88-44C

De Moor B. *An introduction to system identification.* 1988 Integrated European Course in Mechatronics, Leuven, Belgium, May 30- June 3, 1988.

An Introduction to System Identification

Bart De Moor *
ESAT (Electronics, Systems, Automation, Technology)
Katholieke Universiteit Leuven
Kardinaal Mercierlaan 94
3030 Heverlee
tel. 32/16/220931 telex 25941 elekul fax 32/16/221855
email: UUCP:...!kulesat!demoor Bitnet:demoor@blekul60

June 2, 1988

1 Introduction

Inferring models from observations and measurements, and studying their properties, is one of the central issues in all scientific disciplines. *System identification* deals with the problem of building mathematical models of dynamical systems based on observed data from these systems. A system is an object in which variables of different kinds interact. These variables can be divided in *inputs, outputs, disturbances and states.* Typically, *inputs* are those variables that can be manipulated by the modeller and affect the system as external stimuli. *Outputs* are the direct observations. The *disturbances* can be divided into those that are directly measurable and those that are only observed through their influence on the output. Disturbances include measurement noise, uncontrollable inputs, etc.... The *state* of a system is the minimal information that is needed to determine the output, once the inputs and disturbances are known. This paper is organised as follows: In section 2, we discuss in general terms a possible classification of mathematical models and the models that are very well suited for identification. A general discussion on the system identification methodology may be found in section 3, where some *practical guidelines* are given concerning setting up an identification experiment. Input-output models are discussed in section 4. Their identification is analysed in some detail in section 5. In section 6, the most important properties of state space models are summarized and the identification of state space models is surveyed in section 7. Finally, in section 8, several succesful applications of the reviewed identification schemes on industrial plants are presented. The conclusions and references to some interesting books can be found in section 9.

2 The Models.

Mathematical models may be phrased with varying degrees of mathematical formalism. A rough classification can be obtained from the following list of qualifications: *time-invariant /*

*supported by the Belgian I.W.O.N.L

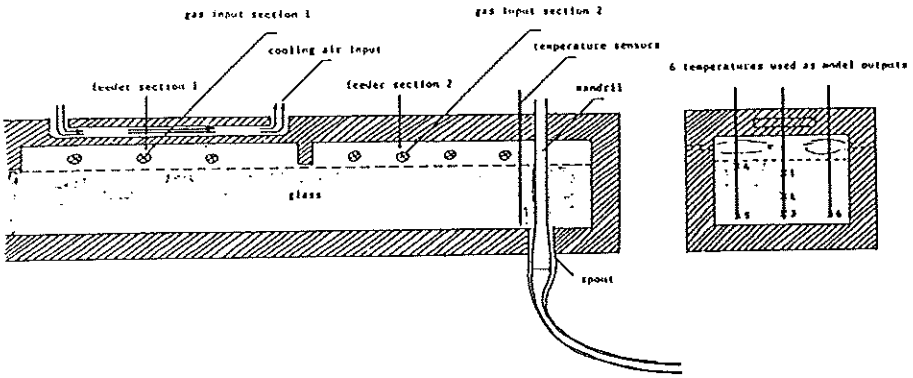

Figure 1: Cross section of a feeder

time-varying, discrete/continuous, linear/non-linear, lumped/distributed.

In this survey, we shall employ only

lumped linear discrete time time-invariant system.

The preference for this model class is dictated by several reasons:

- The *lumpedness* arises from the fact that in most cases, the *sensors* only collect *local* measurements that represent the situation of the system only in the immediate neighbourhood of the sensor. As an example, consider the temperature measurement in a glass feeder by using several sensors. What one really models is the transfer of the inputs (e.g. heating-ventilation) to the outputs, which are the temperatures at the very location of each sensor. Observe that the real system is certainly a *distributed parameter system*, the mathematical model of which is in terms of partial differential equations, when derived from physical laws. Yet, the several sensors represent a spatial *discretization* of the system by a 3-dimensional grid of temperature measurements. Hence the corresponding mathematical model becomes *lumped.*

- While for most physical systems it is most natural to work with a continuous time representation (e.g. differential equations), the increasing use of *digital* measurement and computational equipment, forces the uses of *discrete time models.* Mathematically, it is possible to convert any continuous time behavior into discrete time under fairly general conditions while the reverse is not necessarily true so that discrete time models are slightly more general than continuous time ones. As an example, consider the first order difference equation:

  $$x_{k+1} = -0.5x_k + u_k$$

  There does not exist a first order differential equation that after sampling gives this difference equation. If however the discrete time system matrix A (see section 6) has no eigenvalues on the negative real axis, there exists always a corresponding continuous

time system, which by appropriate sampling, gives the discrete time model. For more details, the interested reader is referred to [1, p.36].

- For *non-linear* systems, we shall exploit the idea of *local linearization.* The behavior is considered to originate in a linear system, within the observation window of finite length. Obviously, the quality and reliability of the derived model will depend upon the relation between the length of the finite window, the number of observations (the sampling rate) and the time constants that characterize the time-variance within this finite window. Observe that the restriction to *linearity* is a self-imposed limitation to the kind of mathematical operations and devices that will be used. Not only is the theory of linear models well developed, the algebraic and numerical tools that are needed are abundantly avalaible and frequently reduce to the solution of a set of (overdetermined) linear equations, or to (generalized) eigenvalue problems.

- For *time varying* systems, we adopt the point of view of *quasi-stationarity*: Since in most practical situations, the observation-window is of finite length and the number of observations is finite as well, the behavior of the system can be *approximated sufficiently well* within this finite window by a time invariant system. An *adaptive strategy* will be developed, which aims at updating the model from one time window to another with minimization of the necessary additional computational complexity. The time variance is then revealed within the observed changes of one model with respect to a previous obtained one.

Mathematical models are derived and applied for several reasons:

Simulation: One may wish to analyse the behavior of a system via simulations under those circumstances where experiments on the *real* system are too dangerous (nuclear power plants, ...), too expensive (loss of production, ...), too time consuming (very slow processes,...), too complicated or simply impossible (ecological systems, ...)

Prediction: In some situations, one is interested in predicting over a certain time horizon, the future behavior of a system, possibly under several different scenarios on the input and disturbance variables. Other applications include so-called *predictive control.*

Optimal Filtering: An appropriate mathematical model may be useful in obtaining information concerning variables that are not directly accessible or observable. This includes estimation of state variables via Luenberger or Kalmam filters.

Control Applications: Once a mathematical model (e.g. a state space model) of a system is available, one can develop controllers for such systems that achieve a certain task in an *optimal* way: one may for instance require that a certain action is to be performed in minimal time or with a minimal consumption of energy. Such problems can be translated in a mathematical framework and then solved by exploiting certain theories that are explicitly based upon the available mathematical models. Besides the optimality of certain control actions, automatic control may also be applied from the point of view of *robustness.* In these applications, one is interested in *decoupling* as much as possible the perturbing action of undesired disturbances from the output or the states. A mathematical model is unavoidable in order to perform the necessary computations. The most simple approach to the design of an automatic control system

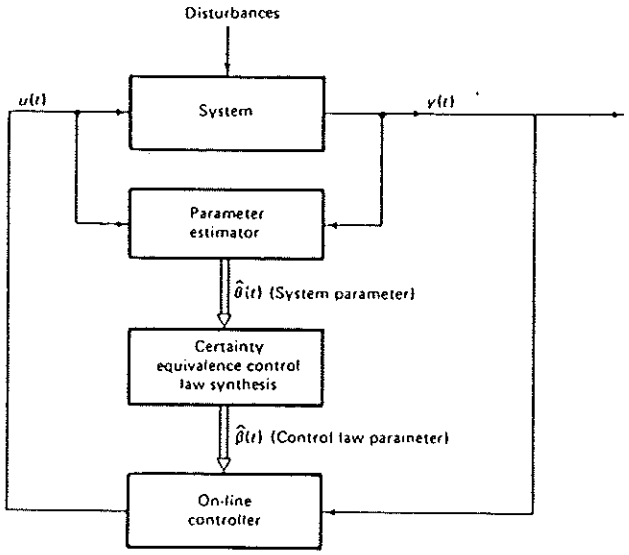

Figure 2: Certainty Equivalence Control Law

is to combine a certain parameter estimation or system identification approach with any control law. This approach of using the estimated model as if it were the true system for the purpose of design, is called the *certainty equivalence principle.* With this approach, one can conceive of generating a wide spectrum of algorithms, depending on which parameter estimation scheme is chosen and which control law is used. Perhaps the best known certainty equivalence stochastic adaptive control law is the *self-tuning regulator* of Astrom and Wittenmark [1]. It combines a least squares procedure for parameter estimation with a one-step ahead minimum variance certainty equivalence controller. The theoretical analysis of such certainty equivalence controllers though is very complicated, due to the strong non-linear action of the parameter estimation scheme.
Finally, control may also be achieved in terms of a so-called *reference model* [1, p.352]. The desired input-output behavior is here modelled via a prespecified mathematical model and the controller tries to adapt the system in order to follow this desired behavior. System identification methods may be applied in order to find an appropriate model for the reference model.

3 The System Identification Methodology

The construction of a model from data involves three basic steps:

1. The data.
2. A set of models.
3. A criterion.

3.1 The Data

It is a *conditio sine qua non*, to have an abundant amount of observations in order to be able to identify a mathematical model from data. However, the acquisition of 'good' data, is not at all a trivial task. The following issues should be kept in mind:

Determination of inputs and outputs: One of the first questions to be solved is the determination of the variables that will be measured in order to obtain a mathematical model of the system under study. The appropriate choice may be determined by the ultimate goal of the model. Another important problem is the determination of inputs and outputs and the question of *causal dependency*. The question here concerns the problem of which variable influences or causes another one.

Chosing the appropriate measurement equipment: This includes the choice of appropriate sensors, the specification of the required and sufficient accuracy, time comstants, the reproducability of the sensor, its drift and offset, the analysis of possible 50 Hz interference etc...

The sampling time and the data storage: The procedure of sampling the data that are produced by the system is inherent in computer based data acquisition systems. It is unavoidable that sampling as such leads to information losses and it is important to select the sampling instances so, that these losses are insignificant. Typically (and most effectively indeed), sampling is carried out at equidistant sampling instants. One can prove mathematically that, if one is sampling at a sampling frequency f_s, no information is 'lost' about frequency components in the signals, which are lower than the so-called Nyquist frequency, which is $f_s/2$. Hence, in order to avoid distortion by the so-called *folding effect*, one should employ *anti-aliasing presampling filters* to the analog signals, in order to eliminate from the signals all frequency components higher than the Nyquist frequency $f_s/2$. Reversely, the sampling rate should be chosen *in principle* twice the highest frequency of interest. However, in practical cases, one advocates a sampling rate which is 4 to 10 times higher than the minimal frequency of interest. Observe that anti-aliasing filtering at the same time induces a considerable noise reduction, since typically, the energy of the noise has important high frequency contents. A detailed analysis and practical guidelines for appropriate determination of the sampling rate, may be found in [1, p.29, p.71] [10, p.385-386].

The data preprocessing: In a lot of applications, especially in industrial environments, it is absolutely necessary to 'clean' the data before any identification approach can be applied. This preprocessing includes the elimination of occasional bursts and outliers, 'peak shaving', trend removal, estimation of drift and offset, periodical interference, the analysis of disturbances, such as day-night phenomena etc... Some useful guidelines and algorithms may be found in [2].

The Estimation of time delays: Time delays are common in mathematical models of industrial processes. As an example, consider the measurement of the tube wall thickness of a glass tube. This can only be measured with suffucient accuracy if the tube itself is sufficiently cooled. This introduces a considerable time delay. While a continuous time system with a time delay is in essence infinite dimensional, the corresponding sampled discrete time system is finite dimensional and can be treated by the introduction of controllable-observable poles in the origin. Details can be found in [1, p.42]. Delays can be estimated via a physical investigation of the origin of the delays, via cross-correlation techniques or from inspection of the impulse responses.

Experimental Design: An important issue which is an explicit part of any identification scheme, is the a priori determination, if possible of the input sequences that will be

applied. In some cases, experimentation on the real plant is impossible and data should be obtained from *normal operating records.* In other cases, the modeller is free to chose his own input sequences. In any case, for a reliable identification, the inputs should satisfy necessary conditions of *persistancy of excitation.* This means, that their excitation of the system, should be sufficiently rich such that all modes of the system are present in the output sequence, so that they can be identified. If for instants, the inputs are all constant, the behavior will soon look as being modelled by a static linear relationship. Another important issue concerns the frequency contents of the input sequence. Considerations on sampling rates, proper excitation and data record lengths strongly suggest that one should not aim at covering more than three decades of time constants in one experiment. For a high frequency model, low frequency dynamics look like integrators while high frequency dynamics for a low frequency model look like static instantaneous direct transmission terms [10].

3.2 The set of Models.

A set of candidate models is obtained by specifying within which collection of models we are going to look for a suitable one. This is no doubt the most important and, at the same time, the most difficult choice of any identification procedure. As much as possible, any *a priori* available information on the system, should be reflected in the choice of a certain model. If for instance, certain *physical* laws are known to hold true for the system, one could impose a certain equation structure and identify some unknwon physical parameters. In other cases, standard linear models may be employed, without reference to the physical background (the *black box* approach). In this paper, we shall review two possible black box models: input-output models (section 4 and 5) and state space models (section 6 and 7).

3.3 The criterion

Having determined the set of models, one should determine within this set, that model that is the 'best' approximation or provides the 'best' explanation of the observed data. The assessment of model quality is typically based upon how the models perform when they attempt to reproduce the measured data. One then has to fix a certain criterion which is to be optimized over all possible models within the model set. The determination of such a criterion is however not a trivial task: Typically, it is the modeller's desire to have models that are as *simple* as possible, i.e. of the least possible *complexity*, yet that at the same time *explain* as much as possible of the observed data, i.e. that minimize the *misfit.* These two requirements are in a certain sense conflicting: Intuitively, it is obvious that a simple model will not be able to explain or simulate complex behaviors while a complex model will explain a lot but will be difficult to identify or to use appropriately. We shall review several possible *complexity* determination criteria and *misfit* quantifications.

While the system identification procedure has the logical flow just described (collect data, fix a model set, pick out the best model), it is possible that the model does not pass the validation test and that one has to go back and revise several steps of the identification procedure. Hence one naturally arrives at a *system identification loop.*

For this reason, good interactive software is an important tool for handling the interactive character of the problem. The qualification 'good' implies the availibility of graphic possibili-

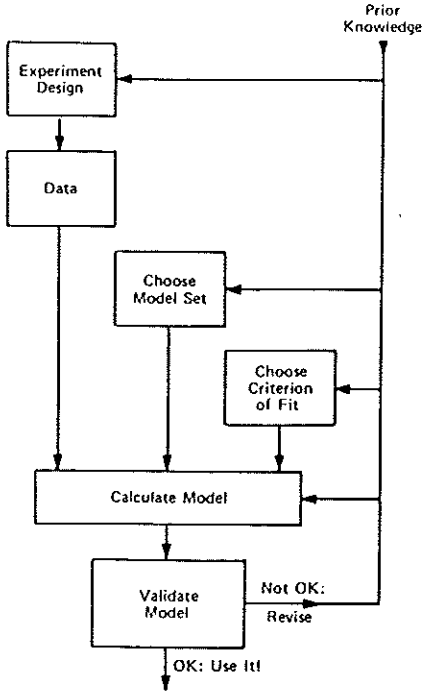

Figure 3: The system identification loop.

ties, the guarantee of numerical reliablitity and acceptable levels of computational speed and memory requirements. As an example of such software packages, one may consider *MATLAB* [11] and *SIMNON* [1], *Matrix-X*, *Control C*, etc...besides the more classical ones like NAG, LINPACK, EISPACK which are of course less interactive.

4 Input-output Models

The basic description of a single input, single output linear system with one additive disturbance is:

$$y_k = G(q)u_k + H(q)e_k$$

where the following notations are applied:

- The output at time k is y_k, while the input is u_k. e_k is a sequence of independent random variables with zero mean.
- q is the *forward shift operator* : $qu_k = u_{k+1}$.
  q^{-1} is the *backward shift operator*: $q^{-1}u_k = u_{k-1}$.
- $G(q)$ is the *transfer function* of the linear system. It is a rational function of the shift operator q^{-1}:

  $$G(q) = \frac{B(q)}{A(q)} = \frac{b_1 q^{-1} + b_2 q^{-2} + \ldots + b_{n_b} q^{-n_b}}{1 + a_1 q^{-1} + a_2 q^{-2} + \ldots + a_{n_a} q^{-n_a}}$$

  $G(q)$ describes the dynamic relation between the input sequence u_k and the output sequence y_k.
- $H(q)$ is a rational function of q as well:

  $$H(q) = \frac{D(q)}{C(q)} = \frac{1 + d_1 q^{-1} + d_2 q^{-2} + \ldots + b_{n_d} q^{-n_d}}{1 + c_1 q^{-1} + c_2 q^{-2} + \ldots + c_{c_a} q^{-n_c}}$$

$H(q)$ represents the transfer function from the unobserved sequence e_k to the output y_k.

The rationale behind this model is the following: The output sequence y_k is being thought as originating in the parallel connection of two systems: The first system models the causal, dynamic dependence of the outputs on the observed or predetermined inputs. Everything what can not be explained by a regression of these output observations on these inputs observations is then contributed to the second system. Obviously, several special cases may arise:

$n_a = n_b = n_d = 0, n_c \neq 0$: The general input-output equation reduces to:

$$y_k + c_1 y_{k-1} + \ldots + c_{n_c} y_{k-n_c} = e_k$$

It is easily seen that the output variables are essentially regressed on themselves, except for one residual term e_k per time instant. Hence the name *auto-regressive model* for this special case, abbreviated as **AR-model.**

$n_a = n_b = n_c = 0, n_d \neq 0$: The output can be written as:

$$y_k = e_k + d_1 e_{k-1} + \ldots + n_d e_{k-n_d}$$

The output at each time instant is described as a weighted average of present and past values of the sequence e_k, hence the name *moving average model* (**MA-model**).

$n_a = n_b = 0, n_c \neq 0, n_d \neq 0$: This is a combination of the two previous cases and is called an *auto-regressive moving average model* **ARMA-model.**

$n_a \neq 0, n_b \neq 0, n_c = n_d = 0$: This model represents a system without unobserved disturbances. It is a *linear difference equation.*

$n_a \neq 0, n_b \neq 0, n_c \neq 0, n_d \neq 0$ and $C(q) = A(q)$: In addition to the auto-regressive, moving average term, one has also a term describing the dynamic tranfer between the predetermined inputs u_k and the observed outputs with the same poles. This model is called **ARMAX**: *an auto-regressive moving average model with eXogeneous inputs.* Observe that there is a close connection with the so-called *Wold Decomposition* from the theory of stationary stochastic processes [7, p.265], which reduces to the observation that a stationary stochastic process may be decomposed into the sum of a deterministic (linearly predictable) and nondeterministic component, which is the output of a linear system driven by white noise.

$n_a \neq 0, n_b \neq 0, n_c \neq 0, n_d \neq 0$: This is one of the most general cases. This model is called the *Box-Jenkins Model Structure* [10].

The presented models can be generalized towards multivariable system representations (systems with several inputs and outputs). However, the corresponding necessary parametrizations are not very elegant and require a lot of notationally complex calculations. Instead of using rational functions, one now should employ rational matrices, i.e. matrices of which the elements are rational functions of the shift operator q. While these models can be embedded in a generalizing approach [7] [10], they lead frequently to numerically unstable behavior because of the bad conditioning of certain canonical parametrizations. This obviously represents a serious disadvantage for the use of input-output model for multivariable systems.

5 Identification of input-output models.

Assume that the system description is given in the form:

$$y_k = G(q)u_k + H(q)e_k$$

and assume that only observations on the variables u_k and y_k are available. The identification problem then reduces to the determination of appropriate transfer functions $G(q)$ and $H(q)$ and an appropriate sequence e_k, i.e. one should find appropriate orders of the polynomials of numerator and denominator of the transfer functions $G(q)$ and $H(q)$ and then compute coefficients a_i, b_i, c_i, d_i, such that the model sufficiently well approximates the input-output data.

5.1 Identifying the parameters.

We shall now derive an identification procedure for the case of a SISO system, described by the equation:

$$y_k = \frac{B(q)}{A(q)}u_k + \frac{D(q)}{C(q)}e_k$$

It is assumed that 'suitable' orders for the polynomials $A(q)$, $B(q)$, $C(q)$, $D(q)$ have been determined. This question will be discussed in section 5.2.
It is convenient to introduce the auxiliary variables:

$$\begin{aligned} w_k &= \frac{B(q)}{A(q)}u_k \\ v_k &= y_k - w_k \end{aligned}$$

It is starightforward to verify that:

$$w_k = b_1 u_{k-1} + \ldots + b_{n_b}u_{k-n_b} - a_1 w_{k-1} - \ldots - a_{n_a}w_{k-n_a}$$

From the definition of v_k it follows that the unobserved input e_k can be written as:

$$\begin{aligned} e_k &= \frac{C(q)}{D(q)}(y_k - \frac{B(q)}{A(q)}u_k) \\ &= \frac{C(q)}{D(q)}v_k \end{aligned}$$

Hence:

$$e_k = v_k + c_1 v_{k-1} + \ldots + c_{n_c}v_{k-n_c} - d_1 e_{k-1} - \ldots - d_{n_d}e_{k-n_d}$$

Now introduce the following 'state' vector:

$$\phi_k^t = (u_{k-1} \ldots u_{k-n_b} - w_{k-1} \ldots - w_{k-n_a} \;\; e_{k-1} \ldots e_{k-n_d} - v_{k-1} \ldots - v_{k-n_c})$$

and the parameter vector:

$$\theta^t = (b_1 \ldots b_{n_b}\, a_1 \ldots a_{n_a}\, d_1 \ldots d_{n_d}\, c_1 \ldots c_{n_c})$$

Observe that the 'state' vector contains only variables up to time $k-1$. It is also called the *regression vector*. One then arrives at the following expression for e_k:

$$e_k = y_k - \theta^t \phi_k$$

Observe that e_k has become a prediction error now. Hence it is natural to try to minimize in some sense this prediction error over all possible parameter vectors θ. This results in the following *least squares criterion:*

$$minimize_{over\ all\ \theta} \frac{1}{2N} \sum_{t=1}^{N} \epsilon_t^2$$

where

- N is the finite length observation horizon.
- $\epsilon_t = y_t - \theta^t \phi_t$ is the predicition error.

Its optimal solution $\hat{\theta}$ is a tedious though straightforward calculation (which may for instance be done via some elementary matrix algebra):

$$\hat{\theta} = [\frac{1}{N} \sum_{t=1}^{N} \phi_t \phi_t^t]^{-1} \frac{1}{N} \sum_{t=1}^{N} \phi_t y_t$$

Statistical properties of this (classical) estimation procedure such as *convergence behavior, consistency properties, asymptotic variance and covariance expressions* for the parameters estimates and *efficiency* are thouroughly investigated and analysed in [1] [7] [10] [12].

5.2 Determination of the model order.

The choice of an appropriate model structure is a difficult question but is most crucial for a succesful identification application. One may start with an inspection of the spectral contents of the transfer function. The resonance peaks and high frequency roll-off and phase shifts may contain useful information concerning the dynamic complexity. Another tool is the analysis of the rank of certain covariance matrices. Here, one tries to determine the order of the polynomials via the rank estimation of certain matrices. Other recently developed statistical order determination tests include *Akaike's Information Theoretic Criterion, Akaike's Final Predicition Error Criterion, Rissanen's Minimal Description Length fromalism*, all of which are discussed in [10, p.417-423] and in the references cited therein.
One may also test for the *whiteness* of the disturbance sequence e_k. Discrete time white noise is a random process of independent identically distributed random variables with a covariance function given by:

$$\begin{aligned} r(t) &= \sigma^2 \quad t = 0 \\ &= 0 \quad t \neq 0 \end{aligned}$$

Its spectral density is given by:

$$\phi(\omega) = \frac{\sigma^2}{2\pi}$$

and is thus constant for all frequencies, which explains the name from analogy with the spectral properties of white light.

Having explained as much 'energy' as possible of the output sequence y_k in terms of the input sequence u_k by an appropriate choice of the orders n_a and n_b, one now turns ones attention to the residual sequence $y_k - G(q)u_k$. This residual sequence is now explained by thinking that it originates in the output of another linear system, that is driven by *white noise*. One mau thus proceed as follows: Choose certain orders n_c and n_d and compute a possible sequence e_k that may have caused the residual sequence $y_k - G(q)u_k$. If this sequence e_k is not sufficiently white, one may decide to increase the complexity of the disturbance term, taking into account some threshold for the finite length of the observation window, by modifying the orders n_c and n_d, until the resulting sequence e_k is 'sufficiently white'. Numerical whiteness tests are e.g. included in MATLAB [11]. This idea of describing stochastic disturbances as linearly filtered white noise goes back to Wold (1938) [9]. However, in practice there will always exist a certain arbitraryness in the decision for what is to be contributed to the sequence u_k and what is to be explained in terms of a disturbance sequence, because of the noise that perturbs the observation and the finite number of observations.

In general, data-aided model structure selection is a largely underdeveloped research field for the input-output models. The cited tests form a serious drawback for a flexible application of the identification strategy based upon the input-output models. Indeed, this identification strategy is very well suited for updating time-variant models. One of its main advantages is its minimal computational complexity, especially in a recursive least squares implementation. The complexity determination tests however are rather time consuming and in some sense, destroy the computational elegancy of this input-output identification scheme. Yet, an appropriate order selection is unavoidable in order to guarantee the numerical stability of the least squares scheme. This can be seen from the expression for the least squares estimate of the parameter vector θ. A necessary condition is the non-singularity of the 'normal' equations' matrix R:

$$R = \frac{1}{N}\sum_{t=1}^{N} \phi_t \phi_t^t$$

If this matrix R is singular or close to a singular matrix, the inversion can not be done in a numerical reliable way. A possible remedy is then a decrease of the order such that the matrix becomes well-conditioned or inserting an additional random signal of limited energy, that sufficiently excites the system, such that the matrix R is non-singular. In any case, strategies to preserve a sufficient degree of *persistancy of excitation* in this input-output identification framework are the subject of vivid research up to date [1] [7] [10].

5.3 Recursive Least Squares Estimation.

In many cases, it is necessary or useful to have a model of the system available while the system is in operation, typically when the system to be modelled is non-linear or time-varying. In these applications, one will perform an on-line adjustment of the model each time a new measurement becomes available. The on-line computation of such a model must be done in such a way that the processing of the measurements from one sample can be completed during one sampling interval. Identification techniques that comply with this requirement

are called *recursive*: They exploit in the adjustment of the model, as much as possible the already obtained model and update the new one by a minimal modification. An interesting feature of the least squares methodology we have been discussing for input-output models is the fact that they are very well suited for a recursive implementation. Instead of giving a complete derivation, we shall restrict ourselves to mentioning only the final recursive scheme. For details, the interested reader is referred to [10] [1] [7].

A recursive linear least squares estimate of the parameter vector θ_k of the input-output model that minimizes the *weighted* least squares criterion:

$$\hat{\theta}_k = min_{over\ all\ \theta} \sum_{i=1}^{N} \lambda^{k-i}(y_i - \phi_i^t\theta)^2$$

is provided by the following recursive scheme:

$$\hat{\theta}_k = \hat{\theta}_{k-1} + L_k(y_k - \phi_k\hat{\theta}_{k-1})$$

$$L_k = \frac{P_{k-1}\phi_k}{\lambda + \phi_k^t P_{k-1}\phi_k}$$

$$P_k = \frac{1}{\lambda}(P_{k-1} - \frac{P_{k-1}\phi_k\phi_k^t P_{k-1}}{\lambda + \phi_k^t P_{k-1}\phi_k})$$

These expressions contain the following information:

- It can be seen that the parameter vector is updated in an additive way. The modification is determined by the *gain* L_k, which weights the amount by which the prediction error will affect the update.
- The covariance matrix P_k of the parameter vector, that reflects the confidence in the parameter estimation, can be determined recursively from the previous estimation of the covariance matrix, taking into account a 'new' regression vector ϕ_k.
- λ is a *forgetting factor*, which is chosen as $0 < \lambda < 1$. It represents an *exponential weighting*, by which older measurements are discounted. If λ is close to 1, the estimation will be *noise insensitive*, while if it is close to 0, the estimation of parameters will adopt itself much faster to modifications of the system but at the same time, it will of course be much more sensitive to perturbations and disturbances. Hence, an appropriate choice of λ will always be a matter of *compromis* between *fast adaptivity* and *noise insensitivity*. An example is given in figure 4. The first figure represents the identification of two parameters (one constant and one time varying) with no forgetting. For the second figure, the forgetting factor is $\lambda = 0.9$.

6 State Space models

A lumped linear time invariant discrete time system may be represented via the state space equations:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + w_k \\ y_k &= Cx_k + Du_k + v_k \end{aligned}$$

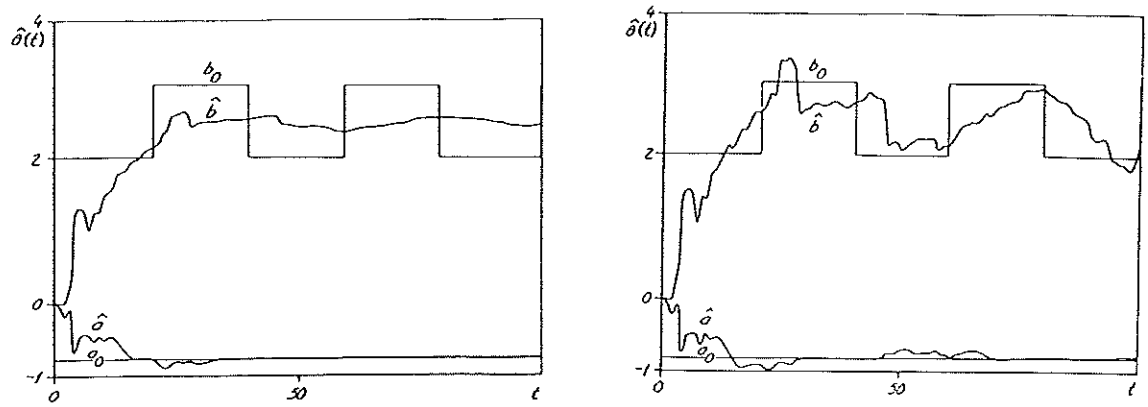

Figure 4: Adaptivity and forgetting factor

The first equation is referred to as the state equation while the second one is called the output equation. u_k is a vector with m components, that contains the m input observations at time k. The vector y_k contains the l outputs. The vector x_k contains the n states at time k. The vectors v_k and w_k both represent disturbances. v_k is referred to as *measurement noise* while w_k is called the *process noise*, which acts on the states. In some applications, typical variables that can be chosen as states are:

system	state 1	state 2	state 3
mechanical	position	velocity	
electrical	voltage	current	
hydraulic	levels	flows	
thermal	temperatures	pressures	densities

Observe however that a state space model is not unique with respect to an observed input-output behavior. Indeed, it is easily verified that an 'equivalent' state space model is obtained by inserting a non-singular matrix T as follows:

$$\begin{aligned} Tx_{k+1} &= (TAT^{-1})(Tx_k) + (TB)u_k + Tw_k \\ y_k &= (CT^{-1})(Tx_k) + Du_k + v_k \end{aligned}$$

It can be observed that the matrices have been changed but that the input-output pairs were not affected. This implies that only for very specific choices of coordinate systems in the vector space of states, the components of the state vectors have physical meaning. However, in most applications. it is not really important whether a state vector component has a real physical meaning.

State space models have certain advantages with respect to input-output models.

- Conceptual problems are most easily studied within the state space framework. Examples include the investigation of properties such as *observability*, *controllability*, *stabilizability*, etc.... [8].

- Optimal control and filtering is completely solvable in terms of state space models (Kalman filter, the linear quadratic regulator, etc...) [8], more specifically, optimal control requires state feedback, which on its turn necessitates the optimal estimation of the state vector.

- In the direct identification of a state space model one may expect to be able to identify observable but non-controllable phenomena such as periodical interference, drift, offsets etc.... Such autonomous phenoma are not identifiable with input-output models. Of particular interest is the definition of *observable excited dimension*:

  Definition 1 Observable excited dimension
  Assume that over a certain time horizon N the states are aggregated in an $n \times N$ matrix X:

  $$X = (x_k \; x_{k+1} \; \ldots \; x_{k+N-1})$$

  The observable excited dimension, denoted by n_{oe} is then simply the algebraic rank of the matrix X.

  A more intuitive interpretation is that the observable excited dimension is nothing else than the numbers of (complex) modes of the system that has been excited and that is observable (and hence may be measured) in the output sequence.

- At the cost of a higher computational complexity, the identification of state space models is numerically more robust than that of input-output models and problems such as *peristancy of excitation of the input signals* are easier to assess. As a matter of fact, recently it has been shown that it is even not needed to know *a priori* which of the variables are to be considered as inputs and which of them are outputs [3].

7 Identification techniques for state space models.

A trivial way to obtain state space models from input-output measurement could consist of a transformation of an obtained input-output model into a state space form. It can be demonstrated that input-output models can be converted into state space models and vice versa [8] via so-called *canonical parametrizations* of state space models. However, an essential drawback of this procedure is the numerical instability of these state space canonical forms: small perturbations in one of the parameters, may influence the quality of the model in a dramatic way. Hereto, state space models should be identified *directly* instead of indirectly. At the cost of a higher computational complexity, state space modelling is characterized by a better numerical behavior from the application of numerically robust techniques. Recall that also the order determination of an input-output model was not a trivial task. It will be shown how this problem is relatively easier for state space identification methods, by using the *singular value decomposition.* Finally, let's mention that adaptive versions of this state space identification may be derived though they are still subject to further research [4].

The *direct* identification of state space models from input-output measurements essentially exploits three basic techniques from numerical linear algebra:

1. The solution of a set of linear equations

2. The eigenvalue decomposition

3. The singular value decomposition

While the first two of them may be qualified as *classical*, the third one was introduced only recently as a practical numerical tool by the development of an efficient algorithm ([6] and the references therein).

7.1 The Singular Value Decomposition.

The singular value decomposition (SVD) is a matrix factorization technique, in which a $p \times q$ matrix X can be factorized as the product of three matrices U, S, V, each of which has some interesting features:

$$X = USV^t$$

In this factorization:

- S is a $p \times q$ diagonal matrix, containing the singular values of the matrix X:

$$S = \begin{pmatrix} s_1 & 0 & \dots & \dots & 0 \\ 0 & s_2 & 0 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & s_r & \dots \\ 0 & \dots & \dots & \dots & 0 \\ 0 & \dots & \dots & \dots & 0 \end{pmatrix}$$

  By convention, the singular values are ordered in non-decreasing order:

$$s_1 \geq s_2 \geq \dots \geq s_r > 0$$

  The smallest non-singular value s_r reveals the *algebraic rank* of the matrix X. This is equivalent with the number of linearly independent rows (columns) in the matrix X. (Recall the a row (column) is linearly independent of all other rows (columns) if it cannot be written as a linear combination of these other rows (columns)).

- The matrix U is a $p \times p$ orthonormal matrix:

$$U^tU = I_p = UU^t$$

  Its columns $u^i, i = 1, \dots, p$ are called the left singular vectors.

- The matrix V is a $q \times q$ orthonormal matrix:

$$V^tV = I_q = VV^t$$

  Its columns $v^i, i = 1, \dots, q$ are called the right singular vectors.

Denoting by U_1 the $p \times r$ submatrix of U consisting of its r first columns and by V_1 the $q \times r$ submatrix of V formed by its first r columns and by S_1 the $r \times r$ right upper part of S, one may write down a *reduced* singular value decomposition of the form:

$$X = U_1 S_1 V_1^t$$

where now S_1 is square diagonal and U_1 and V_1 satisfy the 'reduced' orthonormality conditions:

$$U_1^t U_1 = I_r = V_1^t V_1$$

It is easy to demonstrate that this factorization may also be converted into the following sum, which is called the *dyadic decomposition*:

$$X = u^1 s_1 (v^1)^t + \ldots + u^r s_r (v^r)^t$$

Observe that this decomposes the matrix X of rank r in r rank one matrices of decreasing importance because $s_i \geq s_{i+1}$. This observation is a crucial one for the succesful application of the singular value decomposition in *data reduction applications*. In a lot of these applications, one needs only some of the first terms of the dyadic decomposition in order to *approximate* the matrix X. An important optimality property is the following. Denote by $\|X\|_F^2$ the so called *Frobeniusnorm* of a matrix, which is the sum of its elements squared, then the SVD provides the solution to the following optimization problem:

$$\min_{\hat{X}} \|X - \hat{X}\|_F^2$$

subject to the constraint that

$$rank(\hat{X}) = rank(X) - 1$$

If the dyadic decomposition of the matrix X is given as above, the solution is simply:

$$\hat{X} = \sum_{i=1}^{r-1} u^i s_i (v^i)^t$$

One of the most interesting properties of the singular values is there extreme insensitivity to additive perturbations. This implies that, when the data in the matrix are noisy, the singular values will still reveal the rank of the unperturbed matrix if of course the signal-to-noise ratio is not too bad. As an example, consider the following 50×5 matrix:

$$X = \begin{pmatrix} 1 & 50 & 51 & -102 & 52 \\ 2 & 49 & 51 & -102 & 53 \\ 3 & 48 & 51 & -102 & 54 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 50 & 1 & 51 & -102 & 101 \end{pmatrix}$$

It is easy to verify that the third column is the sum of the first two columns, the 4-th one is 2 times the 3-th one, the last one is the sum of the first and the third column. Hence the algebraic rank of the matrix X is 2. Now random noise is added to this matrix. Each element is corrupted by normally distributed zero mean random noise with variance 1. The singular values of the exact and the perturbed matrix are:

	exact	noisy
1	1005.5	1004.7
2	167.7	167.7
3	0.6E-14	6.57
4	0.3E-14	6.23
5	0.5E-33	5.51

The numerical precision equals $0.1E-15$. Hence, from the 'gap' in the singular spectrum, one may still conclude that the perturbed noisy version is close to a matrix of rank 2. Typical is the so-called noise threshold: The smallest singular values are caused only by the noise and are all of the same order of magnitude.
These properties and more details and algorithms may be found in [6] [3]. This observation together with the data reduction property will be exploited in the identification of state space models.

7.2 Some essential matrix properties for dynamic systems.

The direct identification approach for state space models will be based upon three basic properties. The first one allows to estimate the number of necessary states of the system while the second property allows to obtain a state sequence. The third property then shows how the system matrices and the disturbance vectors w_k and v_k follow from the solution of a set of linear equations.

Consider the vectors w_k that are constructed from input-output pairs by simple concatenation:

$$w_k = \begin{pmatrix} u_k \\ y_k \end{pmatrix}$$

Second, construct 2 so-called block Hankel matrices with these input-output pairs:

$$H_1 = \begin{pmatrix} w_k & w_{k+1} & w_{k+2} & w_{k+3} & \cdots & \cdots & w_{k+j-1} \\ w_{k+1} & w_{k+2} & w_{k+3} & w_{k+4} & \cdots & \cdots & w_{k+j} \\ w_{k+2} & w_{k+3} & w_{k+4} & w_{k+5} & \cdots & \cdots & w_{k+j+1} \\ w_{k+3} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ w_{k+i-1} & w_{k+i} & \cdots & \cdots & \cdots & \cdots & w_{k+i+j-2} \end{pmatrix}$$

The *block dimensions i and j* are user determined dimensions which should be chosen 'sufficiently large' (for more detail, [3]). Observe that the block Hankel structure reveals itself in the repetition of the same vector along the anti-diagonals of the matrix. Since the dimensions of the vectors w_k are $(l+m)\times 1$, the dimensions of the block Hankel matrix are $(l+m)i \times j$.
A second block Hankel matrix is constructed from subsequent observations:

$$H_2 = \begin{pmatrix} w_{k+i} & w_{k+i+1} & w_{k+i+2} & w_{k+i+3} & \cdots & \cdots & w_{k+i+j-1} \\ w_{k+i+1} & w_{k+i+2} & w_{k+i+3} & w_{k+i+4} & \cdots & \cdots & w_{k+i+j} \\ w_{k+i+2} & w_{k+i+3} & w_{k+i+4} & w_{k+i+5} & \cdots & \cdots & w_{k+i+j+1} \\ w_{k+i+3} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ w_{k+2i-1} & w_{k+2i} & \cdots & \cdots & \cdots & \cdots & w_{k+2i+j-2} \end{pmatrix}$$

The block Hankel matrix H_1 is called the *past input-output block Hankel matrix* while H_2 is called the *future input-output block Hankel matrix.* They are both submatrices of the 'big' *input-output block Hankel matrix* H:

$$H = \begin{pmatrix} H_1 \\ H_2 \end{pmatrix}$$

7.3 Estimation of the observable excited dimension

The following theorem allows to estimate the order of the system from the singular values of the input-output block Hankel matrices.

Theorem 1 The rank property
Under fairly general conclusions, it holds that:

$$rank(H_1) = rank(H_2) = mi + n_{oe}$$

and

$$rank(H) = 2mi + n_{oe}$$

where n_{oe} is the excited observable dimension of the state sequence:

$$X = (x_{k+i}\ x_{k+i+1}\ \cdots\ x_{k+i+j-1})$$

For a proof and a detailed analysis of the necessary conditions, the reader is referred to [3].

Observe that the singular value decomposition of for instance the matrix H_1 allows to obtain an estimate of the observable excited dimension n_{oe} from a simple counting of the non-zero singular values. Moreover, in the presence of additive noise, the noise insensitivity of these singular values will still allow to 'recognize' the order of the system. An important result is the property that, when *linear static state variable feedback* is applied, this will cause a decrease of the observable excited dimension. Recall that linear static state veriable feedback consists of the determination of the inputs as a linear static function of the states. This implies that the inputs (or better, those variables that are considered to be inputs) are no longer *independent* variables, but have become completely dependent on other variables of the system. This contradicts the intuitive appreciation of an input as being a completely free variable. That this property is not merely intiution is revealed in the decrease of the observable excited dimension. Hence, when operating under linear static (partial) state variable feedback, the number of non-zero singular values of the input-output block Hankel matrix will be less than when the system operates in *open loop*.

Example
This examples aims at illustrating the several possible situations that may influence the rank input-output block Hankel matrix. Hereto consider the following system with 2 inputs, 1 output and 3 states:

$$A = \begin{pmatrix} \sqrt{(3)}/2 & -1/2 & 0 \\ 1/2 & \sqrt{(3)}/2 & 0 \\ 0 & 0 & -0.92 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 0 \\ 2 & 0 \\ 0 & -3 \end{pmatrix}$$

$$C = (\,1\ 1\ 1\,) \qquad D = (\,0\ 0\,)$$

The initial state equals $x_0 = [\,0\ 0\ 0\,]^t$. The block Hankel parameters are $i = 6$ and $j = 19$. Hence U_h is a 12×19 while Y_h is a 6×19 block Hankel matrix. The system is identified adaptively by a gliding window approach. The window is 24 time steps long: e.g. the first window extends from time 1 to 24, the second from 2 to 25, etc .... In each window the system is considered to be time invariant. The first identification result is obtained at time

step 24. If we refer to a certain time step, this corresponds to the time step at with the identification is obtained, For instance, the identification at time 33, uses data from $k = 10$ to $k = 33$. The history of the system is the following:

Time step 24 - 40 : The first input is zero. The second one is a random signal, Gaussian distributed with mean zero and variance 1. The rank of the input block Hankel matrix is 6. The observable excited dimension is 1.

Time step 41 - 80 : A defect sensor changes the matrix C into $C = [\ 1\ 1\ 0\ ]$ at time 41. This makes the third state unobservable. Between timestep 41 - 64, the rank of the input-output block Hankel matrix reaches a maximum of 11. The reason is that the gliding window identification tries to model the timevariance of the system, by a linear time-invariant system but of higher complexity. Remarkable enough, for this example, it succeeds succesfully: The singular values indicate that the behavior could be modelled by a linear time-invariant system of order 6. The transient phenomenon disappears from time step 64 on, because from then on, the data in the gliding window were generated by the non-observable time invariant system. Observe that the emerging and disappearing of the singular values proceeds rapidly.

Time step 81 - 120 : The second input is activated into a step-signal of amplitude 1 at time 81. This excites the states 1 and 2. Because the corresponding eigenvalues are on the unit circle, the step response is a linear oscillation. Observe that between time steps 81 and 86 the rank of the block input Hankel matrix increases up to its maximum, which is 12. The reason is that the step response gradually moves in into the input block Hankel as the window proceeds. This maximal rank persists until time step 99. At time 100, the rank starts to decrease again because of the constant first input. At time 105, the rank of the input block Hankel equals 7. Meanwhile, it is easily derived that the observable excited dimension equals 2.

Time step 121 - 160 : Because a new sensor has been installed and activated, the system has again become observable at time step 121. However, the new sensor is not of high quality -despite its price!) : it shows drift which happens to be (for this cheap sensor) a first order automous unstable behavior with a pole at 1.01. The identified rank between 121 and 143 reflects the attempt of the identification algorithm to model the time variance (transition from unobservability to observability) by a higher complexity. However, this transient behavior disappears from time step 144 on. Between time 124 and 160, the rank of the concatenated input-output block Hankel matrix equals 11, which can be explained as follows: Rank 1 from the first input (a constant), rank 6 from the second input (random signal),observable excited dimension 2 of the oscillation, observable excited dimension 1 from the third state, which became observable again (the new sensor) and observable excited dimension 1 of the sensor's drift.

Time step 161 - 200 : In an attempt to stop the oscillation, an operator closes the open loop system with a static state variable feedback controller of the form $u_k = -Fx_k$, applied to the first 2 states and the first input, where $F = (\ -0.25\ \ 0.5\ )$. The controller succeeds in damping the oscillation. Again, between 161 and 183, there is a transient rank increase and decrease due to time variance. From 184 on, the rank stabilizes again. The input is no longer independent. Observe that the drift is now very well recognizable

in the output. The input-output rank now equals 10 = 6 (second input) + 2 (first input) + 4 (observable excited dimension) - 2 (dependency of the input).

7.4 Computation of a state sequence.

It will now be demonstrated how a appropriate state sequence can be computed from the past and future input-output block Hankel matrices.

Theorem 2 The state as intersection of past and future.
If the observable excited dimension of the state sequence X equals n_{oe}, where:

$$X = (x_{k+i}\ x_{k+i+1}\ x_{k+i+j-1})$$

then there exist matrices P_1 and P_2, both of rank n_{oe}, such that:

$$X = P_1 H_1 = P_2 H_2$$

For a detailed proof, the reader is again referred to [3]. However, observe that the matrix P_1 makes linear combinations of the rows of the past input-output block Hankel matrix H_1 while the matrix P_2 does the same with the rows of the future input-output block Hankel matrix. These two matrices P_1 and P_2 can be computed from the singular value decomposition of the input-output block Hankel matrix H from the following observation:

$$(P_1\ -P_2)\begin{pmatrix} H_1 \\ H_2 \end{pmatrix} = 0$$

which can be written as:

$$PH = 0$$

with an obvious definition of the matrix P. From this observation, it is seen that the matrix P can be constructed from the left singular vectors of the matrix H, that correspond to zero singular values. It can be shown that a second singular value decomposition then provides the matrices P_1 and P_2 and the state sequence X [3].

7.5 From the state sequence to the model.

By a repeated application of the model equations, it is easy to derive that:

$$\begin{pmatrix} x_{k+1} & x_{k+2} & \cdots & x_{k+p+1} \\ y_k & y_{k+1} & \cdots & y_{k+p} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} x_k & x_{k+1} & \cdots & x_{k+p} \\ u_k & u_{k+1} & \cdots & u_{k+p} \end{pmatrix} + \begin{pmatrix} w_k & w_{k+1} & \cdots & w_{k+p} \\ v_k & v_{k+1} & \cdots & v_{k+p} \end{pmatrix}$$

In these equations p is a user defined parameter which is again 'sufficiently large'. Both input and output variables are known, and in the previous section it was outlined how a state sequence may be obtained. The disturbances v_k and w_k are unknown. One may however obtain an estimation by exploiting the least squares idea as was the case with the input-output models: Hereto, one simply solves in a least squares sense the system of equations:

$$\begin{pmatrix} x_{k+1} & x_{k+2} & \cdots & x_{k+p+1} \\ y_k & y_{k+1} & \cdots & y_{k+p} \end{pmatrix} = \begin{pmatrix} \hat{A} & \hat{B} \\ \hat{C} & \hat{D} \end{pmatrix}\begin{pmatrix} x_k & x_{k+1} & \cdots & x_{k+p} \\ u_k & u_{k+1} & \cdots & u_{k+p} \end{pmatrix}$$

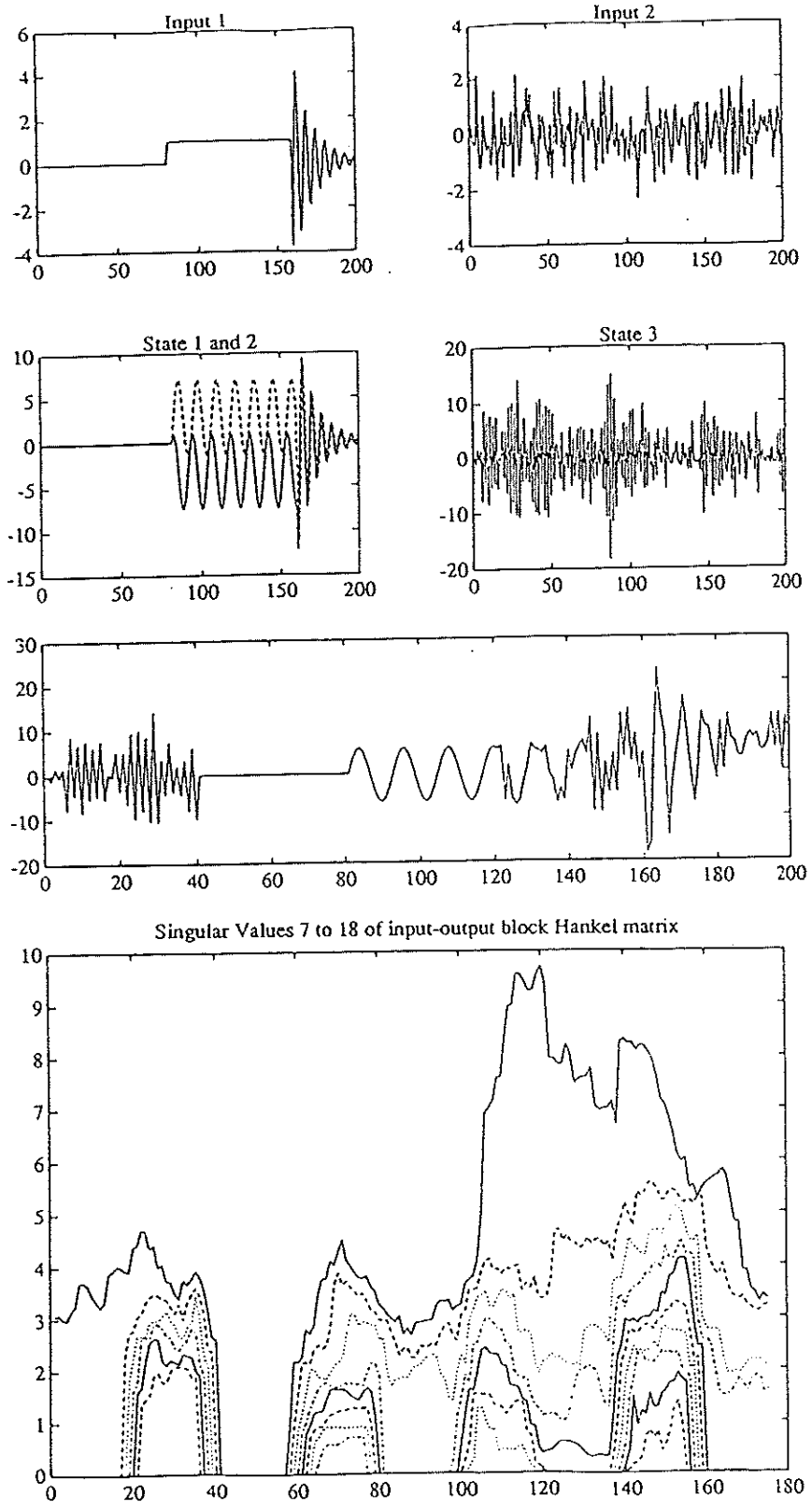

Figure 5: The inputs, the states, the outputs and singular values 7 to 18 of the input-output block Hankel matrix as a function of time for the timevarying 4-th order system.

where the $\hat{}$ denotes the obtained estimates. The residuals:

$$\begin{pmatrix} x_{k+1} \\ y_k \end{pmatrix} - \begin{pmatrix} \hat{A} & \hat{B} \\ \hat{C} & \hat{D} \end{pmatrix} \begin{pmatrix} x_k \\ u_k \end{pmatrix}$$

are then considered to be estimates of the disturbances. In this way, one has obtained a complete identification of the system matrices and the disturbances. Additional properties, more sophisticated approaches based upon the (generalized) singular value decomposition and possible adaptive implementations are derived and analysed in [3] [4].

8 Model Validation

Model validation is concerned with the question whether the model is 'good enough' for its purpose. Good performance will develop a certain confidence in the model: For instance, if the analysed system is known to be time-invariant and approximatively linear, then one expects at least that the model will sufficiently well *predict* the behavior when inputs are applied. Hence a validation criterion could be the energy in the difference between the observed and the simulated output. Other measures could be formulated in terms of frequency behavior, as for instance the accuracy of the amplitude and the phase spectrum of the simulations compared to the observed ones (see also [10] [1] [7]). We shall only present here 3 examples of a succesful identification and corresponding validation.

8.1 A power plant

Input-output measurements on a 120 MW power plant of Pont-Sur-Sambre in France were obtained. The available input data were samples obtained under *normal operating circumstances* of the gas flow, the turbine valves openings, the super heater spray flow, the gas dampers and the air flow. The 3 outputs are the steam pressure, the main steam temperature and the reheat steam temperature. Inputs and outputs are depicted in figure 6. The sampling time was 120 seconds. Using a refinement of the matrix algebra technique presented above [3, p.300], 4 different models were derived with an increasing complexity. It can be concluded from figure 7 that the quality of the model ameliorates as the complexity increases. The block dimensions used were $i = 5, j = 90$.

8.2 A glass production installation

A feeder is the final part of a process installation that is used for melting glass. Its main task is to realize a homogeneous temperature distribution. The reader is invited to consult [2, p.193] for an in depth discussion of the data acquisition and preprocessing procedures. Input 1 is the gas input of the first feeder, input 2 is a cooling air input while input 3 is the gas input of the second feeder. Pseudo-random binary sequences were applied as inputs, ensuring almost surely a sufficient degree of persistancy of excitation. The first 300 samples of these inputs are depicted in figure 8. As outputs of the process, the glass temperature at 6 different locations in a cross section of the feeder was measured. The block dimensions of the input-output block Hankel matrix used, were $i = 10, j = 300$. The singular spectrum of the $2(m+l)i \times j$ (=180 x 300) input output block Hankel matrix is depicted in figure 8. The singular values from the $(2mi+1) = 61$-th on are depicted in the right hand side of figure 8. They allow to determine an approximate order, which for this example was fixed as 4. Results

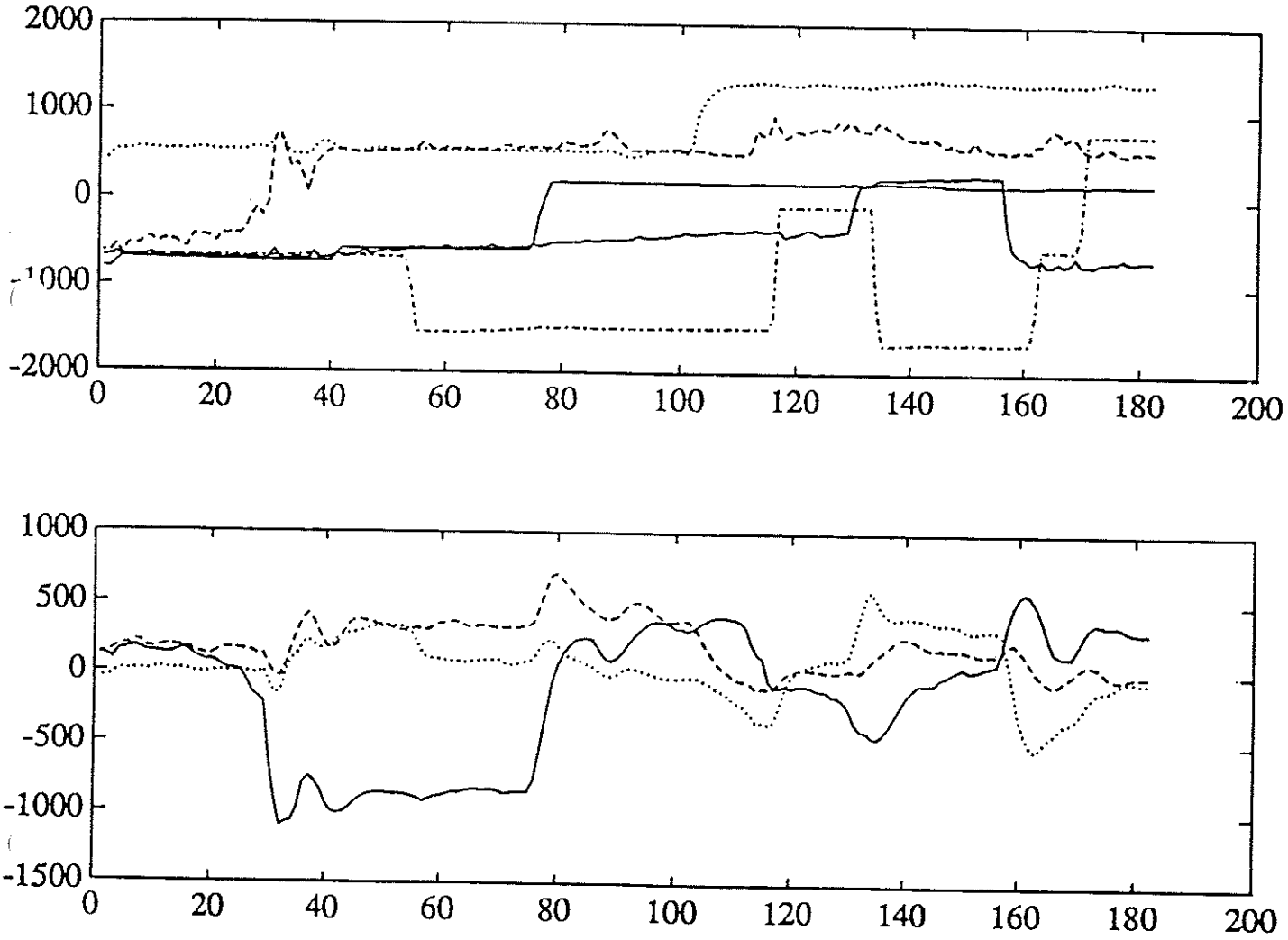

Figure 6: Inputs and outputs of a power plant

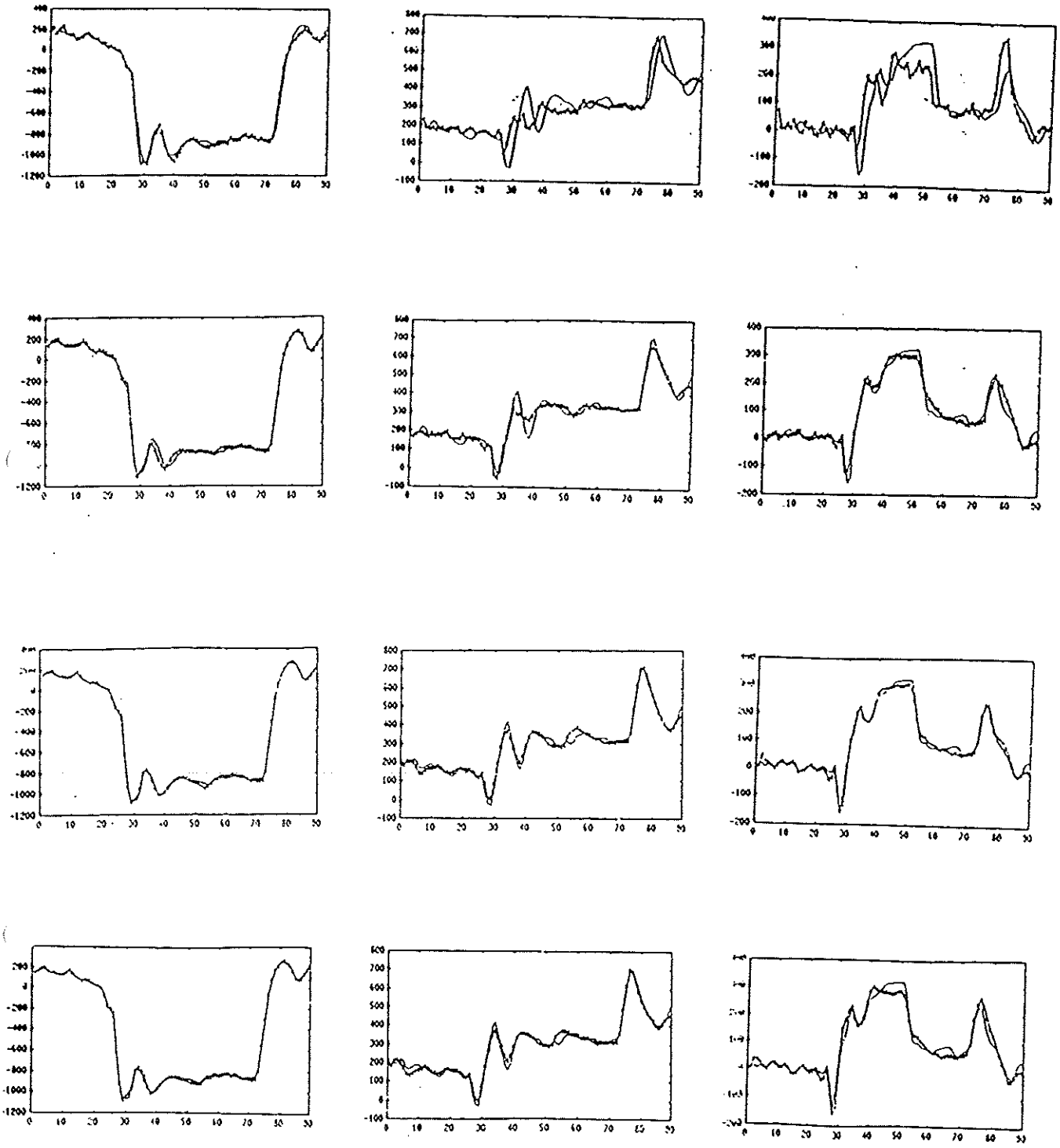

Figure 7: Measured outputs (full line) and simulations (stars) for a (a) first order, (b) 4-th order, (c) 7-th order, (d) 9-th order approximation

of simulations based on the 300 used input-output pairs can be found in figure 9. In figure 9, one also finds a prediction of the first and the 4-th output from time step 800 to 1000. The prediction of output 1 was the worst, while that of output 4 was the best of all predictions. Obviosuly, the model is rather good despite an offset in simulation of the first output.

8.3 A simulated example and the importance of scaling.

Consider the simulated input-output behavior (with PC-Matlab [11]) of a 4-th order system as depicted in figure 10. Observe that the amplitude of the output is at least one order of magnitude larger than this of the input. The singular values of a block Hankel matrix with the input-output vector pairs are depicted in figure 11 from the mi-th singular value on. Figure 11.a represents the singular values from the input-output block Hankel matrix without any scaling. In figure 11.b, the data were scaled such that all rows of the block Hankel matrix have equal energy. In figure 11.c, the data were scaled such that the variance of the noise on input and output is equal. As can be concluded, scaling considerably influences the singular spectrum.
Then the system was identified with the recursive input-output least squares identification scheme of section 5. The phase and amplitude plots of the transfer function of the impulse response for increasing order, can be found in figures 12 (a), (b), (c) , (d).

9 Conclusions

In this survey, we have presented an introduction to the identification of linear models from input-output data. First, we have paid attention to the several possible types of models and their respective properties. Next, attention was paid to some practical issues concerning identification experiments and data acquisition. We have also summarized the main features of two basic identification approaches: one was based on input-output models and least squares algorithms, the other was based upon state space models and matrix algebra. Finally we have provided several examples of identification applications in an industrial environment.

The author is indebted to *Marc Moonen, Jan Swevers, Lieven Vandenberghe and Joos Vandewalle* for the many interesting discussions and stimulating suggestions and for the preparation of some of the figures.

References

[1] Astrom K., Wittenmark B. *Computer Controlled Systems.* Prentice-Hall, Englewood Cliffs, N.J., 1984.

[2] Backx T. *Identification of an industrial process: A Markov Parameter Approach.* Doctoral Thesis, Technische Universiteit Eindhoven, November 1987.

[3] De Moor B. *Mathematical Concepts and Techniques for Modelling of Static and Dynamic Systems.* Doctoral Dissertation, Katholieke Universiteit Leuven, Electrical Department, June 1988.

[4] Moonen M., De Moor B., Vandenberghe L., Vandewalle J. *On- and off-line identification of linear state space models.* Accepted for the International Journal of Control, 1988.

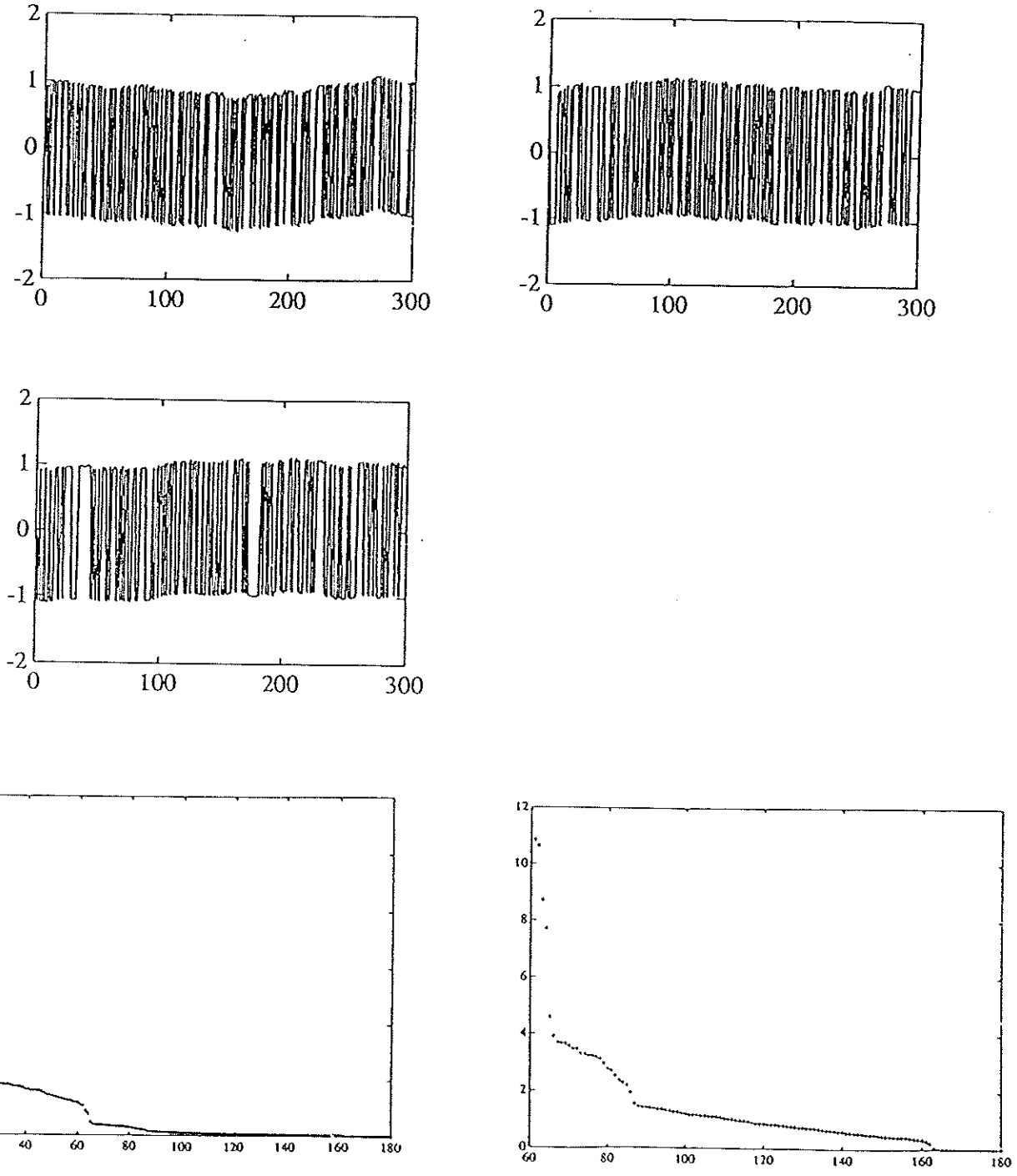

Figure 8: Three inputs of the feeder, singular spectra of the input-output block Hankel matrix.

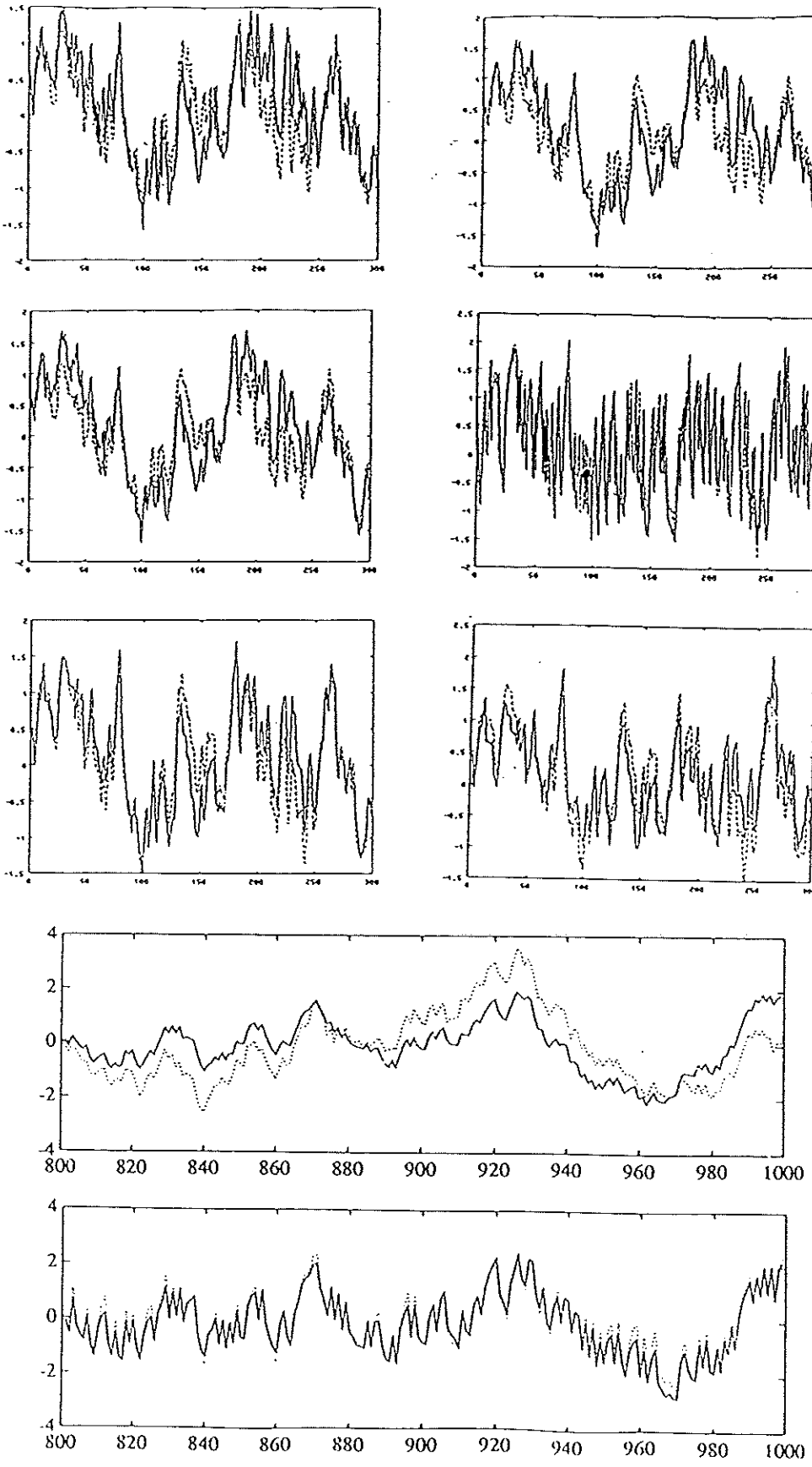

Figure 9: Simulation of the 6 outputs (measured=full line) and predicition of output 1 and 4 (measured=full line)

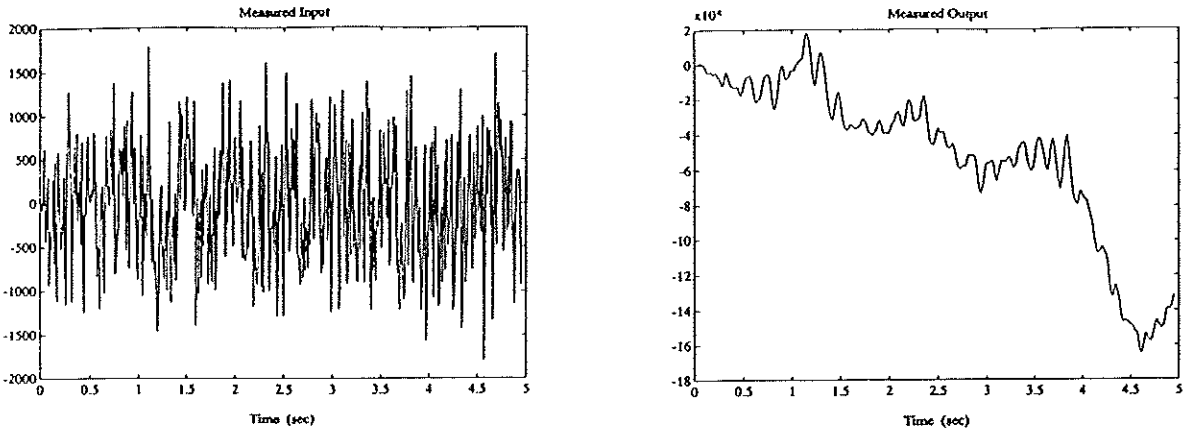

Figure 10: Input-output of a 4-th order system

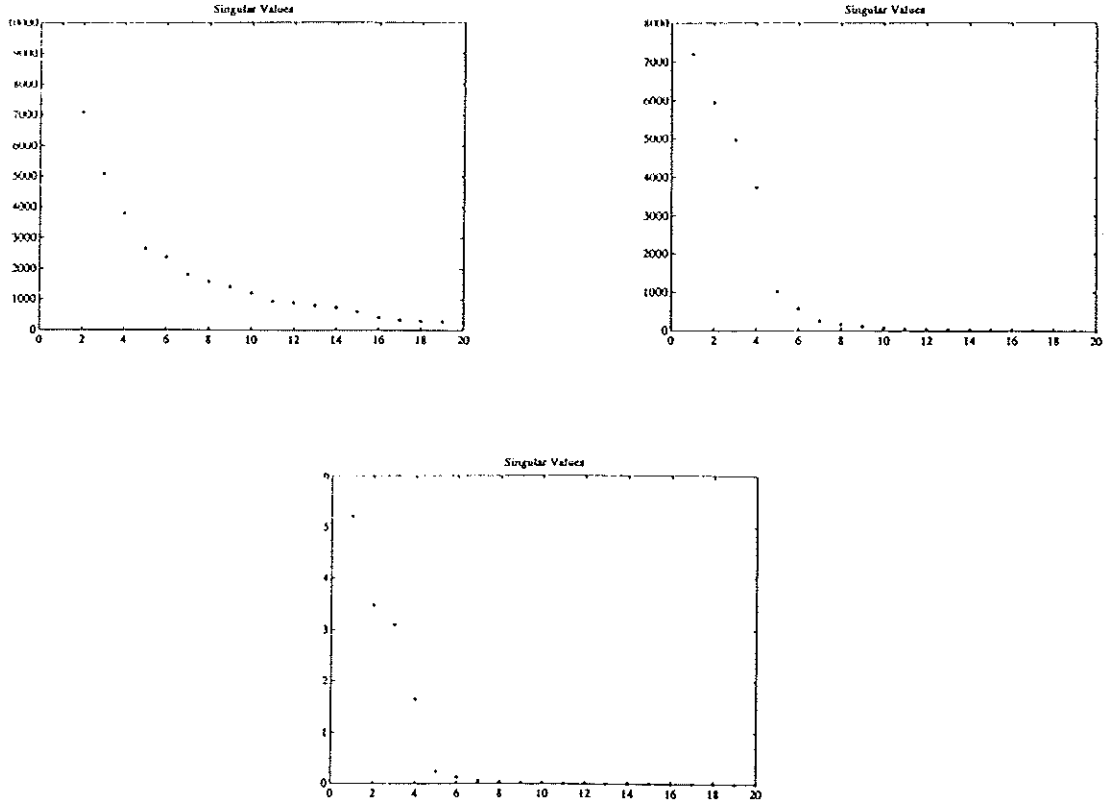

Figure 11: Singular spectra for different scaling strategies.

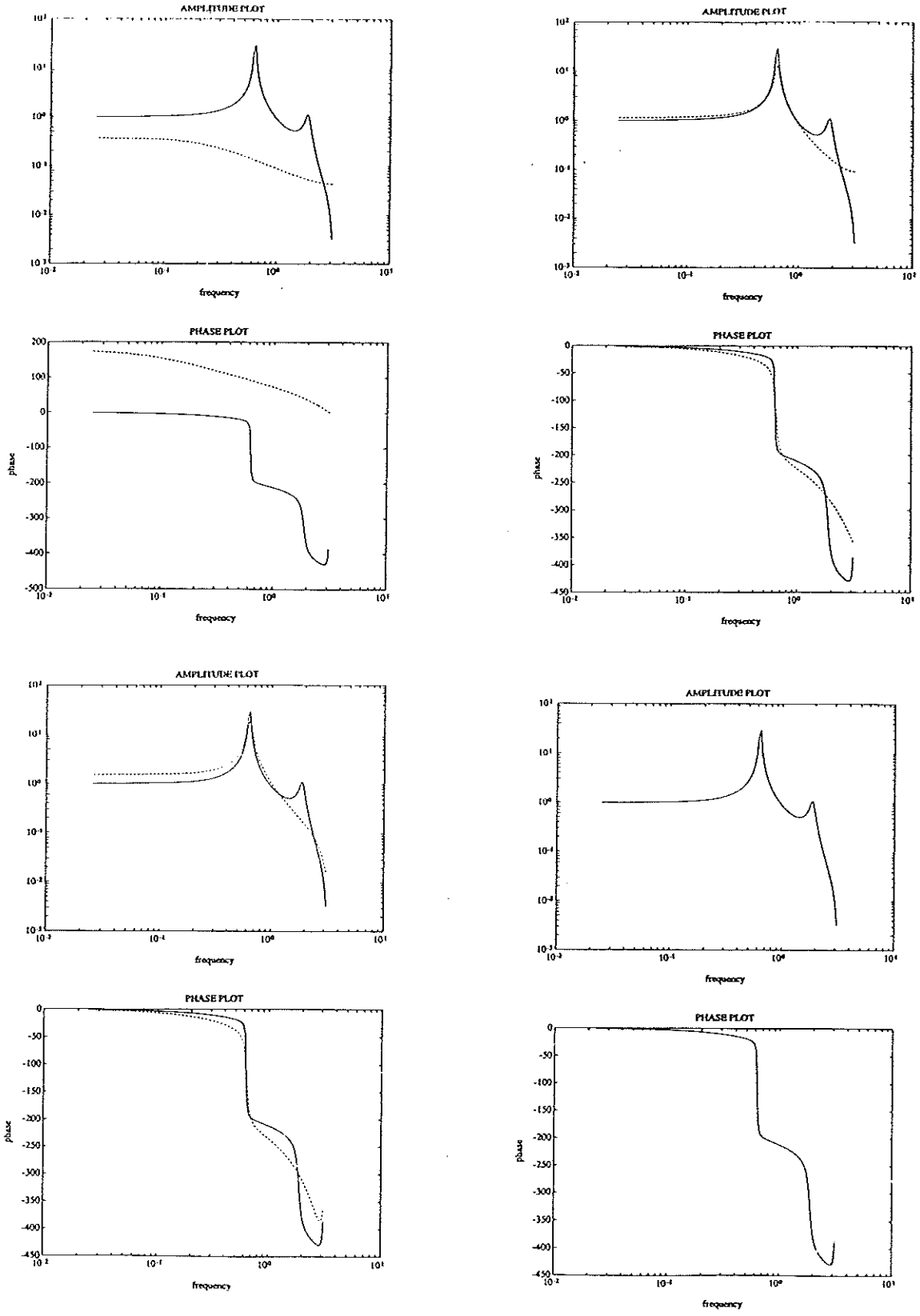

Figure 12: Phase and amplitude plots for (a) first order, (b) 2-th order, (c) 3-th order, (d) 4-th order approximations

[5] Eykhoff P. *System Identification.* Wiley, New York, 1974.

[6] Golub G., Van Loan C. *Matrix Computations.* North Oxford Academic Publ. Co., John Hopkins University Press, 1983.

[7] Goodwin G.C., Sin K.S. *Adaptive filtering, prediction and control.* Prentice-Hall Information and System Sciences Series, (Ed.T. Kailath), Prentice Hall, Inc., Englewood Cliffs, New Jersey 07632, 1984.

[8] Kailath T. *Linear Systems.* Prentice-Hall, Englewood Cliffs, NJ, 1980.

[9] Kailath T. *A view on three decades of linear filtering theory.* IEEE Trans. on Information Theory, vol.IT-20, pp.146-181, 1974.

[10] Ljung L. *System Identification: Theory for the User.* Prentice Hall Information and System Sciences Series (Ed. T. Kailath), Prentice-Hall, Inc., Englewood Cliffs, New Jersey 07632, 1987.

[11] Pro-Matlab, VAX/VMS version 2.1., VMS. December 1986, Mathworks Inc. MA, USA.

[12] Norton J.P. *An introduction to identification.* Academic Press, Harcourt Brace Jovanovich Publishers, London, 1986.