PERSPECTIVES ON THE PROCESS OF IDENTIFICATION

Lennart Ljung
Department of Electrical Engineering, Linköping University
S-581 83 Linköping, Sweden. e-mail ljung@isy.liu.se

Abstract. To identify mathematical models of dynamical systems based on observed inputs and outputs is an important task in many different applications. Here, we shall review some perspectives on this process. The Engineer's perspective is considered to be the most important one.

Keywords. System identification, modeling, dynamical models

1 INTRODUCTION

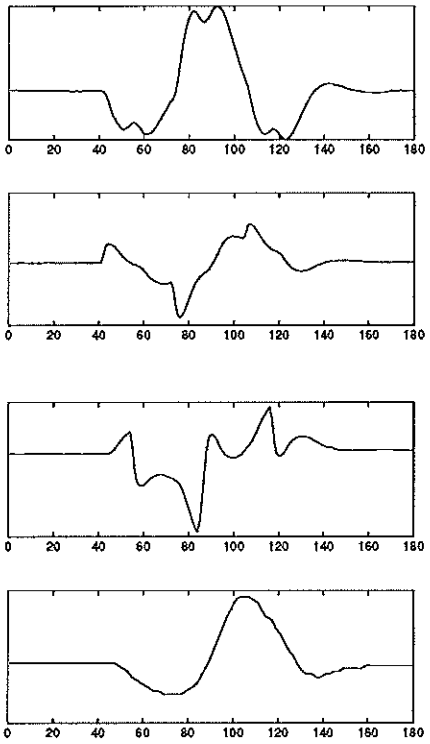

Figure 1: Results from test flights of the new Swedish aircraft JAS-Gripen, developed by SAAB Military Aircraft AB, Sweden. From above:
Pitch rate.
Elevator angle.
Canard angle.
Leading edge flap.

The problem is to use the information in these data to determine the dynamical properties of the aircraft for fine-tuning regulators, for simulations and so on. Of particular interest are the aerodynamical derivatives.

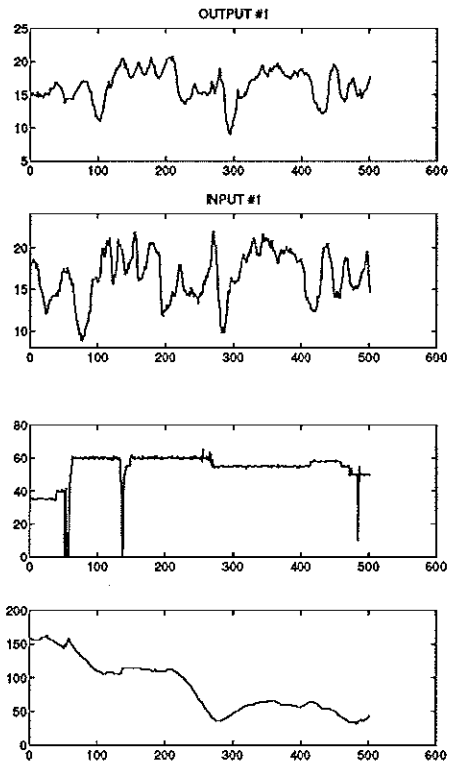

Figure 2: From the pulp factory at Skutskär, Sweden. The pulp flows continuously through the plant via several buffer tanks. From above:
The κ-number of the pulp flowing from a tank.
The κ-number of the pulp flowing into the tank.
Level in the tank.
Flow out from the tank.

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

2 THE ENGINEER'S PERSPECTIVE

An engineer, who is faced with questions such as those described in the introduction has the following perspective:

- How can I best use the information in the observed data to calculate a model of the system's properties?
- How can I know if the model is any good, and how can I trust it for simulation and design purposes?
- How shall I manipulate the input signals (like the rudder angles in the airplane example) to obtain as much information as possible about the system?
- What kind of software support is available for the tasks?

From the engineer's point of view, the end objective with the theory and methodology of System Identification is to provide answers and solutions to these questions. Here we shall give a brief overview of this process of identification, keeping in mind the end objective, even though the tour may take us via more or less sophisticated mathematical frameworks.

3 THE STATISTICIAN'S PERSPECTIVE

The science of inferring models from observations is formally the statistician's area, and it is polite to first consider his perspective. In the traditional probabilistic/statistical framework the observations are viewed as outcomes (realizations) of a random variable (random vector) with an unknown probability density function (pdf). So, let Z^t denote a random vector corresponding to all the observations in one of the figures in the introduction. Let

$$p_{Z^t}(x, \theta) \tag{1}$$

be the pdf of this vector. This function depends on the parameter vector θ, which describes a family of possible pdf:s (i.e. a family of possible probability measures).

The problem in statistical inference is to gain information about the probability measure, i.e., about the parameters θ, from the observation of an outcome of the random vector Z^t. The most celebrated point estimator of θ is the *maximum likelihood* (ML) estimator which selects θ so as to make the actual observation "most likely" according to the pdf, i.e.

$$\hat{\theta}_t = \arg\max_{\theta} p_{Z^t}(z^t, \theta) \tag{2}$$

where z^t is the actually observed data vector ("arg max" denotes the maximizing argument). The ML estimator has of course a long history and has been the subject of extensive analysis, e.g. [6], [29], and [28].

A related, but conceptually somewhat different statistical approach is formed by *non-linear regression*, where a sequence of observations $\{y(t)\}$ is described by

$$y(t) = g_t(\theta_0) + e(t) \tag{3}$$

where $\{g_t(\theta)\}$ is a family of t-dependent functions of the parameter vectors and $e(t)$ is a zero-mean random variable. The time dependence of g could very well be in terms of an observed vector $\varphi(t)$

$$g_t(\theta_0) = g(\theta_0, \varphi(t)) \tag{4}$$

The basic approach to estimate the unknown parameter value θ_0 is to form the *non-linear least squares* estimator

$$\hat{\theta}_t = \arg\min_{\theta} \sum_{k=1}^{t} (y(k) - g(\theta, \varphi(k)))^2 \tag{5}$$

If the $g(\theta, \varphi)$ are linear in θ, we have the familiar linear regression/linear least squares case.

Non-linear regressions have also been the subject of extensive study and analysis in the statistical literature, see, e.g. [12], [10], and [3].

4 THE PATTERN LEARNING PERSPECTIVE

Here is a typical problem in science and human learning: We are shown a collection of vector pairs

$$\{[y(t), \varphi(t)];\ t = 1, \ldots, N\} \tag{6}$$

Call this "the training set". We are then shown $\varphi(N+1)$ and asked to name a corresponding value for $y(N+1)$. The variable t could be thought of as time, but could be anything. The vectors $y(t)$ and $\varphi(t)$ may take values in any sets (finite sets or subsets of $\Re^n$ or anything else) and the dimension of $\varphi(t)$ could very well depend on t (and be unbounded).

This formulation covers most kinds of classification and model building problems. Conceptually, the problem amounts to learning a certain "pattern" from (6) and then applying that pattern to new data. This probably resembles human learning and the scientific process in general terms. For the problems described in the introduction we would let $y(t)$ be the variable to be explained (like the pitch rate at time t) and $\varphi(t)$ contains relevant information about inputs (the rudder angles) and past values of $y(s)$, $s < t$.

The area of *pattern recognition* [7] of course explicitly addresses the problem, but is usually reserved for the case where $y(t)$ assumes values in a finite set (the classes).

Projection methods in statistics like Projection Pursuit [11], or partial least squares, (PLS) [30], view each of the pairs as a point in a $\dim y + \dim \varphi$ dimensional space, and try to find patterns and clusters – mostly linear subspaces – by suitably chosen linear projections. There should be a link between these general projection methods, factor analysis and the new subspace methods for state-space system identification, [20], [17], but that is yet to be explored.

Moreover the general problem of finding patterns in (6) relates to *learning theory*, which deals with how complex patterns can be detected from observations, e.g. [27]. It is for the future to explore and exploit the connection between this theory and the techniques of system identification.

A specific way of finding the pattern in (6) is of course to link $y(t)$ and $\varphi(t)$ by a function g:

$$y(t) \sim g_t(\theta, \varphi(t)) \tag{7}$$

This function may depend on t and also has a number of free parameters θ that are adjusted to reproduce the pairs in the training set (6) as well as possible. Obviously we have then come very close to the non-linear regression formulation (3).

5 THE CONTROL SCIENTIST'S PERSPECTIVE: SYSTEM IDENTIFICATION

If we specify that the sought relationships between the observations shown in Section 1 are of dynamical system character, the control scientist enters the arena. It is then natural to look for models of the general type

$$\begin{aligned} \dot{x} &= f_t(x, u, v, \theta) \\ y &= h_t(x, u, v, \theta) \end{aligned} \tag{8}$$

where θ again denotes a vector of parameters whose values are to be determined from data. Of course, several variants of (8) are possible.

In (8) y denotes the output and u and v the input signals to the dynamical system. It is convenient to distinguish between the *measured input* u , and the *unmeasured input* v. The set of measurements thus consists of u and y:

$$z^t = \{u(s), y(s),\ s < t\} \tag{9}$$

The unmeasured input v is usually thought of as "disturbances and noise". Clearly we need some sort of assumptions about the character of v, in order to proceed to find a good value of θ, based on the information in z^t. There are two basic approaches to such assumptions.

- *Non-probabilistic*: Constrain the set of possible signals $\{v(t)\}$ in some way, like

$$|v(t)| \leq C \quad \forall t \tag{10}$$

  In general we may write for the "allowed" disturbances:

$$v \in V(\theta) \tag{11}$$

- *Probabilistic*: Assign probabilities to the different possible $\{v(t)\}$ sequences. That is, describe $\{v(t)\}$ as a random process with known or parameterized probability distribution:

$$v \text{ has pdf } p_v(\cdot, \theta) \tag{12}$$

The non-probabilistic approach

Given a model description (8) and some constraint on possible $\{v(t)\}$-sequences (such as (10)) the identification problem is conceptually very simple:

Find all those values of θ, such that (8) holds for the observed z^t in (9) and a v subject to (11). This gives

$$\theta \in \Omega_\theta \tag{13}$$

Calculating Ω_θ could of course be an overwhelming task. There is a rather extensive literature on various ways to do this. Most often one has to be content with an overbound $\bar{\Omega}_\theta$

$$\Omega_\theta \subset \bar{\Omega}_\theta,$$

with a more simple version of (8), viz

$$y(t) = \theta^T \varphi(t) + v(t) \tag{14}$$

$$\varphi(t) = [-y(t-1), \ldots, -y(t-n), u(t-1) \ldots, u(t-m)]^T \tag{15}$$

and with the simple constraint (10). The approach has been called "unknown-but-bounded" noise, "set membership identification" and "optimal algorithms". See among many references [21] [18], and [13].

The probabilistic approach

Given a model description like (8) and a probabilistic measure (12) for v, it is, at least conceptually, possible to eliminate the unmeasured x and v and rewrite the expression in innovations form as

$$y(t) = g_t(z^t, \theta) + e(t) \tag{16}$$

The pdf of the *innovations* $e(t)$ will be a function $f_e(\cdot, \theta, t)$.

What we have done by assigning a probability measure to v, is of course indirectly to specify a probabilistic measure (1) for the observations z^t. Indeed, the measure can be explicitly specified as

$$\log p_{Z^t}(x^t, \theta) = \sum_{k=1}^{t} \log f_e(\epsilon(k, \theta), \theta, k) \qquad (17)$$

where

$$\epsilon(t, \theta) = y(t) - g_t(z^t, \theta) \qquad (18)$$

This gives an immediate link between the basic statistical method of maximum likelihood, (2), and innovation representations of dynamical systems. The link to the non-linear regression perspective (3) – (5) is also clear.

A pragmatic approach

A more pragmatic approach to estimating the dynamics of a system is simply to postulate a *predictor model structure*, i.e., look for a description of the observed data within a family of models

$$\hat{y}(t|\theta) = g_t(z^t, \theta) \qquad (19)$$

where the prediction of $y(t)$ is denoted by $\hat{y}(t|\theta)$. The prediction is based on observations available at time $t-1$,

$$z^t = [y(t-1), u(t-1), \ldots, y(0), u(0)] \qquad (20)$$

and is an arbitrary (differentiable) function of these data and of the parameter vector θ. The actual output will then differ from the prediction by an error $e(t)$

$$y(t) = \hat{y}(t|\theta) + e(t). \qquad (21)$$

We then seek that value of θ that has the best track record in achieving good prediction

$$\hat{\theta}_N = \arg\min_{\theta} \frac{1}{N} \sum_{t=1}^{N} \ell(t, \theta, \epsilon(t, \theta)) \qquad (22)$$

$$\epsilon(t, \theta) = y(t) - \hat{y}(t|\theta) \qquad (23)$$

It is clear that by invoking a probabilistic framework, i.e. by assigning a pdf to $\{e(t)\}$ in (21) the pragmatic estimate (22) can be seen as an ML estimate (2), (17) – (18).

It is also clear that by choosing

$$\ell(t, \theta, x) = \begin{cases} 0 & |x| \leq C \\ \infty & |x| > C \end{cases} \qquad (24)$$

the method (22) will pick out those θ which are consistent with the assumption $|e(t)| \leq C$ in (20) for all $0 \leq 1 \leq N$. Thus the non-probabilistic approach also fits into (22).

Most "traditional" control oriented descriptions of System Identification follow this mixture of pragmatic and probabilistic approaches. See e.g. [15] and [23].

6 THE MODEL STRUCTURE

The single most important step in the identification process is to decide upon a model structure such as (8). In practice typically a whole lot of them are tried out and the process of identification really becomes the process of evaluating and choosing between the resulting models in these different structures.

It is natural to distinguish between three types of model structures:

1. Black-box structures
2. Structures from physical modelling
3. Structures from semi-physical modelling

Black-box structures

A black-box structure is one where the parametrization in terms of θ is chosen so that the family of models $\{g_t(z^t, \theta)\, \theta \in D_m\}$ covers as; "many common and interesting" ones as possible. No particular attention to the actual application is then paid. For a linear system (a linear mapping from past data to future ones) we could for example think of choosing the parameters as the impulse response coefficients of a finite impulse response model

$$\hat{y}(t|\theta) = \sum_{k=1}^{M} \theta_k u(t-k) \qquad (25)$$

More common in control applications is the ARX black box structure for linear systems:

$$\begin{aligned} \hat{y}(t|\theta) = &-a_1 y(t-1) - a_2 y(t-2) - \ldots \\ &-a_n y(t-n) + b_1 u(t-1) + \ldots + b_m u(t-m) \end{aligned} \qquad (26)$$

"the mother of all dynamical model structures".

In general we can write a black box structure conceptually as

$$\hat{y}(t|\theta) = \sum_{k=1}^{M} \theta_k h_k(z^{t-1}) \qquad (27)$$

i.e., as some kind of function expansion. In the general case the basis functions $\{h_k\}$ may also depend on θ. In most cases the h_k are also constrained to be functions of a fixed dimensional vector $\varphi(z^t)$ (like

$$\varphi(z^t) = (y(t-1), \ldots, y(t-n), u(t-1), \ldots, u(t-m))$$

Each observation thus corresponds to a point in the space spanned by the φ-vectors, which in the case above is equal to $\Re^{n+m}$. The functions h_k are thus mappings from this space to the space where

the outputs take their values. Depending on the character of these mappings, it is instructive to distinguish between two principally different basis functions:

- Local: Each of the h_k has support only in a small local box in the φ-space. That is, $h_k(\varphi)$ is zero unless φ belongs to a certain neighborhood (that depends on k).
- Global: Each of the h_k have support in the whole φ-space

Among black-box structures that use global basis functions are all the usual linear black box models, Volterra series expansions and so on.

The local basic functions models can be visualized as a multidimensional table: The φ-space has been split up into a number of boxes. A new observation $\varphi(t)$ then falls into one of these boxes, the one corresponding to say h_k, and the predicted output is then taken as θ_k (or possibly interpolated, taking into account a few neighboring boxes). The sizes and locations of the boxes can be determined with the aid of estimation data. The extreme case is when the boxes are determined so that exactly one data point $\varphi(t)$, $t = 1 \ldots, N$ has fallen in each box: this is the so called nearest neighbor approach [26]. All this is well established in the statistical literature under names of "non-parametric regression" and "density estimation" [24], [4].

Neural network model structures, e.g. [19], represent a spectacular revival of these techniques. So called radial basis networks correspond to localized bases (where the "boxes" overlap like Gaussian distribution functions), while the feedforward sigmoid network formally would use global basis functions (although the "dynamic effects" really are localized). Fuzzy modelling [14] is again an example of localized basis functions with typically polynomial interpolation rules, which are inherited from the "membership functions".

It is worth stressing that these new techniques of neural net modelling and fuzzy identification represent a useful revitalization of non-linear black box modelling with some new particular structures, but at the same time they definitely fall into a very old and classical framework of estimation techniques (See, e.g. [16], [2].)

Structures from physical modelling

In case we have physical insight into the properties of the system to be identified, it is natural to exploit this: *"Don't estimate what you already know!"* Basically we then write down those physical laws and relationships that describe the system. Most often they are then summarized in a state space form like (8) where θ denotes unknown physical constants in the description. The identification process is then to estimate these constants. That route takes us from (8) via (16) (explicitly or implicitly) and (22) to the estimate $\hat{\theta}_N$. We are thus entirely within the framework of Section 4. The work to arrive at (8) and then to actually carry out the minimization of (22) can be considerable, though.

Semi-physical model structures

The logical route to utilize available physical knowledge may - as pointed out - be quite laborious. It is then tempting to instead try some simple black-box structures, such as the ARX model (26) *("Try Simple Things First")*. This is quite OK, but it should in any case be combined with physical insight. Here is a toy example to illustrate the point:

"Suppose we want to build a model for how the voltage applied to an electric heater affects the temperature of the room. Physical modelling entails writing down all equations relating to the power of the heater, heat transfer, heat convection and so on. This involves several equations, expressions and unknown heat transfer coefficients and so on. A simple black-box approach would instead be to use, say the ARX-model (26) with u as the applied voltage and y the room temperature. But that's too simple! A moment's reflection reveals that it's the heater power rather than the voltage that gives the temperature change. Thus use (26) with u= squared voltage and y= room temperature"

I would like to coin the term *semi-physical modelling* for introducing non-linear transformation of the raw measurement, based on high-school physics and common sense. The transformed measurements are then used in black-box structures such as the ARX structure.

Clearly semi-physical modelling is in frequent use. It is however also true that many failures of identification are indeed to be blamed on not applying this principle.

Hybrid structures

Of particular current interest is to conceive model structures that are capable of dealing both with dynamic effects, described by differential/difference equations and with logical constraints, "the if:s and the but:s" of the system. Not so many concrete results have yet been obtained in this area, but quite intense work is going on now. We may point to some work on using tree models and pattern recognition for these hybrid

model structures: [25], [22].

7 MODEL VALIDATION

It is not enough to come up with a nominal model $\hat{\theta}_N$ from (22) – we must also have a measure of its reliability. *Model validation* is the process of examining the model, assessing its quality and possibly rejecting its use for the purpose in question. In a sense this could be viewed as the *essential process of identification* – the estimation phase is really just a means to provide candidate models that might pass the needle's eye of validation.

Model validation has at least three different objectives:

1. To decide if the model is "good enough" for the intended application.
2. To decide how "far from the true system description" the model might be.
3. To decide whether the model and the data indeed are consistent with assumptions of the model structure.

These objectives partly overlap, but it is still possible to single out basic techniques:

1. The most obvious and pragmatic way to decide if a model is good enough is to test how well it is able to reproduce validation data (data that were not used to estimate the model) in simulation or prediction. The user can then by eye inspection decide if the fit is "good enough". In my mind this is the prime validation tool.
2. To determine *error bounds* – how far is the true system from the model – is a fundamentally difficult question. If we adopt a probabilistic setting and assume that the true system is to be found within the chosen structure, it becomes a matter to see how much the stochastic disturbances might have affected the model. The covariance matrix of the asymptotic distribution is classically used for the error bounds in this case, This covariance matrix is generally given by

$$\mathrm{cov}\{\hat{\theta}_N\} \sim \frac{1}{N} E e^2(t) [\mathrm{cov}\{\frac{d}{d\theta}\hat{y}(t|\theta)\}]^{-1} \quad (28)$$

   for the structure (21). If we (according to 3) below) cannot disprove that the true system can be represented in the chosen structure it is still reasonable to use the measure (28).

   The remaining cases – no probabilistic setting adopted and/or the used model structure is known to be too simple – has spurred a considerable interest recently. See among many references [9] and [8]. It would lead too far to review that literature here.
3. The test if the data and the model are consistent with the model structure assumptions, is again a more straightforward task. Basically we compute the residuals $y(t) - \hat{y}(t|\hat{\theta}_N) = \epsilon(t)$ from the model and a (validation) data set and check if
   (a) $|\epsilon(t)| < C$ in the deterministic setting (10)
   (b) $\epsilon(t)$ and $u(t-\tau)$ are independent random variables, in the probabilistic setting (u is the input to the system).

The latter test is one of many *residual* analysis tests that can be performed, and this is standard statistical practice, see e.g. [5].

8 THE ENGINEER'S PERSPECTIVE: SOLVING THE PROBLEM

The ultimate objective of the theoretical and mathematical perspectives must of course be to develop methods to solve the engineer's problem as posed in Section 2.

It follows from our discussion that the most essential element in the process of identification – once the data have been recorded – is to try out various model structures, compute the best model in the structures, using (22), and then validate this model. Typically this has to be repeated with quite a few different structures before a satisfactory model can be found.

While one should not underestimate the difficulties of this process, I suggest the following simple procedure to get started and gain insight into the models.

1. Find out a good value for the delay between input and output, e.g. by using correlation analysis.
2. Estimate a fourth order linear model with this delay, using part of the data. Then simulate this model with the input and compare the model's simulated output with the measured output over the whole data record. In MATLAB language this is simple,

```
z = [y u];
compare(z,arx(z(1:200,:),[4 4 1]));
```

If the model/system is unstable or has integrators, use prediction over a reasonably large time horizon instead of simulation.

Now, either of two things happen:

- *The comparison "looks good".* Then we can be confident that with some extra work – trying out different orders, and various noise models – we can fine tune the model and have an acceptable model quite soon. Let me add here that I am amazed by the large amount of applications that fall into this category.
- *The comparison "does not look good".* Then we must do further work. There are three basic reasons for the failure.
    1. *A good description needs higher order linear dynamics.* This is actually in practice the least likely reason, except for systems with mechanical resonances. One then obviously has to try higher order models or focus on certain frequency bands by band pass filtering.
    2. *There are more signals that significantly affect the output.* We must then look for what these signals might be, check if they can be measured and if so include them among the inputs. Signal sources that cannot be traced or measured are called "disturbances" and we simply have to live with the fact that they will have an adverse effect on the comparisons.
    3. *Some important non-linearities have been overlooked.* We must then resort to semi-physical modelling to find out if some of the measured signals should be subjected to non-linear transformations. If no such transformations suggest themselves, we might have to try some non-linear black-box model, like a neural network.

Clearly, this advice do not cover all the art of identification, but it is the best half page summary of the practical process of identification that I can offer.

Example 1 *Aircraft dynamics*

Let us try the recipe on the aircraft data in figure 1! Picking the canard angle only as the input, estimating a fourth order model based on the data points 1 to 80, gives figure 3. (We use 10-step ahead prediction in this example since the models are unstable – as they should be, JAS has unstable dynamics in this flight case). It does not "look good". Let us try alternative 2: More inputs. We repeat the procedure using all three inputs in figure 1. That is , the model is computed as

```
arx([y u1 u2 u3],[4 4 4 4 1 1 1])
```

on the same data set. The comparison is shown in figure 4. It "looks good". By further fine-tuning, as well as using model structures from physical modelling, only slight improvements are obtained.

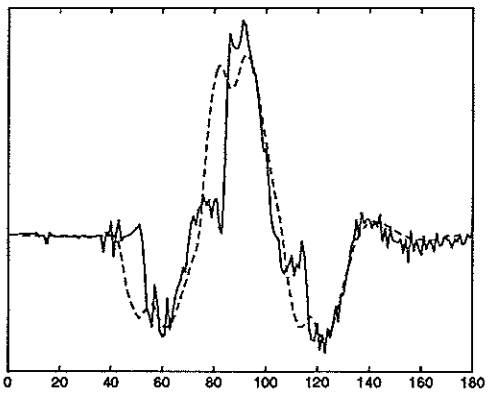

Fig. 3: Dashed line: Actual Pitch rate. Solid line: 10 step ahead predicted pitch rate, based on the fourth order model from canard angle only.

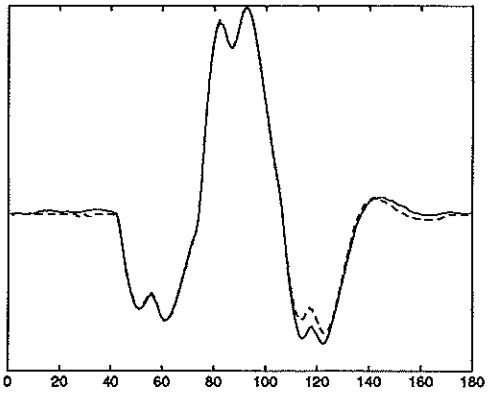

Fig. 4: As figure 3 but using all three inputs.

Example 2 *Buffer vessel dynamics*

Let us now consider the pulp process of figure 2. We use κ-number before the vessel as input and the κ-number after the vessel as output. The delay is preliminarily estimated to 12 samples. Our recipe, where a fourth order linear model is estimated using the first 200 samples and then simulated over the whole record gives figure 5. It does not look good.

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) the vessel would have a dynamics of a pure delay equal to vessel volume divided by flow. 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

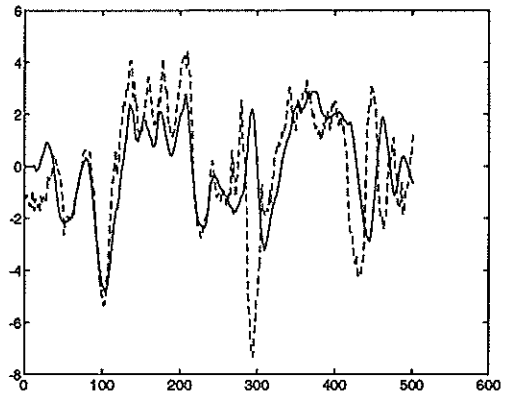

Fig. 5: Dashed line: κ-number after the vessel, actual measurements. Solid line: Simulated κ-number using the input only and a fourth order linear model with delay 12, estimated using the first 200 data points.

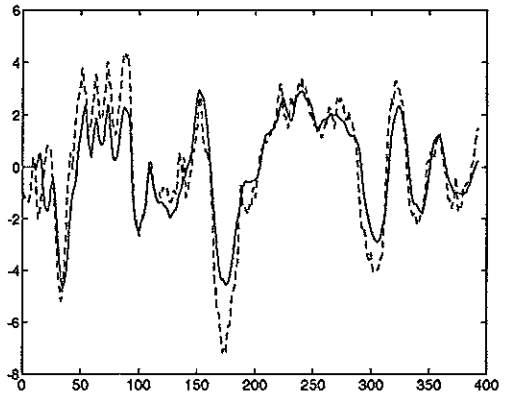

Fig. 6: Same as figure 5 but applied to resampled data

flows divided by volume. In MATLAB terms this will be

```
z = [y,u]; pf = flow./level;
t = 1:length(z)
newt =
table1([cumsum(pf),t],[pf(1):sum(pf)]');
newz = table1([t,z], newt);
```

We now apply the same procedure to the resampled data. This gives figure 6. This "looks good". Somewhat better results can then be obtained by fine-tuning the orders.

9 CONCLUSIONS

The area of process identification is one where real practical applications and rather advanced mathematical tools and perspectives meet. The meeting place is really the software into which many years' research has been packaged. There are now many successful such packages commercially available. They have become standard tools in many industrial applications. This again stresses that it is the engineer's perspective that is the ultimate one in this area.

Acknowledgements

I am happy to acknowledge many years' financial support from the National Swedish Board for Technical Development, and the Swedish Research Council for Engineering Sciences.

I am also grateful to SAAB Military Aircraft AB for allowing me to show the test flight data from JAS-Gripen, and to Skutskär Paper and Pulp Mill for the process data. The analysis of the latter data has been carried out by Torbjörn Andersson and Predrag Pucar, [1].

REFERENCES

[1] T. Andersson and P. Pucar. Estimation of residence time in continuous flow systems with varying flow and volume. Technical Report LiTH-ISY-I-1460, Department of Electrical Engineering, Linköping University, Sweden, 1993.

[2] A.R. Barron. Statistical properteis of artificial neural networks. In *Proceedings of the 28th IEEE Conference on Decision and Control*, pages 280–285, 1989.

[3] R. J. Carroll and D. Ruppert. *Transformation and Weighting in Regression.* Chapman and Hull, New York, 1988.

[4] L. Devroye and L. Györfi. *Non-parametric density estimation.* Wiley, New York, 1985.

[5] N.R. Draper and H. Smith. *Applied Regression Analysis, 2nd ed.* Wiley, New York, 1981.

[6] R.A. Fisher. On an absolute criterion for finnting frequency curves. *Mess. Math.*, 41:155, 1912.

[7] K. Fukunaga. *Introduction to statistical pattern recognition (2nd ed.).* Academic Press, New York, 1990.

[8] G.C. Goodwin, M. Gervers, and D.Q. Mayne. Bias and variance distribution in transfer function estimates. In *Proc. 9th IFAC Symp. Identif. Syst. Parameter Estimation*, Budapest, Hungary, 1991.

[9] G.C. Goodwin, M. Gevers, and B. Ninness. Quantifying the error in estimated transfer functions with application to model order selection. *IEEE Trans. Automatic Control*, 37(7):913–929, 1992.

[10] E.J. Hannan. Nonlinear time series regression. *J. Appl. Prob.*, 8:767–780, 1971.

[11] P.J. Huber. Projection pursuit. *Ann. Statist*, 13:435–475, 1985.

[12] R.I. Jennrich. Asymptotic propertise of nonlinear least squares estimation. *Ann. Autom. Control*, 40(2):633–643, 1969.

[13] R.L. Kosut, M.K. Lau, and S.P. Boyd. Set-membership identification of systems with parametric and nonparametric uncertainty. *IEEE Trans. Autom. Control*, 37(7):929–942, 1992.

[14] C.W. Ku and Y.Z. Lu. Fuzzy model identification and self-learning. *IEEE Trans. on SMC*, 17, 1987.

[15] L. Ljung. *System Identification - Theory for the User*. Prentice-Hall, Englewood Cliffs, N.J., 1987.

[16] L. Ljung and T. Söderström. *Theory and Practice of Recursive Identification*. MIT press, Cambridge, Mass., 1983.

[17] B. Ottersten M. Viberg, L. Ljung and B. Wahlberg. Performance of subspace based state-space, system identification methods. In *Proc. 12th IFAC Congress, Session Th-M-5*, Sydney, Australia, 1993.

[18] M. Milanese and R. Tempo. Optimal algorithms for robust estimation and prediction. *IEEE Trans. Automatic Control*, AC-30:730–738, 1985.

[19] K.S. Narendra and K. Parathasarathy. Identification and control of dynamical systems using neural networks. *IEEE Trans. Neural Networks*, 1:4–27, 1990.

[20] P. Van Overschee and B. De Moor. N4sid: Subspace algorithms for the identification of combined deterministic-stochastic systems . *Automatica, (Special Issue)*, 29(5):to appear, 1993.

[21] F.C. Schweppe. *Uncertain Dynamical Systems*. Prentice-Hall, Englewood Cliffs, 1973.

[22] A. Skeppstedt, L. Ljung, and M. Millnert. Construction of composite models from observed data. *Int. j. Control*, 55(1):141–152, 1992.

[23] T. Söderström and P. Stoica. *System Identification*. Prentice-Hall Int., London, 1989.

[24] C.J. Stone. Consistent non-parametric regression (with discussion). *Ann. Statist.*, 5:595–645, 1977.

[25] J.E. Strömberg, F. Gustafsson, and L. Ljung. Trees as black-box model structures for dynamical systems. In *Proc. 1st European Control conference (ECC'91)*, pages 1175–1180, Grenoble, France, 1991.

[26] T.M.Cover and P.E. Hart. Nearest neigbor pattern classification. *Trans. IEEE Info. Theory*, IT-13:21–27, 1967.

[27] V.N. Vapnik and A. Ya. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theor. Probability Appl.*, 56:269–280, 1971.

[28] A. Wald. Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statis.*, 20:343 – 357, 1949.

[29] P. Whittle. *Hypothesis testing in time series analysis*. Thesis. Uppsala University, Almqvist and Wiksell, Uppsala. Hafner, New York, 1951.

[30] S. Wold, A. Ruhe, H. Wold, and W.J. Dunn III. The collinearity problem in linear regression, the partial least squares(PLS) approach to generalized inverses. *SIAM J Sci. Stat. Computs.*, 5(3):735–743, 1984.

/home/joakim/ulla/Reports/sydney.wri