear unobserved components models. He establishes the connection of these models with the class of linear mixed models, and he relates the methods to those of the semiparametric literature on spline smoothing and ad hoc filtering using penalised least squares. In particular, the equivalence is established between the empirical best linear unbiased predictor, fixed interval smoothing in a state space setting, and penalised least-squares estimation of signals.

The paper of Weinert (2007) focuses on efficient computation for the Whittaker–Henderson smoothing problem. It presents an efficient implementation of the discrete smoothing algorithms, including the computation of the diagonal of the smoothing matrix, which is needed for cross-validation and for the estimation of the effective dimension. His comparison of computational speed shows that the Kalman filter and smoother can outperform most other methods.

Maravall and del Rio (2007) study the aggregation properties of the Hodrick–Prescott filter. This filter, which decomposes a time series into a long-term trend component and a stationary cycle, is a special case of the Whittaker–Henderson smoothing filter; and it has been widely employed in applied macroeconomics. The paper distinguishes between systematic sampling and temporal aggregation (stock and flow variables), and it provides a definitive treatment of the selection of the smoothing parameter for data available at different frequencies of observation.

The paper by Bujosa et al. (2007) deals with a class of unobserved components models where the components of a time series have a so-called dynamic harmonic representation. This is a linear combination of sines and cosines with time-varying coefficients. The coefficients that modulate the trigonometric functions are assumed to follow a generalised random walk. The paper deals with the ARIMA reduced-form representation of the components and it derives an algorithm that identifies the spectral peaks in the empirical spectrum. It estimates the structural parameters of the corresponding components via a frequency-domain method that fits a parametric spectrum to the periodogram of the data.

When a process ceases to be stationary throughout the length of the sampled data, it may be locally stationary or it may be amenable to modelling with time-varying parameters. The paper of Maharaj and Alonso (2007), already mentioned above, concerns locally stationary processes. That of Triantafyllopoulos and Nason (2007) introduces a new class of moving-average processes with time-varying parameters. It considers a multiplicative model that describes the evolution of the squares of the MA parameters.
Long-range dependence is a concept that has been used extensively by econometricians in modelling the volatility of financial returns. Ledesma et al. (2007) invoke the concept for modelling packet traffic in communication networks. They propose two new methods to approximate the power spectrum of a fractional Gaussian-noise process: the first method employs a second-order approximation and the second method is based on splines. They show that the two methods are faster and have a better accuracy than existing methods. When it is placed beside the econometric applications, this paper serves to illustrate the manner in which the techniques of signal extraction are applicable in disparate subject areas. It is this kind of fruitful interdisciplinarity that has attracted the editors to the project of compiling this special issue.

Another topic in signal extraction, which is of increasing importance, is the detection of structural breaks. The topic is touched on in a number of papers, but it is the primary focus of the final paper by Fried, which deals with the robust detection of level shifts in a time series (Fried, 2007).

**References**


D.S.G. Pollock

*Queen Mary College, University of London, Mile End Road, London E1 4NS, UK*

*E-mail addresses: pollock@sigmapi.u-net.com*

Tommaso Proietti

*SEFeMEQ, University of Rome ‘Tor Vergata’, Via Columbia 2, 00133 Rome, Italy*

*E-mail addresses: tomaso.proietti@uniroma2.it*