

Subspace system identification

J. Poshtan and H. Mojallali

Abstract: We give a general overview of the state-of-the-art in subspace system identification methods. We have restricted ourselves to the most important ideas and developments since the methods appeared in the late eighties. First, the basis of linear subspace identification are summarized. Different algorithms one finds in literature (Such as N4SID, MOESP, CVA) are discussed and put into a unifying framework. Further, a comparison between subspace identification and prediction error methods is made on the basis of computational complexity and precision of methods by applying them to a glass tube manufacturing process.

Keywords: subspace methods, system identification, state-space methods, multivariable systems, linear algebra.

1 Introduction

Mathematical models of dynamical systems are used for analysis, simulation, prediction, optimization, monitoring, fault detection, training and control. There are several approaches to generate a model of a system. One could for instance start from first principles, such as writing down basic physical or chemical laws that generate behavior of the system. This so called white box approach works for simple examples, but its complexity increases rapidly for real-world systems. In some cases, equations of the system are known up to within some unknown parameters, which are estimated using some parameter-estimation method (gray box modelling). Both, two aforementioned techniques require the extraction of system equations which is difficult in MIMO systems due to system variables interaction and often leads to several trial and error stages. Another approach is provided by system identification, in which measurements or observations are first collected from the system, and then modeled using a so called black-box identification approach. Such an approach basically consists of first defining a parameterization of the model, and then determining the model parameters in such a way that the measurements are explained as accurately as possible by the model. In system identification, parameterization and model selection are important. [1], [2], [9], [10], [11] represent some techniques for model parameterization. [8] represents a kind of parameterization called observability model, in which the user should ultimately select a desired model from model sets. All of proposed techniques, with their advantages and disadvantages, for starting require prior knowledge such as system order. Therefore, usage of the these techniques take a lot of time and operations and several trial and error stages must be done. The beginning of the 1990s witnessed the birth of a new type of linear system identification algorithms, called subspace methods. Subspace methods basically originate in a good combination between system theory, geometry and numerical linear algebra. It is shown that subspace methods calculate a good state

space model without any priori knowledge of the system. This paper consists of the following sections:

First of all, we briefly review the main concepts and algorithms of linear subspace system identification (section 2). Different methods found in literature are presented and put into a unifying framework. Further we comment on the comparison between prediction error methods and subspace identification methods by applying them to a glass tube manufacturing process.

2 An overview of the theory [6]

Models and/or systems can be roughly divided into classes, such as linear and nonlinear, time-invariant or time-varying, discrete-time or continuous-time, with lumped or with distributed parameters, etc. While at first sight, the class of linear time-invariant models with lumped parameters seems to be rather restricted, it turns out in practice that many real-life input-output behaviors of practical, industrial processes can be approximated very well by such models. Linear subspace identification methods are connected with systems and models of the form:

$$x_{k+1} = Ax_k + Bu_k + w_k \quad (1)$$

$$y_k = Cx_k + Du_k + v_k \quad (2)$$

with

$$E \left[\begin{pmatrix} w_p \\ v_p \end{pmatrix} \begin{pmatrix} w_q^t & v_q^t \end{pmatrix} \right] = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \delta_{pq} \geq 0 \quad (3)$$

The vectors $u_k \in \mathbb{R}^{m \times 1}$ and $y_k \in \mathbb{R}^{l \times 1}$ are the measurements at time instant k of respectively the m inputs and l outputs of the process. The vector x_k is the state vector of the process at discrete time instant k , $v_k \in \mathbb{R}^{l \times 1}$ and $w_k \in \mathbb{R}^{n \times 1}$ are unobserved vector signals, called the measurement noise and the process noise, respectively. It is assumed that they are zero mean, stationary white noise vector sequences and uncorrelated with the inputs u_k . $A \in \mathbb{R}^{n \times n}$ is the system matrix, $B \in \mathbb{R}^{n \times m}$ is the input matrix, $C \in \mathbb{R}^{l \times n}$ is the output matrix while $D \in \mathbb{R}^{l \times m}$ is the direct feed-through matrix. The matrices $Q \in \mathbb{R}^{n \times n}$, $S \in \mathbb{R}^{n \times l}$ and $R \in \mathbb{R}^{l \times l}$ are the covariance matrices of the noise sequences w_k and v_k . In subspace identification, it is typically assumed that the number of available data points goes to infinity,

Iranian Journal of Electrical & Electronic Engineering, 2005.

Paper first received 27th May 2002 and in revised form 3th February 2004.

J. Poshtan and H. Mojallali are with the Department of Electrical Engineering, Iran University of Science and Technology, Narmak, Tehran 16844, Iran.

and that the data is ergodic. We are now ready to state the main problem treated:

Given a number of measurements of the input u_k and the output y_k generated by the unknown system (1)-(3);

Determine the order n of the unknown system, the system matrices A, B, C, D up to within a similarity transformation, and the matrices Q, S, R .

In this section, we will first describe the general concepts in subspace identification. Further, the two basic steps all subspace methods consist of are presented. Finally, the different algorithms existing in the literature are analyzed in a unifying framework.

Subspace identification algorithms always consist of two steps. The first step makes a projection of certain subspaces generated from the data, to find an estimate of the extended observability matrix and/or an estimate of the states of the unknown system. The second step then retrieves the system matrices from either this extended observability matrix or the estimated states. We will come back to this in section 2.2.2, where we describe different subspace identification methods and fit them into a unifying framework.

2.1 The subspace structure of linear systems

The following input-output matrix equation[4], played a very important role in the development of subspace identification:

$$Y_f = \Gamma_i X_i + H_i^d U_f + H_i^s M_f + N_f \quad (4)$$

The different terms in this equation are now defined: The extended observability matrix Γ_i is defined as:

$$\Gamma_i = \begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{i-1} \end{pmatrix} \quad (5)$$

The deterministic lower block triangular Teoplitz matrix H_i^d is defined as:

$$H_i^d = \begin{pmatrix} D & 0 & 0 & \dots & 0 \\ CB & D & 0 & \dots & 0 \\ CAB & CB & D & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ CA^{i-2}B & CA^{i-3}B & CA^{i-4}B & \dots & D \end{pmatrix} \quad (6)$$

The stochastic lower block triangular Teoplitz matrix H_i^s is defined as:

$$H_i^s = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ C & 0 & 0 & \dots & 0 \\ CA & C & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ CA^{i-2} & CA^{i-3} & CA^{i-4} & \dots & 0 \end{pmatrix} \quad (7)$$

The input and output block Hankel matrices are defined as:

$$U_{0|i-1} = \begin{pmatrix} u_0 & u_1 & \dots & u_{j-1} \\ u_1 & u_2 & \dots & u_j \\ \dots & \dots & \dots & \dots \\ u_{i-1} & u_i & \dots & u_{i+j-2} \end{pmatrix} \quad (8)$$

$$Y_{0|i-1} = \begin{pmatrix} y_0 & y_1 & \dots & y_{j-1} \\ y_1 & y_2 & \dots & y_j \\ \dots & \dots & \dots & \dots \\ y_{i-1} & y_i & \dots & y_{i+j-2} \end{pmatrix} \quad (9)$$

where we assume that $j \rightarrow \infty$ throughout the paper. For convenience and short hand notation, we call:

$$U_p = U_{0|i-1}, \quad U_f = U_{i|2i-1} \\ Y_p = Y_{0|i-1}, \quad Y_f = Y_{i|2i-1}$$

where the subscript p and f denote respectively the past and the future. The matrix containing the inputs U_p and outputs Y_p will be called W_p :

$$W_p = \begin{pmatrix} Y_p \\ U_p \end{pmatrix}$$

The block Hankel matrix formed with the process noise w_k and the measurement noise v_k are defined respectively as $M_{0|i-1}$ and $N_{0|i-1}$ in the same way. Once again, we define for short hand notation:

$$M_p = M_{0|i-1}, \quad M_f = M_{i|2i-1} \\ N_p = N_{0|i-1}, \quad N_f = N_{i|2i-1}$$

We finally denote the state sequence X_i as:

$$X_i = (x_i \ x_{i+1} \ x_{i+2} \ \dots \ x_{i+j-1}) \quad (10)$$

Definition(Orthogonal projection)

The orthogonal projection of the row space of A into the row space of B is denoted by A/B and defined as:

$$A/B = AB^+B$$

A/B^+ is the projection of the row space of A into B^\perp , the orthogonal complement of the row space of B , for which we have

$$A/B^\perp = A - A/B$$

2.2 The two basic steps in subspace identification

In this section we will explore the two main steps that all subspace algorithms consist of (see Figure 1). The first step always performs a weighted projection of the row space of the previously defined data Hankel matrices. From this projection, the observability matrix Γ_i and/or an estimate \tilde{X}_i of the state sequence X_i can be retrieved. In the second step, the system matrices A, B, C, D and Q, S, R are determined. As shown in Figure 1, a clear distinction can be made between the algorithms that use the extended observability matrix Γ_i to obtain the state space matrices, and those using the estimated state sequence \tilde{X}_i .

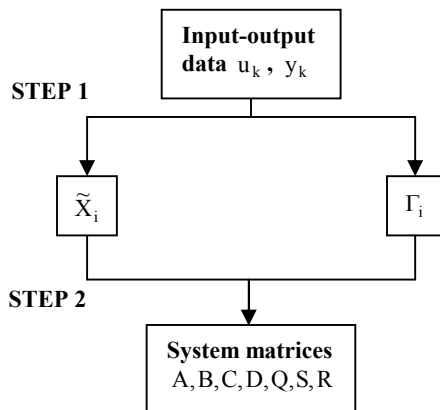


Fig. 1 Two main steps in subspace algorithms.

First step: finding the state sequence and/or the extended observability matrix

In this section, we show how an orthogonal projection with data block Hankel matrices forms one of the key elements in subspace system identification algorithms. All subspace methods start from the previously presented matrix input-output equation (4), from which it can be observed that the block Hankel matrix containing the future outputs Y_f is related in a linear way to the future input block Hankel matrix U_f and the future state sequence X_i . The basic idea of subspace identification now is to try to recover the $\Gamma_i X_i$ -term of this equation. This is a particularly interesting term since either the knowledge of Γ_i or X_i leads to the system parameters (see next section). Moreover, $\Gamma_i X_i$ is a rank-deficient term (of rank n , i.e. the system order!) which means that once $\Gamma_i X_i$ is known, Γ_i and X_i can be simply found from a SVD.

How can an estimate of $\Gamma_i X_i$ be extracted from the above equation? For this we need the previously defined notion of orthogonal projection. By projecting the row space of Y_f into the orthogonal complement U_f^\perp of the row space of U_f we find:

$$Y_f / U_f^\perp = \Gamma_i X_i / U_f^\perp + H_i^s M_f / U_f^\perp + N_f / U_f^\perp$$

Since it is assumed that the noise is uncorrelated with the inputs, we have that:

$$M_f / U_f^\perp = M_f, \quad N_f / U_f^\perp = N_f$$

Therefore:

$$Y_f / U_f^\perp = \Gamma_i X_i / U_f^\perp + H_i^s M_f + N_f$$

The following step consists in weighting this projection to the left and the right with some matrices W_1 and W_2 :

$$W_1 \cdot Y_f / U_f^\perp \cdot W_2 = W_1 \cdot \Gamma_i X_i / U_f^\perp \cdot W_2 + W_1 \cdot (H_i^s M_f + N_f) \cdot W_2$$

Of course, the inputs U_f and the weighting matrices W_1 and W_2 cannot be chosen arbitrarily but they should satisfy the following 3 conditions:

$$1. \text{rank}(W_1 \Gamma_i) = \text{rank} \Gamma_i \quad (11)$$

$$2. \text{rank} X_i = \text{rank} (X_i / U_f^\perp \cdot W_2) \quad (12)$$

$$3. W_1 \cdot (H_i^s M_f + N_f) \cdot W_2 = 0 \quad (13)$$

The first two conditions guarantee that the rank- n property of $\Gamma_i X_i$ is preserved after projection onto U_f^\perp and weighting by W_1 and W_2 . The third condition expresses that W_2 should be uncorrelated with the noise sequences w_k and v_k .

If these three conditions are satisfied, we have that:

$$O_i = W_1 \cdot Y_f / U_f^\perp \cdot W_2 = W_1 \cdot \Gamma_i X_i / U_f^\perp \cdot W_2 \quad (14)$$

with SVD:

$$O_i = (U_1 \quad U_2) \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix}$$

The following important properties can now be stated:

$$\begin{aligned} \text{rank} O_i &= n \\ W_1 \Gamma_i &= U_1 S_1^{1/2} \\ X_i / U_f^\perp \cdot W_2 &= S_1^{1/2} V_2^T \end{aligned}$$

Obviously, the singular value decomposition of the matrix O_i delivers the order n of the system. Moreover, from the left singular vectors corresponding to non-zero singular values, the extended observability matrix Γ_i can be found (up to a similarity transformation), whereas the right singular vectors contain information about the states X_i . If the weighting matrix W_2 is such that it has j columns, the matrix

$$\tilde{X}_i = X_i / U_f^\perp \cdot W_2 \quad (15)$$

can indeed be considered as an estimate of the state sequence X_i . It was shown [11] that, for a particular choice of W_2 , \tilde{X}_i is a kalman filter estimate of X_i . One might wonder about the effect of choosing the weights W_1 and W_2 in (14). Without going into details here, it suffices to say that, by choosing appropriate weighting matrices W_1 and W_2 , all subspace algorithms for LTI systems can be interpreted in the above framework, including N4SID, MOESP and CVA (see Table 1). At this point, a clear distinction can be made between the algorithms that start from Γ_i to find the system matrices A, B, C, D (MOESP) and those that use \tilde{X}_i (N4SID, CVA).

Second step: finding the state space model

We have found how an estimate \tilde{X}_i of the state sequence X_i and the extended observability matrix Γ_i can be retrieved from the weighted projection (14) of the future outputs Y_f into the orthogonal complement of future inputs U_f . In what follows, we discuss the two classes of subspace identification algorithms mentioned above. The first class uses the state estimates \tilde{X}_i (the right singular vectors) to find the state space model. Algorithms that follow this approach are N4SID and CVA. The second class of algorithms uses the extended observability matrix Γ_i (i.e. the left singular vectors) to find the model parameters.

Table 1 This table interprets different existing subspace identification algorithms in a unifying framework

Acronym	W_1	W_2
N4SID	I_{li}	$(W_p / U_f^\perp)^+ W_p$
CVA	$[(Y_f / U_f^\perp) \cdot (Y_f / U_f^\perp)^T]^{-1/2}$	$[(W_p / U_f^\perp)^+ \cdot (W_p / U_f^\perp)^+]^{-1/2}$
MOESP	I_{li}	$(W_p / U_f^\perp)^+ \cdot (W_p / U_f^\perp)$

Algorithms using an estimate \tilde{X}_i of the state sequence
 The estimated state sequence \tilde{X}_i can be interpreted as the solution of a bank of Kalman filters, working in parallel on each of the columns of the matrix W_p . Besides \tilde{X}_i , we also need the state sequence \tilde{X}_{i+1} . This sequence can be obtained from a O_{i+1} projection and weights W_1, W_2 in (14) based on $W_{0|i}, Y_{i+1|2i-1}$ and $U_{i+1|2i-1}$ (see section 2.1 for notations). This leads to the sequence O_{i+1} and the Kalman filter states \tilde{X}_{i+1} :

$$O_{i+1} = W_1 \cdot \frac{Y_{i+1|2i-1}}{U_{i+1|2i-1}^\perp} \cdot W_2 = W_1 \cdot \Gamma_{i-1} \tilde{X}_{i+1} \cdot W_2$$

System model: The state space matrices A, B, C and D can now be found by solving a simple set of over-determined equations in a least-squares sense:

$$\begin{pmatrix} \tilde{X}_{i+1} \\ Y_{ij} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \tilde{X}_i \\ U_{ij} \end{pmatrix} + \begin{pmatrix} \rho_w \\ \rho_v \end{pmatrix} \quad (16)$$

with obvious definitions for ρ_w and ρ_v as residual matrices. This reduces to

$$\min_{A, B, C, D} \left\| \begin{pmatrix} \tilde{X}_{i+1} \\ Y_{ij} \end{pmatrix} - \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \tilde{X}_i \\ U_{ij} \end{pmatrix} \right\|_F^2$$

Noise model: The noise covariance Q, S and R can be estimated from the residuals ρ_w and ρ_v as:

$$\begin{pmatrix} Q & S \\ S^T & R \end{pmatrix}_i = \frac{1}{j} \begin{bmatrix} \rho_w \\ \rho_v \end{bmatrix} \begin{bmatrix} \rho_w^T & \rho_v^T \end{bmatrix} \geq 0$$

where the index i denotes a bias induced for finite i , which disappears as $i \rightarrow \infty$. As is obvious by construction, this matrix is guaranteed to be positive semi-definite. This is an important feature since only positive definite covariances can lead to a physically realizable noise model.

Algorithms using the extended observability matrix Γ_i

contrary to the previous class of algorithms, here the system matrices are determined in two separate steps: first, A and C are determined from Γ_i while in a second step B and D are computed.

Determination of A and C

The matrices A and C can be determined from the extended observability matrix in different ways. All the methods, make use of the shift invariance property of the matrix Γ_i , which implies that (Kung,1978)(following Matlab notations):

$$A = \underline{\Gamma}_i^+ \cdot \Gamma_i, \quad C = \Gamma_i(1:1,:)$$

Determination of B and D

After the determination of A and C , the system matrices B and D have to be computed. Here we will

only sketch one possible way to do so. From the input-output equation (4), we find that:

$$\underbrace{\Gamma_i^\perp Y_f U_f^+}_{\in \mathbb{R}^{(l-n) \times m}} = \underbrace{\Gamma_i^\perp}_{\in \mathbb{R}^{(l-n) \times li}} \cdot \underbrace{H_1^d}_{\in \mathbb{R}^{li \times m}}$$

where $\Gamma_i^\perp \in \mathbb{R}^{(l-n) \times li}$ is a full row rank matrix satisfying $\Gamma_i^\perp \cdot \Gamma_i = 0$. Here once again the noise is cancelled out due to the assumption that the input u_k is correlated with the noise. Observe that with known matrices $A, C, \Gamma_i^\perp, U_f$ and Y_f , this equation is linear in B and D .

3 An industrial glass tube manufacturing process identification

3.1 Process description [4]

The outline of the process is shown in Figure 2. By indirect electric heating the glass is melted and it flows down through a ring shaped hole along a mandrill. Shaping of the tube takes place at, and just below the end of the mandrill. The glass tube is pulled down due to gravity and a drawing machine. Two measures of dimension, wall thickness and diameter of the tube, are the most important quantities to be regulated (and thus to be identified), hence they are the outputs of the process. The mandrill gas pressure and the drawing speed can affect the wall thickness and the diameter most directly and easily, so they are good candidates for inputs. Other variables such as the power supplied to melt the glass, the pressure in the melting vessel and the room temperature will be considered as disturbances. Therefore the process can be modeled as a 2-input 2-output process with disturbances (combined deterministic-stochastic model). For the purpose of identification, the process was excited with an orthogonal white PRBN sequence. The two inputs and outputs were recorded. The diameter is measured in two orthogonal directions and averaged out. The wall thickness is measured in four directions in a plane, and once again, these measurements are averaged out. The input and output signals are scaled so that the original signals can not be retrieved. The input and output signals are made zero mean, filtered with a third order low pass butterworth filter with a cutoff frequency equal to $(f_s/2)/10$ where f_s is the sampling frequency. The outputs are then detrended, as typically the thickness has a trend in it. Then the delays between the inputs and outputs are determined through a simple correlation analysis. The signals are corrected for these delays. The first 500 samples are neglected (transient effects).

3.2. Subspace Identification vs. Prediction Error Method

In this section, we will compare the identification results obtained using the three subspace algorithms with the results from the classical identification technique (PEM). We used signals of 1326 samples from this system, The first 900 samples for identification and the remaining 426 samples for validation. The results are summarized in table 2. For subspace algorithms, the

order can be determined through the inspection of a singular value plot (Figure 3). In noise-free case, the model order is equal to the number of nonzero singular values [4]. In a noisy case such as our process, however, the model order was chosen to be equal to 8 resulting the least prediction error. Among the three subspace algorithms, CVA led to the least prediction error for this case study. For using PEM classical identification, an initial model is required. Therefore, for this purpose PEM was initialized with the model from N4SID algorithm. Input signals used for identification are shown in Figure 4. In this Figure, measured and predicted output signals for CVA model are shown. It should be noted that for the prediction of the output signal, a Kalman filter was used for the following model:

$$x_{k+1} = Ax_k + Bu_k + Ke_k$$

$$y_k = Cx_k + Du_k + e_k$$

Meanwhile, the error related to a deterministic prediction is shown in table 2.

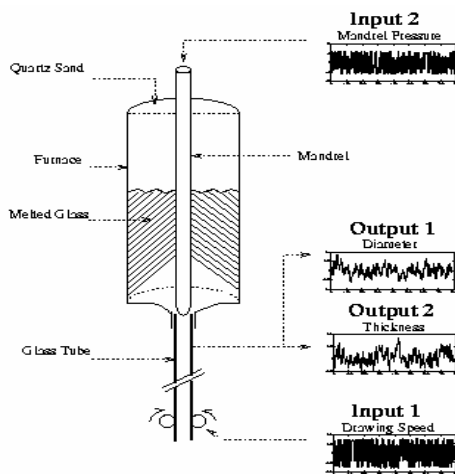


Fig. 2 The glass tube production process

Table 2 Comparison of three subspace algorithm and PEM classical identification. The second row indicates the chosen order. The third row indicates the number of block rows that is used in the past and future block Hankel matrices of subspace algorithms. The fourth row indicates number of iterations related to each algorithm. The fifth row shows the number of floating point operations. The sixth shows consuming time for these algorithms in second. The seventh row is the errors (in percent) between the measured validation outputs and the simulated outputs using only the deterministic subsystem. The eighth row shows the error between measured and kalman filter one step ahead prediction. With $[y_k^v]_i$ the i th validation output channel and $[y_k^s]_i$ the i th simulated

output channel, total error is defined as ($N_v = 426$):
$$\varepsilon = 100 \sqrt{\frac{\sum_{k=1}^{N_v} ([y_k^v]_i - [y_k^s]_i)^2}{\sum_{k=1}^{N_v} ([y_k^v]_i)^2}} \%$$

algorithm	N4SID	MOESP	CVA	PEM
order	8	8	8	8
Number of block rows	20	20	20	20
iterations	-	-	-	10
flops	61980981	81280318	63415534	314708169
time	4	5	4	25
Prediction error(deterministic)	30.8	29.1	43.7	34.4
Prediction error(Kalman filter)	14	13.5	13	14

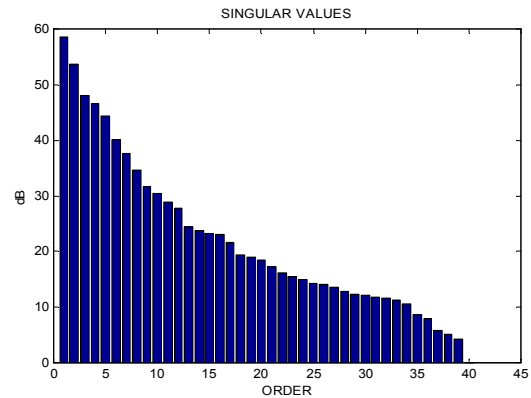


Fig. 3 Singular values as a function of the model order. The model order is chosen to be 8.

4 Conclusion

In this paper, we have given a brief overview of the linear subspace system identification methods. We made a clear distinction between methods using the states and methods starting from the observability matrix to recover the system parameters. Further, a direct comparison was made between subspace algorithms and PEM identification method on the basis of accuracy and computation time. Therefore, all methods were applied to a real-life process. The conclusion is that subspace methods calculate a state space model without a-priori fixed parametrization, and this is a fast and numerically reliable way. Subspace methods and prediction error methods are complementary in the sense that a good initial model can be quickly obtained with subspace methods while a further optimization of parameters (if possible) can be done with prediction error methods.

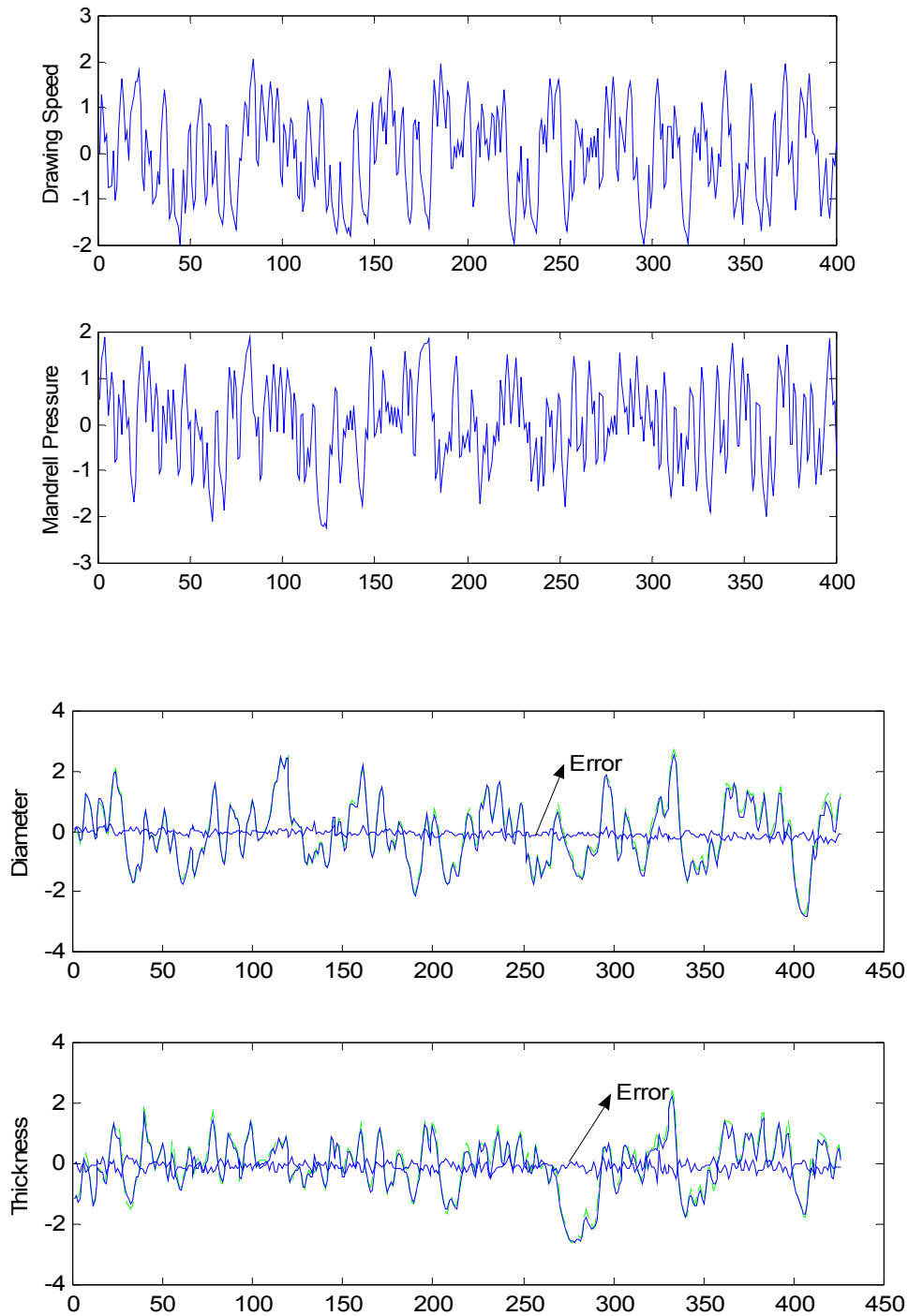


Fig. 4 Input and output signals used for identification. The inputs are drawing speed and mandrill pressure. The outputs are diameter and thickness. The dotted lines shows one step ahead predicted outputs from CVA model. Prediction errors have shown in figure.

5 References

- [1] C.T. Chou, Geometry of linear systems and identification, PHD Thesis, Trinity college, Cambridge, England (March 1994).
- [2] C.T. Chou and J.M. Maciejowsky , System identification using balanced parametrizations, IEEE Trans. On Automatic Control, 42 (7) 956-974 (1997).
- [3] P.T. Kabamba, Balanced forms: Canonicity and parametrization, IEEE Trans. On Automatic Control, 30(11) 1106-1109 (1985).
- [4] B. DeMoor, P. DeGersem, B. Deschutter, W. Favoreel, DAISY: A database for identification of systems, Journal A, Special Issue on CACSD (Computer Aided Control Systems Design), 38 (3) 4-5 (1997).

- [5] B. DeMoor, Mathematical concepts and techniques for modeling of static and dynamic systems, PHD Thesis, Department of Electrical Engineering, Katholieke Universiteit Leuven, Belgium(1988).
- [6] W. Favoreel, B. DeMoor and P. Vanoverschee, Subspace state space system identification for industrial processes, Journal of Process Control, 10 140-155 (2000).
- [7] W.E. Larimore , Canonical variate analysis in identification, filtering and adaptive control, in proc. 29th Conference on Decision and Control, Hawaii, 596-604 (1983).
- [8] L. Ljung , System Identification-Theory for the User, Prentice Hall , Englewood cliffs, N.J (1999).
- [9] T. Mckelvey, Identification of state space models from time and frequency data, PHD Thesis, No. 380, Departmet of Electrical Engineering, Linkoping, Sweden (1995).
- [10] B.C. Moore, Principal component analysis in linear systems: Controllability, observability and model reduction, IEEE Trans. On Automatic Control, 26(1) 17-32 (1981).
- [11] P. Vanoverschee and B. DeMoor, N4SID: subspace algorithms for the identification of combined deterministic and stochastic systems, Automatica, 30(1) 75-93 (1994).
- [12] M. Verhaegen and P. Dewilde , Subspace identification, part I: the output-error state space model identification class of algorithms, Int. J. Control , 56 1187-1210, (1992).