# System Identification and Simple Process Models

Lennart Ljung,

Division of Automatic Control
Department of Electrical Engineering
Linköpings universitet, SE-581 83 Linköping, Sweden
WWW: http://www.control.isy.liu.se
E-mail: Ljung@isy.liu.se, @isy.liu.se

3rd December 2003

AUTOMATIC CONTROL
COMMUNICATION SYSTEMS
LINKÖPING

Report no.: LiTH-ISY-R-2464
Submitted to International Symposium on Advanced Control of Industrial Processes, Kumamoto, Japan, June 2002

# SYSTEM IDENTIFICATION AND SIMPLE PROCESS MODELS

**Lennart Ljung**

*Division of Automatic Control, Linköping University, SE-58183 Linköping, Sweden. E-mail:* `ljung@isy.liu.se`

Abstract: System Identification is the art and science of building mathematical models of dynamical processes, from observed input and output signals. The presentation will give a state-of-the art overview of approaches, methods and models for this. Particular attention will be paid to the connection to techniques of building simple process models (such as gain + time constant + delay).

Keywords: Parameter estimation; Process Models; System Identification;

## 1. INTRODUCTION: PROCESS MODELS AND PROCESS KNOWLEDGE

System Identification is about building models of dynamical systems and processes. It is matter of matching possible prior knowledge about the process to information contained in observed input-output data to construct a process description that is capable of capturing and reproducing the basic behaviors of the process.

Depending on the intended use of the model one may demand quite accurate models, perhaps in the form of high order linear or nonlinear state-space models. In other cases it may be sufficient to capture basic dynamic behavior like dominating time-constant, static gain and possible dead-time.

Prior knowledge about the process could come in various forms, and depending on how it blends with the empirical information in observations we talk about models in different shades of gray:

- White-box models: This is the case when a model is perfectly known; it has been possible to construct it entirely from prior knowledge and physical insight.
- Grey-box models: This is the case when some physical insight is available, but several parameters remain to be determined from observed data. It is useful to consider two sub-cases:
  - Physical Modeling: A model structure can be built on physical grounds, which has a certain number of parameters to be estimated from data. This could, e.g., be a state space model of given order and structure.
  - Semi-physical modeling: Physical insight is used to suggest certain nonlinear combinations of measured data signal. These new signals are then subjected to model structures of black box character.
- Black-box models: No physical insight is available or used, but the chosen model structure belongs to families that are known to have good flexibility and have been "successful in the past".

A typical linear black-box model could be the ARX model:

$$\begin{aligned} y(t) + a_1 y(t-1) + \ldots a_n y(t-n) = \\ b_1 u(t-1) + b_2 u(t-2) + \ldots \\ + b_n u(t-n) + e(t) \end{aligned} \tag{1}$$

where $y(t)$ denotes the output at time $t$, $u(t)$ is the input (control signal) at time $t$ and $e(t)$ is a sequence of independent random variables, *unpredictable (white) noise*. The *order n* is another variable to choose, along with the parameters $a_i, b_i$.

Another linear black box model, which has a slight shade of gray is the simple process model

$$G(s) = \frac{K}{1 + sT_{p1}} e^{-sT_d} \tag{2}$$

This is a very common model in process control for PID tuning, etc. Its use relies upon some prior knowledge: There are no essential under-damped modes in the process.

Common black-box non-linear models are ANN (artificial Neural Net) models, where past inputs and outputs are subjected to a network on nonlinear transformations to construct the future output (in contrast to (1) which uses a linear such transformation.)

A simple example of a gray, semi-physical model could be

$$y(t) = ay(t-1) + bu(t-1)^2 \tag{3}$$

where $y$ could be the temperature of a heated liquid and $u$ could be the current in a heating device (so that $u^2$ is proportional to the heating power).

In this contribution we shall consider how grey and black box models of various kinds can be identified from measured data. In particular we will study simple process models and how their estimation fits into a general framework.

In Section 2 we shall first review some frequently used process models, while in Section 3 we consider general linear time-invariant models. Section 4 deals with non-linear models, primarily of Black-box character. The basic principles for estimating parameteric models are dealt with in Section 5, while essential issues of model quality and model seclection are described in Section 6. We return to simple process models in Section 7 and illustrate several of the issues discussed in the paper by numerical examples in Section 8.

The general prespective of this paper follows the book (Ljung, 1999). See also (Söderström and Stoica, 1989) and (Ljung and Glad, 1994). More details about process models in this persepctive are given in (Ljung, 2002*a*).

## 2. IDENTIFICATION FOR CONTROL: SIMPLE PROCESS MODELS

Perhaps the most commonly used process model is

$$G(s) = \frac{K}{1 + sT_{p1}} e^{-sT_d} \tag{4}$$

Among variants of this model, we can have a model without delay ($T_d = 0$):

$$G(s) = \frac{K}{1 + sT_{p1}} \tag{5}$$

and/or introduce an enforced integration (self-regulating process)

$$G(s) = \frac{K}{s(1 + sT_{p1})} e^{-sT_d} \tag{6}$$

Moreover, on can postulate two real poles with or without a zero

$$G(s) = \frac{K(1 + sT_z)}{(1 + sT_{p1})(1 + sT_{p2})} e^{-sT_d} \tag{7}$$

A further possibility is to allow resonant poles ("under-damped models"):

$$G(s) = \frac{K(1 + sT_z)}{1 + 2\zeta sT_r + (sT_r)^2} \tag{8}$$

Clearly a variety of models can be defined based on these components.

Several papers and books discuss how to estimate models like (4) from transient response data (e.g. (Åström and Hägglund, 1995), (Rake, 1980), (Ziegler *et al.*, 1943)). Most of the classical methods are graphical or semi-graphical, like finding the steepest tangent to the step response and calculate its intersection with the time axis, etc, or computing areas below the response curve and so on. See e.g. (Åström and Hägglund, 1995) for a recent overview of such approaches.

The estimation of process models of the kind described in this section is supported in the most recent version of MATLAB's System Identification Toolbox, using more standard parameter estimation schemes, (Ljung, 2002*b*).

## 3. A GENERAL MODEL DESCRIPTION I: LINEAR MODELS

### 3.1 *The Generic LTI Model*

The vast amount of possible models may cause confusion about what the basic principles are, and there is need for some uniform notions. One such notion is the arbitrarily parameterized, discrete time *linear time invariant* (LTI) model:

$$y(t) = G(q,\theta)u(t) + H(q,\theta)e(t) \tag{9}$$

where $G$ and $H$ are transfer functions in shift operator $q$ and $e$ is white noise. The parameterization of the transfer function could be chosen in many different ways. A common one is to let $\theta$ be the coefficients of numerator and denominator polynomials in $q$.

### 3.2 *Some Specific Examples*

A typical example is the ARX-model (1), which corresponds to

$$\theta = \left[a_1 \ \ldots \ a_n, b_1, \ldots, b_n\right]^T \tag{10a}$$

$$G(q,\theta) = \frac{b_1 q^{n-1} + \ldots + b_n}{q^n + a_1 q^{n-1} + \ldots + a_n} \tag{10b}$$

$$H(q,\theta) = \frac{q^n}{q^n + a_1 q^{n-1} + \ldots + a_n} \tag{10c}$$

By introducing a numerator in $H$ we would obtain an ARMAX model, while allowing a separate denominator in $H$ (i.e. different from that in $G$, gives rise to the BOX-JENKINS model. Generally, models that do not have a special description of the noise term, that is models with $H(q,\theta) \equiv 1$, are called *Output Error Models*.

A general discrete time state-space model takes the form

$$x(t+1) = A(\theta)x(t) + B(\theta)u(t) + w(t) \quad (11a)$$
$$y(t) = C(\theta)x(t) + D(\theta)u(t) + v(t) \quad (11b)$$

where $w$ and $v$ are white noises with certain covariances matrices. Computing the Kalman filter gain $K$ for this model, allows an alternative representation,

$$x(t+1) = A(\theta)x(t) + B(\theta)u(t) + K(\theta)e(t) \quad (12a)$$
$$y(t) = C(\theta)x(t) + D(\theta)u(t) + e(t) \quad (12b)$$

which is equivalent to (11a) as regards to the second order probabilistic properties. The matrices $A,B,C,D,K$ in these state-space models may depend on parameters $\theta$ in an arbitrary manner.

Clearly, (12) is equivalent to (9) with

$$G(q,\theta) = C(\theta)(qI - A(\theta))^{-1}B(\theta) \quad (13a)$$
$$H(q,\theta) = C(\theta)(qI - A(\theta))^{-1}K(\theta) + I \quad (13b)$$

Other examples are formed by any of the process models of Section 2, which can be written as

$$G(s,\theta) \quad (14)$$

where $\theta$ comprises the model parameters $K, T_{p1}, T_d$ etc. To estimate the parameters we have collected a data set

$$Z^N = \{u(1), y(1), \ldots, u(N), y(N)\} \quad (15)$$

of sampled inputs and outputs. Suppose the sampling interval is constant and equal to $T$. The model (14) is sampled with this sampling interval, according to the input intersample behavior (zero-order-hold, first-order-hold, band-limited) giving the discrete time model

$$y(t) = G_T(q,\theta)u(t) + \text{possible noise} \quad (16)$$

A similar result is obtained if we start from a general, physically parameterized state-space model in continuous time.

So linear, time invariant models of any color essentially end up in (9). The somewhat "esoteric" noise term $He$ could really primarily be seen as vehicle to come up with reasonable ways of guessing are predicting future outputs.

### 3.3 *Predictors for LTI models*

The logical one-step ahead predictor related to (9) is

$$\hat{y}(t|\theta) = H^{-1}(q,\theta)G(q,\theta)u(t) + (I - H^{-1}(q,\theta))y(t) \quad (17)$$

It is the assumed that $H$ is normalized to be *monic* (starting with the identity matrix, so that $I - H^{-1}(q,\theta)$ contains a delay). This means that (17) only involves past values of the output, so that it makes sense as a predictor.

The right hand side of (17) is a linear function of past data $Z^{t-1}$:

$$\hat{y}(t|\theta) = g(\theta, Z^{t-1}) \quad (18)$$

with $g$ linear. The step to general non-linear predictor (regression) models is then immediate: just let $g$ be non-linear.

## 4. MODEL STRUCTURES II: NON-LINEAR BLACK BOX MODELS

In this section we shall describe the basic ideas behind model structures that have the capability to cover any non-linear mapping from past data to the predicted value of $y(t)$. We defined a general model structure as a parameterized mapping in (18). We shall consequently allow quite general non-linear mappings $g$. This section will deal with some general principles for how to construct such mappings, and will cover Artificial Neural Networks as a special case. See (Sjöberg *et al.*, 1995), (Juditsky *et al.*, 1995) and (Ljung, 1999), Chapter 5 for more comprehensive surveys.

### 4.1 *A basic structure*

Now, the model structure family (18) is really too general, and it turns out to be useful to write $g$ as a concatenation of two mappings: one that takes the increasing number of past observations $Z^{t-1}$ and maps them into a finite dimensional vector $\varphi(t)$ of fixed dimension and one that takes this vector to the space of the outputs:

$$\hat{y}(t|\theta) = g(\theta, Z^{t-1}) = g(\varphi(t), \theta) \quad (19)$$

where

$$\varphi(t) = \varphi(Z^{t-1}) \quad (20)$$

Let the dimension of $\varphi$ be $d$. We shall call this vector the *regression vector* and its components will be referred to as the *regressors*. We also allow the more general case that the formation of the regressors is itself parameterized:

$$\varphi(t) = \varphi(Z^{t-1}, \eta) \quad (21)$$

which we for short write $\varphi(t,\eta)$. For simplicity, the extra argument $\eta$ will however be suppressed.

The choice of the non-linear mapping in (18) has thus been reduced to two partial problems for dynamical systems:

(1) How to choose the non-linear mapping $g(\varphi)$ from the regressor space to the output space (*i.e.*, from $R^d$ to $R^p$).
(2) How to choose the regressors $\varphi(t)$ from past inputs and outputs.

The second problem is the same for all dynamical systems, and it turns out that the most useful choices of regression vectors are to let them contain past inputs and outputs, and possibly also past predicted/simulated outputs. We now turn to the first problem.

### 4.2 *Non-Linear Mappings: Possibilities*

Now let us turn to the nonlinear mapping

$$g(\varphi,\theta) \tag{22}$$

which for any given $\theta$ maps from $R^d$ to $R^p$. For simplicity we will use $p=1$, i.e., the output is scalar-valued. At this point it does not matter how the regression vector $\varphi=(\varphi_1,\dots,\varphi_d)^T$ was constructed. It is just a vector that lives in $R^d$.

It is natural to think of the parameterized function family as function expansions:

$$g(\varphi,\theta)=\sum\alpha_k g_k(\varphi)\ . \tag{23}$$

We refer to $g_k$ as *basis functions*, since the role they play in (23) is similar to that of a functional space basis. In some particular situations, they do constitute a functional basis. Typical examples are wavelet bases (see below).

We are going to show that expansion (23) with different basis functions, plays the role of a unified framework for investigating most known nonlinear black-box model structures.

Now, the key question is: How to choose the basis functions $g_k$? The following facts are essential to understand the connections between most known nonlinear black-box model structures:

- All the $g_k$ are formed from one "mother basis function", that we generically denote by $\kappa(x)$.
- This function $\kappa(x)$ is a function of a scalar variable $x$.
- Typically $g_k$ are dilated (scaled) and translated versions of $\kappa$. For the scalar case $d=1$ we may write
$$g_k(\varphi)=g_k(\varphi,\beta_k,\gamma_k)=\kappa(\beta_k(\varphi-\gamma_k)) \tag{24}$$
We thus use $\beta_k$ to denote the dilation parameters and $\gamma_k$ to denote translation parameters.

For any of the described choices the resulting model becomes

$$g(\varphi,\theta)=\sum_{k=1}^{n}\alpha_k\kappa(\beta_k(\varphi-\gamma_k)) \tag{25}$$

the exact interpretations of the argument $\beta_k(\varphi-\gamma_k)$ when $\varphi$ is a column vector could vary: On possibility, known as the *ridge approach*, is to let $\beta$ be a row vector (and hence $\beta\gamma$ is a scalar) so that the term in question is constant in the null space of $\beta$. Another approach, the *radial approach*, is to interpret $\beta(\varphi-\gamma)$ as the distance between $\varphi$ and $\gamma$ measured in a quadratic norm determined by $\beta$ (a matrix or a vector or a scalar). A third interpretation, called the *tensor approach*, is to interpret the term as the product of $\kappa(\beta_i(\varphi_i-\gamma_i))$ where $i$ ranges over the rows of the vectors.

In any case, the expansion is entirely determined by

- the scalar valued function $\kappa(x)$ of a scalar variable $x$
- the interpretation of $\kappa(\beta(\varphi-\gamma))$ in case $\varphi$ is a vector.

The parameterization in terms of $\theta$ can be characterized by three types of parameters:

- The *coordinates* $\alpha$
- The *scale* or *dilation* parameters $\beta$
- The *location* parameters $\gamma$

The essential message of this general description is that most commonly used non-linear model structures fit into this template:

- Sigmoidal Artificial Neural Networks (ANN) with one hidden layer correspond to $\kappa$ given by (29) and the ridge approach.
- Radial Basis Neural Networks are obtained by the radial approach, and typically $\kappa$ given by (27).
- Wavenet networks are formed with $\kappa$ being a mother wavelet and typically the radial approach
- Kernal methods correspond to a fixed choice of dilation parameter $\beta$ and a given grid of location parameters.
- The nearest neighbor method is obtained by letting $\kappa$ be an indicator function and by picking $\beta$ and $\gamma$ so that exactly one observation falls into each term of (25) and letting $\alpha$ be the value of the corresponding observation.
- Fuzzy or neuro-fuzzy models correspond to letting $\kappa$ be the membership functions and $\alpha$ be "the weighted value of applicable rule conclusions", see Section 5.6 in (Ljung, 1999).

### 4.3 *Why does it work?: A Scalar Example*

The function expansion (25) is able to approximate any (reasonable) function arbitrarily well by taking $n$ sufficiently large. To get an intuitive feel for this important property, let as assume $\varphi$ is scalar and take $\kappa$ as the unit interval indicator function:

$$\kappa(x)=\begin{cases}1 & \text{for } 0\le x<1\\ 0 & \text{else}\end{cases} \tag{26}$$

Then take, for example, $\gamma_k=k$, $\beta_k=1/\Delta$ and $\alpha_k=f(k\Delta)$. Then (23), (24) gives a piecewise constant approximation of any function $f$. With the given choices the approximation will be constant over intervals of length $\Delta$. Any sufficiently smooth function can be approximated arbitrarily well by a piecewise constant function, so this shows the power of the function expansion (25).

Clearly we would have obtained a quite similar result by a smooth version of the indicator function, e.g., the Gaussian bell:

$$\kappa(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{27}$$

If instead $\kappa$ is taken to be the unit step function

$$\kappa(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases} \tag{28}$$

we just have a variant of (26), since the indicator function can be obtained as the difference of two steps. A smooth version of the step, like the *sigmoid* function

$$\kappa(x) = \sigma(x) = \frac{1}{1+e^{-x}} \tag{29}$$

will of course give quite similar results.

## 5. GENERAL PARAMETER ESTIMATION TECHNIQUES

In this and the following sections we shall deal with issues that are independent of model structure. Principles and algorithms for fitting models to data, as well as the general properties of the estimated models are all model-structure independent and equally well applicable to, say, ARMAX models, Neural Network models and simple process models.

### 5.1 *Fitting Models to Data*

We have characterized in general terms a model structure as a parameterized predictor, (18):

$$\hat{y}(t|\theta) = g(\theta, Z^{t-1}) \tag{30}$$

that depends on the unknown parameter vector and past data $Z^{t-1}$ (see (15).

*We now need a method to determine a good value of* $\theta$, based on the information in an observed, sampled data set (15).

A procedure with some degrees of freedom is the following one

(1) From observed data and the predictor $\hat{y}(t|\theta)$ form the sequence of prediction errors,

$$\varepsilon(t,\theta) = y(t) - \hat{y}(t|\theta), \quad t = 1,2,\ldots N \tag{31}$$

(2) Possibly filter the prediction errors through a linear filter $L(q)$,

$$\varepsilon_F(t,\theta) = L(q)\varepsilon(t,\theta) \tag{32}$$

so as to enhance or depress interesting or unimportant frequency bands in the signals. We shall call $L$ the *Focus Filter*.

(3) Choose a scalar valued, positive function $\ell(\cdot)$ so as to measure the "size" or "norm" of the prediction error:

$$\ell(\varepsilon_F(t,\theta)) \tag{33a}$$

(4) Minimize the sum of these norms:

$$\hat{\theta}_N = \arg\min_\theta V_N(\theta, Z^N) \tag{33b}$$

where

$$V_N(\theta, Z^N) = \frac{1}{N}\sum_{t=1}^{N} \ell(\varepsilon_F(t,\theta)) \tag{33c}$$

This procedure is natural and pragmatic – we can think of it as "curve-fitting" between $y(t)$ and $\hat{y}(t|\theta)$. It also has several statistical and information theoretic interpretations. Most importantly, if the noise source in the system is supposed to be a sequence of independent random variables $\{e(t)\}$ each having a probability density function $f_e(x)$, then (33b) becomes the Maximum Likelihood estimate (MLE) if we choose

$$L(q) = 1 \quad \text{and} \quad \ell(\varepsilon) = -\log f_e(\varepsilon) \tag{34}$$

The MLE has several nice statistical features and thus gives a strong "moral support" for using the outlined method. Another pleasing aspect is that the method is independent of the particular model parameterization used (although this will affect the actual minimization procedure). For example, the method of "back propagation" often used in connection with neural network parameterizations amounts to computing $\hat{\theta}_N$ in (33b) by a recursive gradient method.

### 5.2 *Regularization*

The parameteriztion may sometimes involve many parameters. This is often the case for linear black-box models. In those cases it is quite helpful to condsider a *regularized* version of the criterion (33c):

$$\begin{aligned} \hat{\theta}_N &= \arg\min_\theta W_N(\theta, Z^N) \\ W_N(\theta, Z^N) &= V_N(\theta, Z^N) + \delta\|\theta - \theta^\dagger\|^2 \end{aligned} \tag{35}$$

where $V_N$ is given by (33c), $\delta$ is the regulariztion parameter and $\theta^\dagger$ as a fixed parameter value (often the origin).

## 6. MODEL QUALITY

### 6.1 *Basic Asymptotic Results*

An essential question is, of course, what properties the estimate resulting from (33) will have. These will naturally depend on the properties of the data record $Z^N$ defined by (15). It is in general a difficult problem to characterize the quality of $\hat{\theta}_N$ exactly. One normally has to be content with the asymptotic properties of $\hat{\theta}_N$ as the number of data, $N$, tends to infinity.

It is an important aspect of the general identification method (33b) that the asymptotic properties of the resulting estimate can be expressed in general terms for arbitrary model parameterizations.

The first basic result is the following one:

$$\hat{\theta}_N \to \theta^* \quad \text{as} \quad N \to \infty \quad \text{where} \tag{36a}$$

$$\theta^* = \arg\min_{\theta} E\ell(\varepsilon_F(t,\theta)) \tag{36b}$$

That is, as more and more data become available, the estimate converges to that value $\theta^*$, that would minimize the expected value of the "norm" of the filtered prediction errors. This is in a sense *the best possible approximation* of the true system that is available within the model structure. The expectation $E$ in (36b) is taken with respect to all random disturbances that affect the data and it also includes averaging over the input properties. This means in particular that $\theta^*$ will make $\hat{y}(t|\theta^*)$ a good approximation of $y(t)$ with respect to those aspects of the system that are enhanced by the input signal used.

The second basic result is the following one: If $\{\varepsilon(t,\theta^*)\}$ is approximately white noise, then the covariance matrix of $\hat{\theta}_N$ is approximately given by

$$E(\hat{\theta}_N - \theta^*)(\hat{\theta}_N - \theta^*)^T \sim \frac{\lambda}{N}[E\psi(t)\psi^T(t)]^{-1} \tag{37a}$$

$$\text{where} \quad \lambda = E\varepsilon^2(t,\theta^*) \tag{37b}$$

$$\psi(t) = \frac{d}{d\theta}\hat{y}(t|\theta)|_{\theta=\theta^*} \tag{37c}$$

Think of $\psi$ as the sensitivity derivative of the predictor with respect to the parameters. Then (37) says that the covariance matrix for $\hat{\theta}_N$ is proportional to the inverse of the covariance matrix of this sensitivity derivative. This is a quite natural result.

**Note:** For all these results, the expectation operator $E$ can, under quite general conditions, be replaced by the limit of the sample mean, that is

$$E\psi(t)\psi^T(t) \leftrightarrow \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\psi(t)\psi^T(t) \tag{38}$$

□

The results (36) and (37) are general and hold for all model structures, both linear and non-linear ones, subject only to some regularity and smoothness conditions. They are also fairly natural, and will give the guidelines for all user choices involved in the process of identification. See (Ljung, 1999) for more details around this.

### 6.2 *A Characterization of the Limiting Model in a General Class of Linear Models*

Let us apply the general limit result (36) to the linear model structure (9). If we choose a quadratic criterion $\ell(\varepsilon) = \varepsilon^2$ (in the scalar output case) then this result tells us, in the time domain, that the limiting parameter estimate is the one that minimizes the filtered prediction error variance (for the input used during the experiment.) Suppose that the data actually have been generated by

$$y(t) = G_0(q)u(t) + v(t) \tag{39}$$

Let $\Phi_u(\omega)$ be the input spectrum and $\Phi_v(\omega)$ be the spectrum for the additive disturbance $v$. Then the filtered prediction error can be written

$$\varepsilon_F(t,\theta) = \frac{L(q)}{H(q,\theta)}[y(t) - G(q,\theta)u(t)] =$$

$$\frac{L(q)}{H(q,\theta)}[(G_0(q) - G(q,\theta))u(t) + v(t)] \tag{40}$$

By Parseval's relation, the prediction error variance can also be written as an integral over the spectrum of the prediction error. This spectrum, in turn, is directly obtained from (40), so the limit estimate $\theta^*$ in (36b) can also be defined as

$$\theta^* = \arg\min_{\theta}\Bigg[\int_{-\pi}^{\pi}|G_0(e^{i\omega}) - G(e^{i\omega},\theta)|^2 \times \frac{\Phi_u(\omega)|L(e^{i\omega})|^2}{|H(e^{i\omega},\theta)|^2}d\omega + \int_{-\pi}^{\pi}\Phi_v(\omega)|L(e^{i\omega})|^2/|H(e^{i\omega},\theta)|^2 d\omega\Bigg] \tag{41}$$

If the noise model $H(q,\theta) = H_*(q)$ does not depend on $\theta$ (as in the output error model $H(q,\theta) \equiv 1$) the expression (41) thus shows that the resulting model $G(e^{i\omega},\theta^*)$ will give that frequency function in the model set that is closest to the true one, in a quadratic frequency norm with weighting function

$$Q(\omega) = \Phi_u(\omega)|L(e^{i\omega})|^2/|H_*(e^{i\omega})|^2 \tag{42}$$

This shows clearly that the fit can be affected by the choice of prefilter $L$, the input spectrum $\Phi_u$ and the noise model $H_*$.

### 6.3 *Measures of Model Fit*

Some quite general expressions for the expected model fit, that are independent of the model structure, can also be developed.

Let us measure the (average) fit between any model (30) and the true system as

$$\bar{V}(\theta) = E|y(t) - \hat{y}(t|\theta)|^2 \tag{43}$$

Here expectation E is over the data properties (i.e. expectation over "$Z^\infty$" with the notation (15)). Recall that expectation also can be interpreted as sample means as in (38).

Before we continue, let us note the very important aspect that the fit $\bar{V}$ will depend, not only on the model and the true system, *but also on data properties*, like input spectra, possible feedback, etc. We shall say that the fit depends on the *experimental conditions*.

The estimated model parameter $\hat{\theta}_N$ is a random variable, because it is constructed from observed data, that can be described as random variables. To evaluate the model fit, we then take the expectation of $\bar{V}(\hat{\theta}_N)$ with respect to the estimation data. That gives our measure

$$F_N = E\bar{V}(\hat{\theta}_N) \qquad (44)$$

In general, the measure $F_N$ depends on a number of things:

- The model structure used.
- The number of data points $N$.
- The data properties for which the fit $\bar{V}$ is defined.
- The properties of the data used to estimate $\hat{\theta}_N$.

The rather remarkable fact is that if the two last data properties coincide, then, asymptotically in $N$, (see, e.g., (Ljung, 1999), Chapter 16)

$$F_N \approx \bar{V}_N(\theta^*)(1+\frac{dim\theta}{N}) \qquad (45)$$

Here $\theta^*$ is the value that minimizes the expected criterion (36b). The notation $dim\theta$ means the number of estimated parameters. The result also assumes that the criterion function $\ell(\varepsilon) = \|\varepsilon\|^2$, and that the model structure is successful in the sense that $\varepsilon_F(t)$ is approximately white noise.

Despite the reservations about the formal validity of (45), it carries a most important conceptual message: If a model is evaluated on a data set with the same properties as the estimation data, then *the fit will not depend on the data properties*, and it will depend on the model structure *only in terms of the number of parameters used and of the best fit offered within the structure.*

The expression can be rewritten as follows. Let $\hat{y}_0(t|t-1)$ denote the “true” one step ahead prediction of $y(t)$, and let

$$W(\theta) = E|\hat{y}_0(t|t-1) - \hat{y}(t|\theta)|^2 \qquad (46)$$

and let

$$\lambda = E|y(t) - \hat{y}_0(t|t-1)|^2 \qquad (47)$$

Then $\lambda$ is the *innovations* variance, i.e., that part of $y(t)$ that cannot be predicted from the past. Moreover $W(\theta^*)$ is the *bias error*, i.e. the discrepancy between the true predictor and the best one available in the model structure. Under the same assumptions as above, (45) can be rewritten as

$$F_N \approx \lambda + W(\theta^*) + \lambda\frac{dim\theta}{N} \qquad (48)$$

The three terms constituting the model error then have the following interpretations

- $\lambda$ is the unavoidable error, stemming from the fact that the output cannot be exactly predicted, even with perfect system knowledge.
- $W(\theta^*)$ is the bias error. It depends on the model structure, and on the experimental conditions. It will typically decrease as $dim\theta$ increases.
- The last term is the *variance error*. It is proportional to the number of estimated parameters and inversely proportional to the number of data points. It does not depend on the particular model structure or the experimental conditions.

## 7. PARTICULAR ASPECTS FOR PROCESS MODELS

A leading idea with simple process models like (4) is –precisely– that they are simple. They cannot therefore be expected to capture the true system behavior in all its aspects. Often the model is used just to tune a PI or PID regulator, and then only certain model aspects have to be captured. It is thus important to analyze in what way the models are approximating the true system.

If the true system is supposed to be linear, the model fit is explicitly described by the general result (41):

Suppose the true frequency function for the sampled system is $G_0^{(T)}(e^{i\omega})$ (or, more generally, the frequency function of the linear time invariant second order equivalent of the true system, see (Ljung, 2001)). Then for $H = 1$ we have from (41):

$$\hat{\theta}_N \rightarrow \arg\min_{\theta} \int_{-\pi}^{\pi} |G_T(e^{i\omega},\theta) - G_0^{(T)}(e^{i\omega})|^2 \times \Phi_u(\omega)|L(e^{i\omega})|^2 d\omega \qquad (49)$$

Here $L$ is the "focus filter" in (32), and $\Phi_u$ is the input spectrum. The expression describes exactly in what way the simple process model like (4) approximates the true system. We also see how the focus filter may steer the fit to important frequency ranges.

In a sense, (49) also explains the success of simple process models. Even a three-parameter model like (4) has substantial "local flexibility". The delay term may pick up the true system’s phase, even if there is no dead-time in the system. For successful control design it is often sufficient to have a rough picture of the Nyquist curve in a limited, but important, frequency region, and (49) illustrates how this can be achieved.

### 7.1 *Frequency Domain Aspects on the Model Fit*

Note that the fit in (49) is affected both by the focus filter $L$ and by the input spectrum. This means that the input used will affect the resulting model. Even if the experiment is just a step response, the length of the measurement period will affect to result. That will be illustrated in the next section. However, whatever input characteristic is at hand, the focus filter $L$ can always be used to make sure that the model matches the true system at suitable frequency ranges.

### 7.2 *Time Delays That are not an integer number of the sampling interval*

It is important to view the time delay $T_d$ in (4)–(7) is a freely adjustable parameter, not necessarily corresponding to a physical dead-time. That will allow useful flexibility in fitting the phase of the model to that of the true system in important frequency ranges. Now, allowing $T_d$ not to be an integer of the sampling interval in the step from (14) tp (16) means a bit more complicated sampling algorithm, but is still a well defined process, described in most textbooks.

### 7.3 *Algorithmic Aspects*

Dealing with the particular model structures of Section 2 in the minimization (32) requires some attention to obtain numerical efficiency. The most important one is to secure good initial parameter values (in particular of the delay) in order to come into the domain of attraction of the global minimum of the criterion.

Such and other algorithmic aspects are discussed in (Ljung, 2002*a*).

C

## 8. EXAMPLES

### 8.1 *An 8th order system: Experiment length*

In (Åström and Hägglund, 1995) the following model is considered:

$$G(s) = \frac{1}{(1+s)^8} \tag{50}$$

Step response data were simulated for this system with sampling interval 0.3 seconds. Several different lengths of the measurement record were tested: 9 seconds of input = 0 followed by 9, 12 and 45 seconds with the input = 1. The process model (4) was fitted to these data using (32), giving the models

$$G(s) = \frac{K}{1+T_{p1}s}e^{-T_d s} \tag{51}$$

with

| Response | $K_p$ | $T_{p1}$ | $T_d$ | curve |
|---|---|---|---|---|
| short | $2.9 \cdot 10^5$ | $21 \cdot 10^5$ | 3.99 | A |
| mid | 2.30 | 14.8 | 4.07 | B |
| long | 1.006 | 3.27 | 5.10 | C |

As a comparison, the long response data were also fitted to a first order model without delay. That gave the model

$$G(s) = \frac{1.04}{1+8.8s} \tag{52}$$

The step responses and the Nyquist plots of these models are compared with the true system in Figures 8.1 and 8.1.

From this we note a few things:

- For the longer step response, the simple process model with delay gives a good approximation, despite the substantial difference in model complexity,

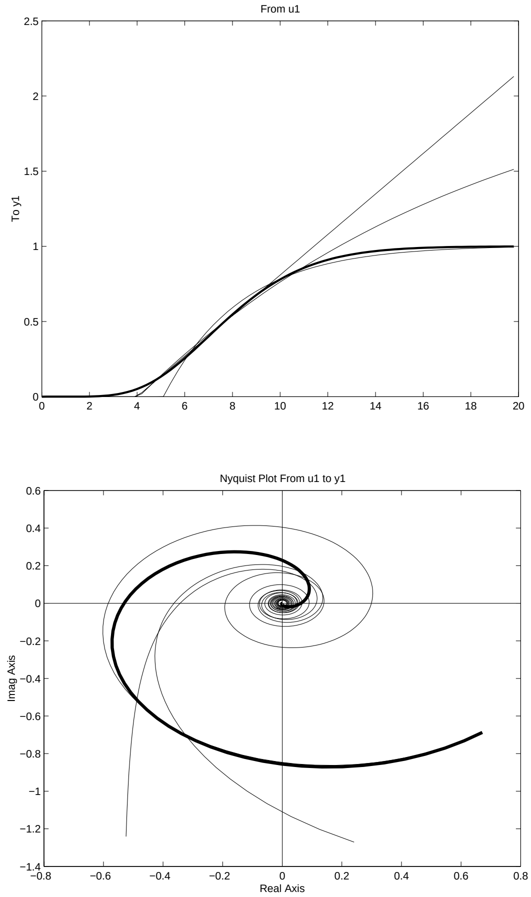


Fig. 1. Step responses and Nyquist curves for the true system (thick line), for the simple process models (51) (The step response thin lines show from top to bottom the responses A, B, and C, respectively, The Nyquist thin line curves are the same models in order with starting point from left to right).

- Although there is no dead time in the true system, the presence of the delay term in the simple model is most essential to obtain a good fit. Allowing this delay to be a continuous parameter and not just an integer times the sampling interval is also essential.
- The resulting model depends to a large extent on the length of the step response. The reason is given by equation (49): The different experiment lengths give different input spectra.

### 8.2 *An 8th order system: Effect of prefiltering*

One might wish to have a better agreement around the phase crossover frequency, and for that reason we choose a focus filter, and applied it to the longer experiment data:

```
[a,b] = butter(5,[0.25 0.55]*0.3/pi)
mf = procmod(data,'p1d','focus',[a,b])
```

The resulting model is

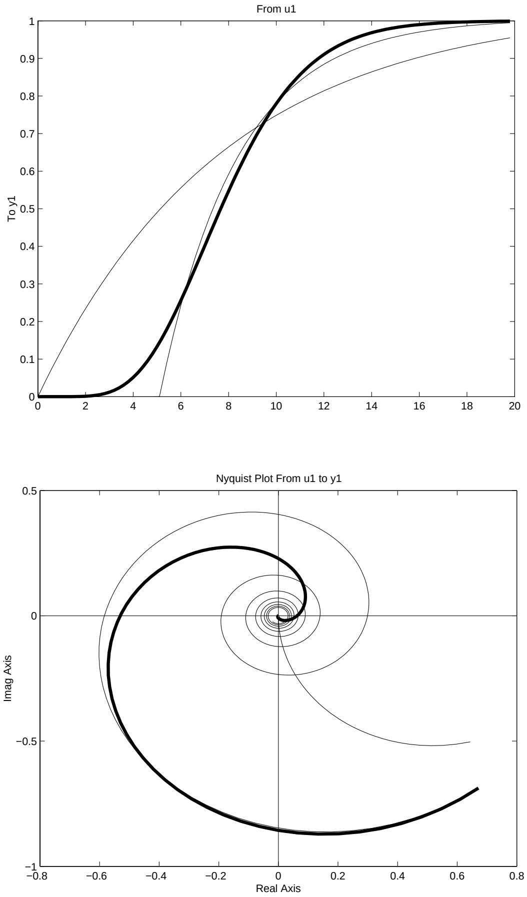


Fig. 2. Step responses and Nyquist curves for the true system (thick line), for the simple process model (52) and the curve C from the table. (In the step response plot (52) is the curve with no delay, and in the Nyquist plot it is the curve that lives entirely in the fourth quadrant.)

$$G(s) = \frac{1.0399}{1+3.9609s}e^{-5.0202s} \tag{53}$$

and its time and frequency responses are also shown in Figure 8.2. The fit around the phase cross-over between the model (53) and the true system is now very close.

## 8.3 *An 8th order system: Intersample Behavior*

For the same system (50), also a simulated experiment with a saw-tooth input was performed. This gave the data of Figure 4. The sampling interval was here 2 seconds, which is pretty slow by not unreasonable, since the solution time of the true step response is about 14 seconds. The input was a continuous saw-tooth signal, so that it was piece-wise linear between the sampling instants, “first-order-hold”. A model of the kind (4) was fitted to the data, giving the result

$$G(s) = \frac{1.0}{1+2.87s}e^{-4.07s} \tag{54}$$

under the assumption that the input was piece-wise constant (“zero-order-hold”). This means that conven-

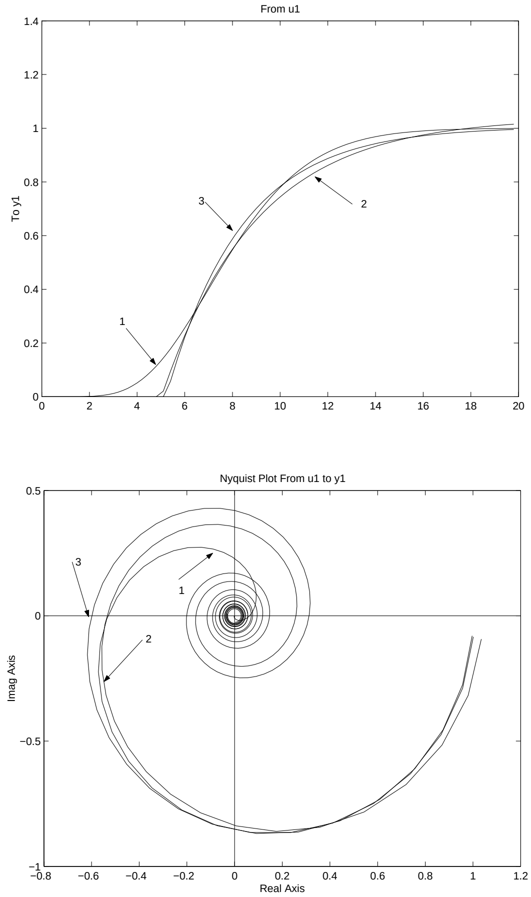


Fig. 3. Step responses and Nyquist curves for the true system (curve 1), for the simple process model (53) (curve 2) and the model C in the table (curve 3).

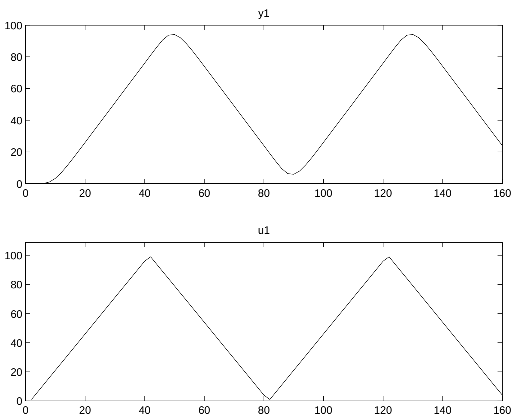


Fig. 4. Input-Output data for saw-tooth experiment .

tional sampling was applied when computing (16) from (14). However, telling the identification process the true nature of the input intersample behavior:

```
dat = iddata(y,u,2,'Intersample','foh');
m = procmod(dat,'p1d');
```

gave the following model

$$G(s) = \frac{1.0}{1+2.93s}e^{-5.12s} \tag{55}$$

The step responses and Nyquist plots of these models are given in Figure 8.3. It is clear that the model that is based on correct input intersample information is superior.

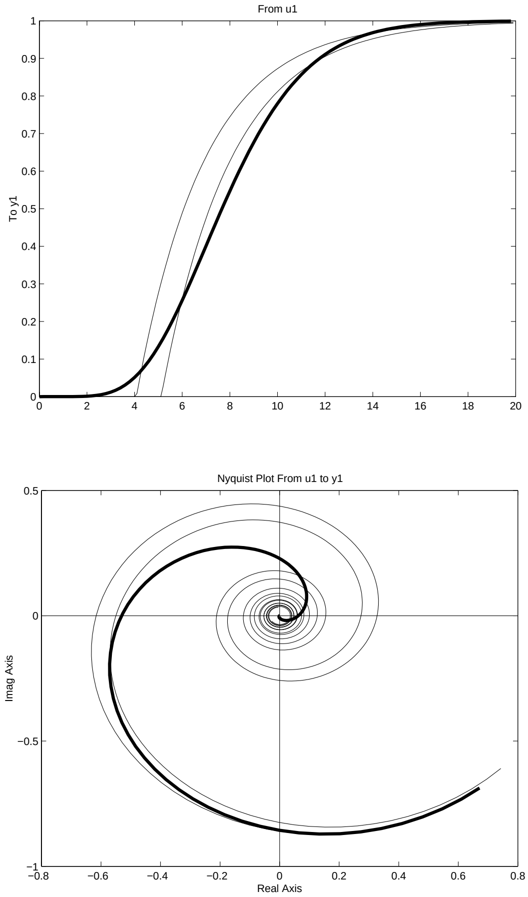


Fig. 5. Step responses and Nyquist curves for the true system (thick line), for the process model (54) (step starts at 4, and outermost Nyquist curve) and the model (55) (step starts at 5 and Nyquist curve closer to the true one.).

### 8.4 *Buffer Vessel Dynamics*

This example concerns a typical problem in process industry. It is taken from the pulp factory in Skutskär, Sweden. Wood chips are cooked in the digester and the resulting pulp travels through several vessels where it is washed, bleached etc. The data that we study are the same as in the paper (Andersson and Pucar, 1995).

The pulp spends about 48 hours total in the process, and knowing the residence time in the different vessels is important in order to associate various portions of the pulp with the different chemical actions that have taken place in the vessel at different times. Figure 8.4 shows data from one buffer vessel. We denote the measurements as follows:

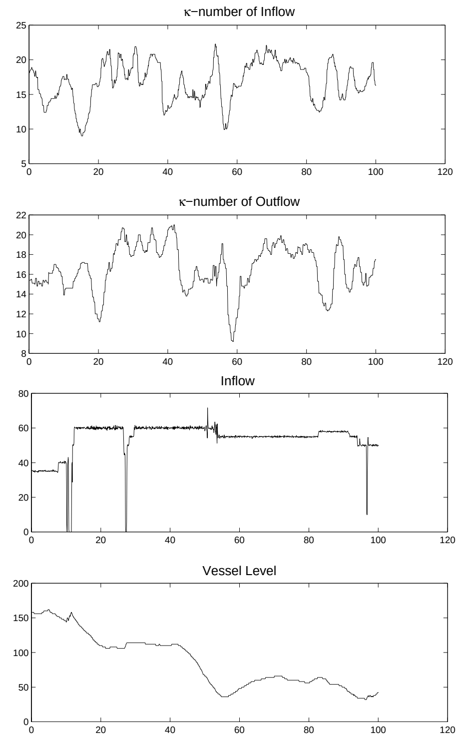


Fig. 6. From the pulp factory at Skutskär, Sweden. The plots show The $\kappa$-number of the pulp flowing into a buffer vessel. The $\kappa$-number of the pulp coming out from the buffer vessel. Flow out from the buffer vessel. Level in the buffer vessel. The sampling interval is 4 minutes, and the time scale shown in hours.

$y(t)$ : The $\kappa$-number of the pulp flowing out
$u(t)$ : The $\kappa$-number of the pulp flowing in
$f(t)$ : The output flow
$h(t)$ : The level of the vessel

The problem is to determine the residence time in the buffer vessel. (The $\kappa$-number is a quality property that in this context can be seen as a marker allowing us to trace the pulp.)

To estimate the residence time of the vessel it is natural to estimate the dynamics from $u$ to $y$. That should show how long time it takes for a change in the input to have an effect on the output.

*SISO Data.* We can visually inspect the input-output data and see that the delay seems to be at least an hour or two. The sampling rate may therefore be too fast and we resample the data (decimate it) by a factor of 3, thus giving a sampling interval of 12 minutes. We proceed by removing the means from the $\kappa$-number signals, split into estimation and validation data and estimate simple process models. This gives the model

$$G(s) = \frac{0.818}{1+676s} e^{-480s} \qquad (56)$$

A comparison between the output of this model and the measured output is shown in Figure 8.4. The fit cannot be considered as good, since the model output

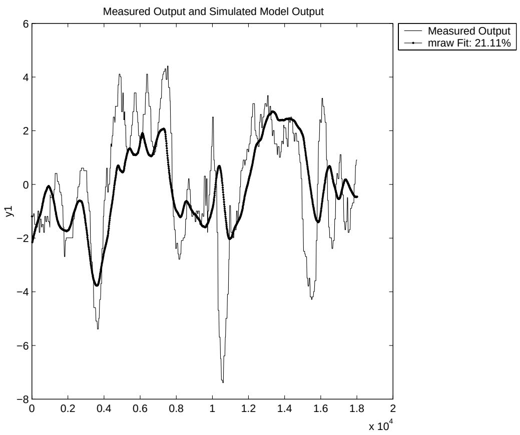


Fig. 7. Comparison between the model output (thick line) and the measured output (thin line) for the buffer vessel data. The model is estimated using the first half of the data. The data used here are the detrended, decimated data, without any further transformations.

is not only very smoothed out, but also out of phase with dips and peaks, so the residence time is not captured very well.

*Three Inputs One Output.* According to the normal routines in practical system identification, we should then contemplate if there are more input signals that may affect the process. Yes, clearly the flow and level of the vessel should have something to do with the dynamics, so we include these two inputs. We then estimate a process model of the kind (4), but with three inputs. That means that we have a three input, one-output model made up from three individual transfer functions like (4). The corresponding comparison is shown in Figure 8.4. This does not look good at all.

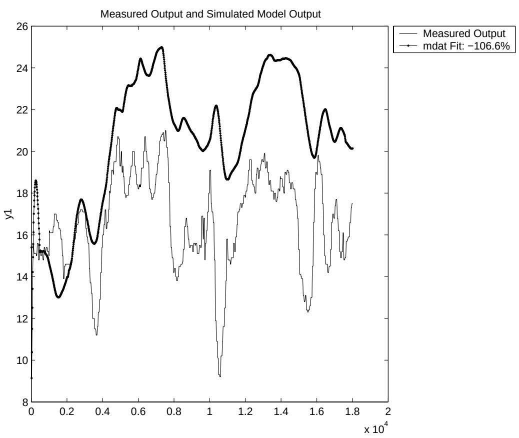


Fig. 8. Comparison between the model output (thick line) and the measured output (thin line) for the buffer vessel data. The model is estimated using the first half of the data. The data used here are the detrended, decimated data, using input $\kappa$-number, flow, and level as inputs, without any further transformations.

*Non-linear Black Boxes.* One can also try various non-linear black box models of the type described in Section 4, for these data with three inputs and one output, but it seems to be difficult to get any reasonable result that way.

*Semi-physical Modeling.* At this point we need to think more and apply some simple semi-physical modeling to move to a grey-box model. Some reflection shows that this process indeed must be non-linear (or time-varying): the flow and the vessel level definitely affect the dynamics. For example, if the flow was a plug flow (no mixing in the vessel) the vessel would have a dynamics of a pure delay equal to vessel volume divided by flow. This ratio, which has dimension time, is really the natural time scale of the process, in the sense that the delay would be constant in this time scale for a plug flow, even if vessel flow and level vary.

Let us thus resample the date accordingly, i.e. so that a new sample is taken (by interpolation from the original measurement) equidistantly in terms of integrated flow divided by volume. In MATLAB terms this will be

```
z = [y,u]; pf = f./h;
t =1:length(z)
newt = interp1(cumsum(pf+0.0001),...
t,[pf(1):sum(pf)]’ );
newz = interp1(t,z, newt);
y1=newz(:,1); u1=newz(:,2)
```

(The small added number to `pf` is in order to overcome those time points where the flow is zero.) The resampled data are shown in Figure 8.4 We now apply

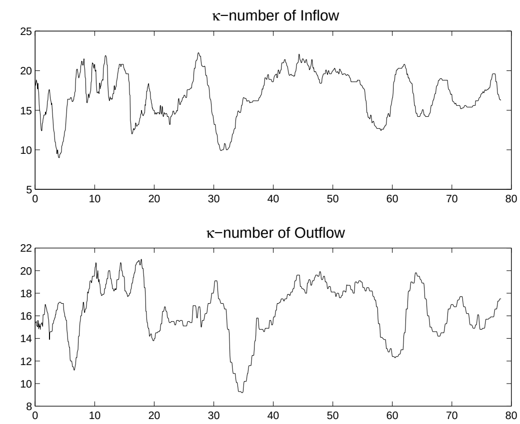


Fig. 9. The input and output $\kappa$-numbers resampled according to the text.

the same procedure to the resampled data $u_1$ to $y_1$. The best process model fit was obtained for

$$G(s) = \frac{0.8116}{1+110.28s} e^{-369.58s}$$

The comparison is shown in Figure 8.4. This “looks good”, since the peaks and dips of the model output perfectly fit the corresponding times of the true output. The residence time is thus very well captured. Without the fundamental non-linear transformation that corresponds to the resampling to a time-invariant model such a good model would be impossible to obtain. In principle, this nonlinearity could have been picked up within a nonlinear black-box model, but evidently the number of data points (800) in the estimation data was not sufficient for this.

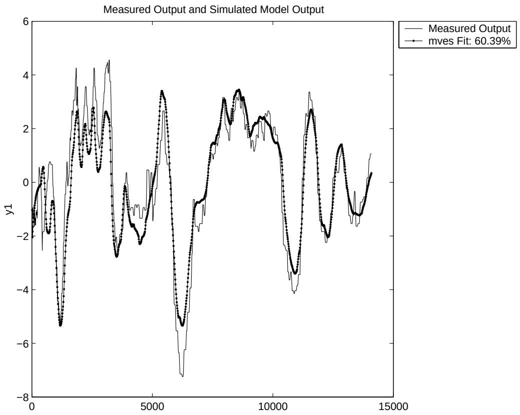


Fig. 10. Comparison between the model output (thick line) and the measured output (thin line) for the buffer vessel data. The model is estimated using the first half of the data. The data used here are the detrended, decimated data, resampled as described in the text.

## 9. CONCLUSIONS

A major purpose of this contribution has been to point to the underlying basic features of all identification methods. Most of the commonly used methods can be seen as "curve-fitting" between the predicted output and the measured one. The only difference in this perspective is the parameterization of the predictor. Algorithms, basic asymptotic results and the major considerations that the user is faced with all have a common ground.

We have illustrated this common ground by describing both general linear models and a large family of popular non-linear black box models. In the course of this we have illustrated and displayed the basic mechanisms of such black box models and the reasons for their general approximation capabilities.

Simple Process Models have been of particular interest in this contribution. We have shown how also these simple "three-parameter-models" fit into the general framework. In particular, experimental conditions, focus filters, and intersample behaviours have to be treated with care, along the lines of the general identification theory, also for static-gain + dominating time constant + delay models.

A punch line has been that even if you build simple models it may be rewarding or even necessary to do the "semi-physical homework" to find a reasonable model, with a suitable shade of grey.

## 10. REFERENCES


Andersson, T. and P. Pucar (1995). Estimation of residence time in continuous flow systems with dynamics. *Journal of Process Control* **5**, 9–17.

Juditsky, A., H. Hjalmarsson, A. Benveniste, B. Delyon, L. Ljung, J. Sjöberg and Q. Zhang (1995). Nonlinear black-box modeling in system identification: Mathematical foundations. *Automatica* **31**(12), 1724–1750.

Ljung, L. (1999). *System Identification - Theory for the User*. 2nd ed.. Prentice-Hall. Upper Saddle River, N.J.

Ljung, L. (2001). Estimating linear time invariant models of non-linear time-varying systems. *European Journal of Control* **7**(2-3), 203–219. Semi-plenary presentation at the European Control Conference, Sept 2001.

Ljung, L. (2002*a*). Identification for control: Simple process models. In: *The IEEE 2002 Conference on Decision and Control*. Las Vegas. submitted for publication.

Ljung, L. (2002*b*). *System Identification Toolbox for use with* MATLAB. *Version 6.*. 6th ed.. The MathWorks, Inc. Natick, MA.

Ljung, L. and T. Glad (1994). *Modeling of Dynamic Systems*. Prentice Hall. Englewood Cliffs.

Rake, H. (1980). Step response and frequency response methods. *Automatica* **16**, 519–526.

Sjöberg, J., Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P.Y. Glorennec, H. Hjalmarsson and A. Juditsky (1995). Nonlinear black-box modeling in system identification: A unified overview. *Automatica* **31**(12), 1691–1724.

Söderström, T. and P. Stoica (1989). *System Identification*. Prentice-Hall Int.. London.

Åström, K. J. and T. Hägglund (1995). *PID Controllers: Theory, Design, and Tuning*. 2nd ed.. Instrument Society of America. Triangle Research Park, N.C.

Ziegler, J. G., N. B. Nichols and N.Y. Rochester (1943). Process lags in automatic-control circuits. *Trans. ASME* **65**, 443–444.

**Abstract**

**Keywords:**

**Avdelning, Institution**
Division, Department

Division of Automatic Control
Department of Electrical Engineering

**Datum**
Date

2003-12-03

**Språk**
Language

- ☐ Svenska/Swedish
- ☒ Engelska/English
- ☐ .......................

**Rapporttyp**
Report category

- ☐ Licentiatavhandling
- ☐ Examensarbete
- ☐ C-uppsats
- ☐ D-uppsats
- ☒ Övrig rapport
- ☐ .......................

**ISBN**
..........

**ISRN**
..........

**Serietitel och serienummer**
Title of series, numbering

**ISSN**
1400-3902

LiTH-ISY-R-2464

**URL för elektronisk version**

http://www.control.isy.liu.se

**Titel**
Title

System Identification and Simple Process Models

**Författare**
Author

Lennart Ljung, ,

**Sammanfattning**
Abstract

.

**Nyckelord**
Keywords