Identification of System Behaviours by Approximation of Time Series Data

Wolfgang Scherrer[a]   Christiaan Heij[b*]

version 19-03-97

a Institut für Ökonometrie, Operations Research und Systemtheorie, Technische Universität Wien, Austria

b Econometric Institute, Erasmus University Rotterdam, Netherlands

Abstract

The behavioural framework has several attractions to offer for the identification of multivariable systems. Some of the variables may be left unexplained without the need for a distinction between inputs and outputs; criteria for model quality are independent of the chosen parametrization; and behaviours allow for a global (i.e., non-local) approximation of the system dynamics. This is illustrated with a behavioural least squares method with an application in dynamic factor analysis.

Keywords: behaviour, linear system, system identification, least squares, factor models.

*Corresponding author, email: heij@few.eur.nl

1 Introduction

Principally, behaviours focus our attention on the primary level of system dynamics, that is, on empirical observations. In a sense the behaviour consists of the objective information on the system. Other issues, like input-output decomposition and parametric representation, are secondary as they depend on the subjective choices of the user. This explicit distinction between data information, model representation, and model use is one of the charms of the behavioural approach in system theory.

Several behavioural methods for system identification have been proposed since the introduction of this framework by Willems in [18]. One approach is based on state models, realization theory and model reduction as in [19], see for example [3, 4, 16]. Subspace methods as in [14] can be seen as approximate versions of the realization method in [19]. Earlier roots of these ideas are in [1, 2] for stochastic systems. Another approach is more equation oriented as in [20], see for example [6, 17], and this related to prediction error methods [9]. A third approach, based on least squares approximation of behaviours, is developed in [11], see also [8, 13].

In this paper we consider the behavioural least squares method (BLS), also called global total least squares. In a sense, this method adheres most strictly to the behavioural principle of focusing on the external system characteristics. Subspace methods and realization based methods for data modelling employ states as secondary objects in their approximation, and equation oriented procedures are faced with the choice of canonical parameters. In BLS, the model quality is expressed simply as the least squares distance between the empirical data and the system behaviour.

In Section 2 we briefly review the BLS method. This is applied in Section 3 to dynamic factor models for stationary processes, and topics for future research are outlined in Section 4.

2 Behavioural Least Squares

We assume that the reader is familiar with the behavioural framework as developed in [18, 19, 20, 21]. The empirical data are denoted by w, a q-dimensional vector time series observed in discrete time over the interval $T \subset Z$. A behaviour is a subset $\mathcal{B} \subset (R^q)^Z$ of the set of all time series over the full time axis Z. We will only consider behaviours that correspond to linear systems, that is, $\mathcal{B}$ is a linear, shift invariant set that is closed in the topology of pointwise convergence.

In general, unless the observed data are very structured, linear systems that model the observed data without error will be very complex. If $T = Z$, generically the only linear system with $w \in \mathcal{B}$ is given by $\mathcal{B} = (R^q)^Z$. If T is a finite interval of length N, then it can be shown that there always exists a linear system $\mathcal{B}$ with $p = 1$ output and $m = q - 1$ inputs and state dimension $n \leq (N + 1)/(q + 1)$ with the property that $w \in \mathcal{B}_T$, the restriction of $\mathcal{B}$ to T.

Such systems are not helpful in describing the data in a less complex way, and this is only possible by allowing for some kind of approximation. We measure the approximation error by the squared distance between the data and the linear system, that is

$$d_T(w, \mathcal{B}) = \min\{\| w - \hat{w} \|; \ \hat{w} \in \mathcal{B}_T\} \tag{1}$$

where $\| w - \hat{w} \|^2 = \sum_{t \in T} \sum_{j=1}^{q} \{w_j(t) - \hat{w}_j(t)\}^2$. Because of the foregoing result, we consider only approximating systems with $p > 0$ and $n < N$, in which case the dimension is given by

$$\dim(\mathcal{B}_T) = mN + n \tag{2}$$

The complexity of a system is defined by the pair (m, n), that is, by the number of inputs and the number of states of the system. The behavioural least squares problem is to determine the optimal system for given complexity, that is,

$$d_T(w, m, n) = \min\{d_T(w, \mathcal{B}); \quad \mathcal{B} \text{ has complexity } (m, n)\} \tag{3}$$

For a structural analysis of this minimization problem we refer to [7] and for a Gauss-Newton algorithm to [11, 12]. In practice one can determine the errors $d_T(w, m, n)$ for a range of complixities (m, n) and choose the model that provides an acceptable trade-off between complexity and fit. This choice is subjective, no formal criteria have been developed until now.

As compared with other methods for system identification, the above BLS approach has the following characteristic features. The model quality, in terms of fit and complexity, is defined on the observational level, in terms of the data w and the behaviour $\mathcal{B}$. Further, all variables are treated in a symmetric way, and the error criterion measures the global misfit of models in the sense that $\hat{w}$ is required to satisfy all the system restrictions. In comparison, more conventional identification procedures like prediction error methods start from an input-output decomposition of the system variables and consider the one-step-ahead forecasting quality, a local criterion of fit. The equation oriented methods in [6, 20] are symmetric, but with a local criterion. The realization based behavioural methods in [3, 4, 16] first determine an exactly fitting system, that is, with $d_T(w, \mathcal{B}) = 0$, and then employ model reduction procedures to lower the complexity. The main distinction with BLS is that the realization step neglects the energy content of the data. That is, in this step all directions of the signal space that are present in the data are incorporated in the system, whereas in (1) we may neglect low energy directions without much loss in fit. Stated otherwise, BLS is more robust with respect to data variations.

3 Factor Models for Stationary Processes

Suppose that the data are generated by a full rank stationary process w with spectral density S that is bounded on the unit circle. A factor model is a representation

$$w = \hat{w} + \tilde{w} \tag{4}$$

characterized by the condition that the factor process $\hat{w}$ has less degrees of freedom than w. The process $\tilde{w}$ is the error resulting from the approximation of w by the reduced process $\hat{w}$. We denote by $\hat{S}$ the spectrum of $\hat{w}$ and by $\mathcal{B}(\hat{w})$ the factor behaviour, that is, the smallest linear system with the property that $\hat{w} \in \mathcal{B}$ almost surely.

In the non-dynamic case, S is the covariance matrix of w and factor models have the property that the covariance matrix $\hat{S}$ has reduced rank. If we fix the allowed rank m of this matrix, i.e., the dimension of $\mathcal{B}$, then the model that minimizes the error $E \parallel \tilde{w}(t) \parallel^2$ is given by principal components. If the covariance matrix has eigenvalue decomposition $S = \sum_{j=1}^{q} \lambda_j u_j u_j^*$, with $\lambda_1 \geq \cdots \geq \lambda_m > \lambda_{m+1} \geq \cdots \geq \lambda_q > 0$, then the principal component model of complexity m is given by

$$\hat{w} = \sum_{j=1}^{m} u_j u_j^* w, \quad \tilde{w} = \sum_{j=m+1}^{q} u_j u_j^* w \tag{5}$$

with behaviour $\mathcal{B}(\hat{w}) = \mathrm{span}\{u_j; j = 1, \cdots, m\}$. In the dynamic case we can apply this decomposition frequency-wise, as is proposed in [5]. Under certain regularity conditions, the functions u_j can be chosen to be analytic so that the factor process $\hat{w}_{PC} = \sum_{j=1}^{m} u_j(\sigma) u_j(\sigma)^* w$ and the error process $\tilde{w}_{PC} = \sum_{j=m+1}^{q} u_j(\sigma) u_j(\sigma)^* w$ are well-defined, with σ the shift operator on Z. The resulting error $E\|\tilde{w}_{PC}(t)\|^2 = \sum_{j=m+1}^{q} \int_{-\pi}^{\pi} \lambda_j(e^{-i\mu}) d\mu$ is minimal among all factor models with $\hat{w}$ of rank at most m. The disadvantage of this model is that the restrictions on the factor process $\hat{w}$ are in general non-rational. This means that it is hard to give an explicit description of the factors. More precisely, in general the factor behaviour $\mathcal{B}(\hat{w}_{PC}) = (R^q)^Z$ so that, in the sense of linear systems, the factor process $\hat{w}_{PC}$ is not simpler than the original process w.

Suppose that the allowed complexity (m, n) of the factor behaviour has been fixed. There are now several ways to find a model that satisfies this restriction, according to different ways of approximation. An obvious method similar to principal components is to find a solution for

$$\min\{d(w, \mathcal{B}); \quad \mathcal{B} \text{ has complexity } (m, n)\} \tag{6}$$

where

$$d(w, \mathcal{B}) := \min\{(E \parallel w(t) - \hat{w}(t) \parallel^2)^{1/2}; \quad \mathcal{B}(\hat{w}) = \mathcal{B}\} \tag{7}$$

This can be seen as the 'infinite sample' analogon of BLS in (1) and (3), see [7] for further details.

An alternative approach, somewhat in the spirit of the realization approach, is to construct the principal component model in a first step and then to approximate this by a system of given complexity. The factor process of the principal component model (5) satisfies the restrictions $U(\sigma)\hat{w}_{PC} = 0$, where U is the (in general non-rational and non-controllable) $(q - m) \times q$ matrix function with rows $u_j^*, j = m + 1, \cdots, q$. The idea is to approximate U by a rational function $\hat{U}$, based on the frequency data $U(e^{-i\mu_j})$, $\mu_j = 2\pi j/n_f$, $j = 0, \ldots, n_f - 1$. For

4

this purpose we extend frequency domain algorithms in [10] for causal transfer functions and in [15] for miniphase spectral factors to the non-causal case. To describe this in more detail, let $G(e^{-i\mu}) = C(e^{i\mu}I - A)^{-1}B + D$ be a rational transfer function with n_1 stable and n_2 unstable poles so that

$$G(e^{-i\mu}) = C_1(e^{i\mu}I_{n_1} - A_1)^{-1}B_1 + C_2(e^{i\mu}I_{n_2} - A_2)^{-1}B_2 + D \qquad (8)$$

with A_1 an $n_1 \times n_1$ stable matrix and A_2 an $n_2 \times n_2$ antistable matrix. Now G can be reconstructed from the frequency data $G(e^{-i\mu_j})$, $\mu_j = 2\pi j/n_f$, $j = 0, \ldots, n_f - 1$, by the following subspace algorithm (assuming that $n_f > 2(n_1+n_2)$ and, for notational simplicity, that n_f is even). The inverse discrete Fourier transform of $G(e^{-i\mu_j})$ is given by

$$\bar{G}_0 = D + M_0, \quad \bar{G}_j = C_1 A_1^{j-1} M_1 + C_2 A_2^{-n_f/2+1} A_2^{j-1} M_2, \; j = 1, \ldots, n_f - 1 \quad (9)$$

with M_i matrices that depend on A, B, C and n_f. The $(n_f/2 \times n_f/2)$ block Hankel matrix H built up from $\bar{G}_j$, $j = 1, \ldots, n_f - 1$, has rank $n_1 + n_2$ and the parameters in (8) can be obtained from H by the following subspace method. A singular value decomposition of H is used to determine matrices K and L, both with $n_1 + n_2$ columns, such that $H = KL^T$. Then A is obtained from the regression of $K_{2:(n_f/2)}$ onto $K_{1:(n_f/2-1)}$, where for example $K_{1:(n_f/2-1)}$ denotes the matrix consisting of the first $(n_f/2 - 1)$ block rows of K. After transforming A to block diagonal form $A = \text{diag}(A_1, A_2)$ and transforming and partitioning the columns of $K = [K_1, K_2]$ correspondingly, C_1 is the first block row of K_1 and C_2 the $(n_f/2)$-th block row of K_2. Then B and D are obtained by regressing $G(e^{-i\mu_j})$ on $C(e^{i\mu_j}I - A)^{-1}B + D$, with A and C given. This realization procedure can also be applied for the identification of reduced order systems, by using a singular value approximation of H. We call this the subspace identification method for frequency data (SIFD).

We apply SIFD in two methods to approximate the principal component model $U(\sigma)\hat{w}_{PC} = 0$. The first method, that we will call the input-output method (IOM), is defined as follows. Let the observed variables be decomposed in m inputs and $p = q - m$ outputs, with a corresponding decomposition $U = [U_1, -U_2]$ of the columns of U, then the transfer function is given by $G = U_2^{-1}U_1$. This transfer function is in general non-rational and non-causal. For given frequency data $G(e^{-i\mu_j})$, $j = 0, \cdots, n_f - 1$, a rational approximation $\hat{G} = P^{-1}Q$ of McMillan degree n is obtained by SIFD. The corresponding behaviour $\mathcal{B}_{IOM}$ is described by the polynomial relations $\hat{U}(\sigma)\hat{w} = 0$ where $\hat{U} = [Q, -P]$.

A disadvantage of IOM is the arbitrary selection of inputs and outputs. The second method, that we will call the iterative relation method (IRM), does not require this selection. In IRM the relation U is approximated, taking into account that U is only defined up to left multiplication with a $p \times p$ unitary matrix function V. The system $\mathcal{B}_{IOM}$ can be represented as $\mathcal{B}_{IOM} = \ker(U_n)$ with U_n a $q \times m$ isometric rational matrix function of McMillan degree n, see [11]. Now for each frequency μ_j an orthogonal $p \times p$ matrix V_j is determined such that $\|U_n(e^{-i\mu_j}) - V_j U(e^{-i\mu_j})\|$ is minimal. Then a rational approximation of the frequency data $V_j U(e^{-i\mu_j})$ is obtained by SIFD with a corresponding isometric

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} -0.1965 & 0.0000 & -0.3143 & 0.8770 & -0.4949 \\ 0.5200 & -0.2538 & -0.6434 & -0.6621 & -0.0280 \\ 0.0952 & 0.7954 & 0.2210 & -0.2319 & -0.6759 \\ -0.4646 & -0.2678 & -0.1856 & 1.0000 & 0.0000 \\ 0.1833 & 0.3058 & 0.4734 & 0.0000 & 1.0000 \end{bmatrix}$$

n		$N = \infty$	Mean	StDev	Median	Min	Max
1	BLS	0.8747	0.8771	0.0020	0.8765	0.8747	0.8871
	IOM	1.0685	1.0607	0.0574	1.0712	0.8847	1.2341
	IRM	0.8754	0.9666	0.0985	0.8897	0.8747	1.1732
2	BLS	0.8495	0.8693	0.0052	0.8683	0.8504	0.8899
	IOM	0.8761	0.9356	0.0812	0.8978	0.8508	1.2128
	IRM	0.8703	0.8784	0.0367	0.8714	0.8513	1.0854

Table 1: Simulation of a process with $q = 2$ variables and state dimension $n = 3$. The matrix shows the parameters of the data generating process $\sigma x = Ax + B\varepsilon$, $w = Cx + D\varepsilon$, with ε a two-dimensional white noise process with unit covariance matrix. The error of the principal component model with $m = 1$ is 0.8209. The column with $N = \infty$ relates to the case where the true spectral density is used in IOM and IRM and where for BLS (7) is minimized. The next columns show summary statistics on the errors $d(w, \mathcal{B})$ in (7) of factor models of complexity $m = 1$ and $n = 1, 2$ obtained in 500 simulation runs with samples of length $N = 1024$. For IOM and IRM $n_f = 16$ frequencies were used, the results are similar for larger values.

rational function $\hat{U}_n$. The foregoing steps are iterated with this new approximation $\hat{U}_n$, until convergence is reached. If $\hat{U}$ is the final approximation, then the identified factor behaviour is described by $\hat{U}(\sigma)\hat{w} = 0$.

We compare these three approaches (BLS, IOM, IRM) by means of a simulation. We consider two situations, one where the process spectrum S is known and another where the empirical data consists of a finite sample of length N of the process, in which case the spectrum is estimated by smoothing the periodogram. In table 1 we summarize results of simulations obtained from a randomly chosen data generating process with $q = 2$ variables and $n = 3$ states, that is, S is a 2×2 rational matrix with rank 2 and McMillan degree 6. For this process, behaviours are estimated with complexities $m = 1$ and $n = 1, 2$. Reported are results on the corresponding errors (7).

The results indicate that BLS can give reasonably good approximations of stationary processes by factor processes and that the methods IOM and IRM perform less well. This is partly because the approximation step of these algorithms, i.e., the rational approximation of U, is not directly related to the criterion $d(w, \mathcal{B})$. Alternative input-output schemes might give better results, but in any case the choice of inputs and outputs remains arbitrary in this setting.

Concluding, for the estimation of dynamic factor models BLS seems to outperform IOM and IRM. A more complete analysis and development of alterna-
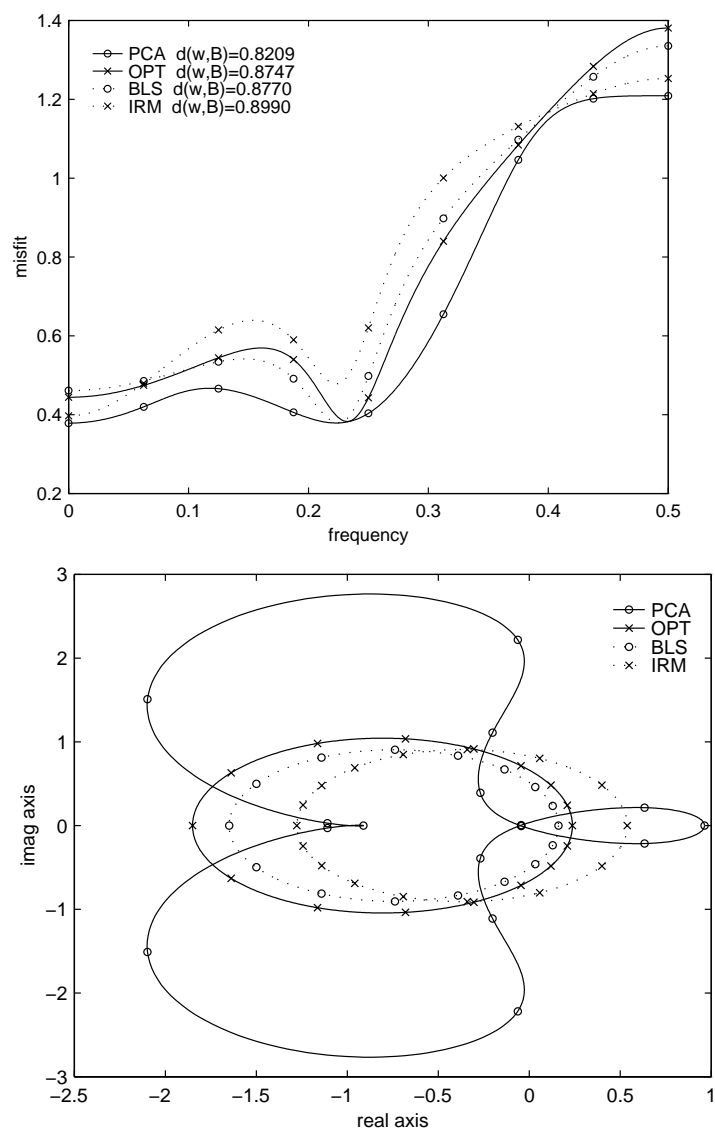
Figure 1: Four factor models for the process described in table 1: PCA (the principal component model), OPT (the optimal model with $m = 1$, $n = 1$), BLS and IRM. The models are determined for a random sample of length $N = 1024$, using $n_f = 16$ frequencies. The first plot shows the frequency-wise squared approximation error (the trace of the spectrum $\tilde{S}$ of the error process $\tilde{w}$, which integrated over $[-\pi, \pi]$ gives the error (7)). The PCA model gives a lower bound for the error of all models with $m = 1$. The second plot is the Nyquist plot of the corresponding transfer functions $G = P^{-1}Q$, where $[Q, -P]\hat{w} = 0$ describes the corresponding factor behaviours. In the PCA model G is non-rational, whereas for the other three models G has McMillan degree $n = 1$.
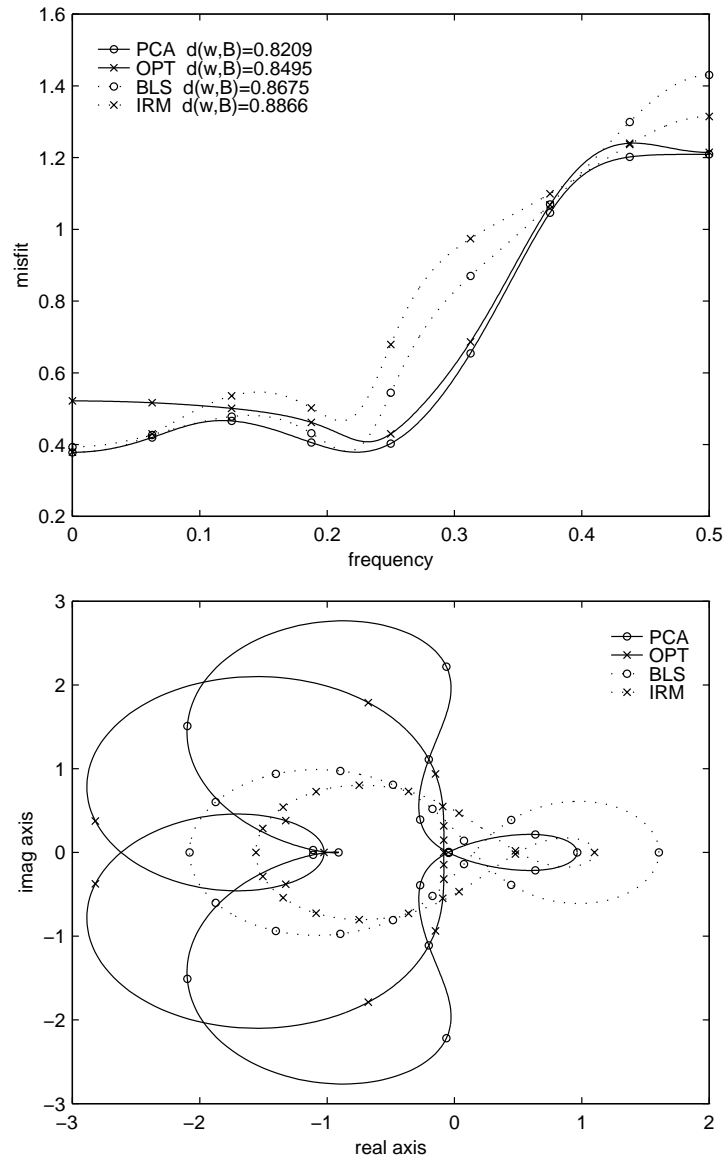
Figure 2: Results for the model in table 1 for the case $m = 1$, $n = 2$, see figure 1 for more details.

8

tive methods are topics for further investigation.

4   Conclusion

In this paper we discussed the behavioural approach for the identification of linear systems. The behavioural least squares (BLS) method expresses model quality on the observational level, without the need to choose inputs and outputs or a parametric representation of the model. The criterion function evaluates the global fit of the model, that is, it not only considers the local restrictions but also all behavioural restrictions over time intervals of arbitrary length.

We applied BLS for the identification of dynamic factor models, with the advantages that all variables are treated in a similar way (no inputs and outputs) and that the obtained model for the factors is a linear system (as compared to the infinite dimensional system obtained by principal components). The results are merely indicative of the possible uses of behavioural identification. A more thorough comparison with existing methods is needed before more general conclusions can be drawn. With respect to the application considered in this paper, this concerns the development of methods to reduce a given spectrum to one of given rank and McMillan degree.

References

[1] H. Akaike, Stochastic theory of minimal realization, *IEEE Transactions on Automatic Control* 19, 1974, pp. 667-674.

[2] H. Akaike, Canonical correlation analysis of time series and the use of an information criterion, in R. Mehra and D. Lainiotis (eds.), *System Identification: Advances and Case Studies*, Academic Press, New York, 1976, pp. 27-96.

[3] A.C. Antoulas, A new approach to modeling for control, *Linear Algebra and its Applications* 203-204, 1994, pp. 45-65.

[4] A.C. Antoulas and J.C. Willems, A behavioural approach to linear exact modeling, *IEEE Transactions on Automatic Control* 38, 1993, pp. 1776-1802.

[5] D.R. Brillinger, *Time Series: Data Analysis and Theory*, Holt, Rinehart and Winston, New York, 1975.

[6] C. Heij, *Deterministic Identification of Dynamical Systems*, Lecture notes in control and information sciences 127, Springer, Berlin, 1989.

[7] C. Heij and W. Scherrer, Consistency of system identification by global total least squares, *Report* 9635, Econometric Institute, Erasmus University Rotterdam, 1996. Submitted for publication.

[8] C. Heij, W. Scherrer and M. Deistler, System identification by dynamic factor models, *SIAM Journal on Control and Optimization* 35, 1997, to appear.

[9] L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, Englewood Cliffs, New Jersey, 1987.

[10] T. McKelvey, *Identification of State Space Models from Time and Frequency Data*, Linköping Studies in Science and Technology, Dissertation no. 380, Linköping, 1995.

[11] B. Roorda, *Global Total Least Squares*, Tinbergen Institute Research Series 88, Thesis Publishers, Amsterdam, Netherlands, 1995.

[12] B. Roorda, Algorithms for global total least squares modelling of finite multivariable time series, *Automatica* 31, 1995, pp. 391-404.

[13] B. Roorda and C. Heij, Global total least squares modelling of multivariable time series, *IEEE Transactions on Automatic Control* 40, 1995, pp. 50-63.

[14] P. Van Overschee and B. De Moor, Subspace algorithms for the stochastic identification problem, *Automatica* 29, 1993, pp. 649-660.

[15] P. Van Overschee, B. de Moor, W. Dehandschouter and J. Swevers, A subspace algorithm for the identification of discrete time frequency domain power spectra, *Report* ESAT-SISTA/TR, Leuven, 1996.

[16] S. Weiland and A.A. Stoorvogel, Optimal Hankel norm identification of dynamical systems, *Automatica* 33, 1997, to appear.

[17] E. Weyer, *System Identification in the Behavioural Framework*, Thesis, Department of Engineering Cybernetics, The Norwegian Institute of Technology, Trondheim, Norway, 1992.

[18] J.C. Willems, From time series to linear system, part I: Finite dimensional linear time invariant systems, *Automatica* 22, 1986, pp. 561-580.

[19] J.C. Willems, From time series to linear system, part II: Exact modelling, *Automatica* 22, 1986, pp. 675-694.

[20] J.C. Willems, From time series to linear system, part III: Approximate modelling, *Automatica* 23, 1987, pp. 87-115.

[21] J.C. Willems, Paradigms and puzzles in the theory of dynamical systems, *IEEE Transactions on Automatic Control* 36, 1991, pp. 259-294.