

**EXPERIENCE AND CHALLENGES IN  
IDENTIFICATION OF NON-LINEAR SYSTEMS:  
BIOCHEMICAL AND ENVIRONMENTAL  
APPLICATIONS**

**K.J. Keesman \***

*\* Systems and Control Group, Wageningen University  
THE NETHERLANDS  
Tel.: +31 317 483780, Fax: +31 317 414957  
E-mail address: karel.keesman@wur.nl*

Abstract: In this short paper for the panel discussion on "Experience and challenges in identification of non-linear systems" some major issues with respect to identification of non-linear biochemical and environmental systems are presented. ©IFAC 2006

Keywords: identification, non-linear systems, biotechnology, environmental engineering

## 1. INTRODUCTION

Identification of non-linear (static) systems has a long tradition, although it mainly focussed on the parameter estimation problem. The Gauss, Newton(-Raphson) and steepest descent method are therefore known for a long time. However, in practice, and especially in the pre-computer era, people frequently used graphical methods on *e.g.* log-log paper to obtain (rough) parameter estimates. Hence, transformation of the problem into a linear regression problem was a well-known and widely accepted method. After the introduction of computers and the development of simulation models, at least initially, most often manual calibration was applied. Nowadays, it is still applied among practitioners to obtain a first guess. It is only since the early 90's that computational software related to the aforementioned numerical methods has been more and more used in solving non-linear estimation problems (*e.g. leastsq* or *lsqnonlin* in Matlab, *Goal.Seek* in Excel, *etc.*) and it is expected that this line will be continued.

Some key problems in non-linear estimation are: (i) presence of local minima, (ii) high computational effort (most often in a small region

around local/global minimum, (iii) specification of constraints to preserve physical interpretation of model. Moreover, practical identification of non-linear systems, apart from the parameter estimation problem and including model selection, model quality assessment and so on, is most often further hindered by limited amounts of data and/or limited prior system's knowledge.

In Section 2 the mathematical modeling of biochemical and environmental systems and related identification problems are further discussed. Then in Section 3 some solutions to the key problems in the identification of non-linear systems are suggested. The paper finishes with some concluding remarks.

## 2. MATHEMATICAL MODELING

Many mathematical modeling approaches exist, ranging from *data-based* (using non-linear regressions, neural networks, *etc.*) to *physical* modeling (using first principles and constitutive relationships). For biochemical and environmental applications, the grey-box or *experimental* modeling approach is usually our starting point. Let us illustrate this to a biochemical process. The dynamics

of a component ( $x$ ), such as substrate, in a continuously stirred tank bio-reactor are described by

$$\frac{dx}{dt} = -\xi(x) + \frac{u_1}{V}(u_2 - x) \quad (1)$$

see e.g. Bastin and Dochain (1990), where  $\xi(\cdot)$  is the reaction kinetics term,  $u_1$  the flow rate,  $u_2$  the concentration of  $x$  in the so-called influent and  $V$  the volume of the reactor. Hence, bilinear terms appear in (1). The reaction term is frequently described in terms of so-called Monod or Michaelis-Menten kinetics, *i.e.*

$$\xi(x) = \mu \frac{x}{K + x} \quad (2)$$

where  $\mu$ , the kinetic reaction rate, is a function of the maximum reaction rate, the biomass concentration and some conversion factor. The term  $\frac{x}{K+x}$ , with  $K$  the half-saturation constant, describes reaction limitations at low concentrations of  $x$ , for which also alternative expressions exist. Clearly, (1) needs further adjustment for reactors with different flow conditions. For instance, in batch reactors  $u_1 = 0$  and in fed-batch reactors  $V$  becomes a state variable instead of a constant. Hence, (1) and (2) are elementary building blocks in bio-reactor modeling and serve as a starting point for identification. Key problems in the identification of biochemical processes are:

- badly identifiable parameters,
- existence of local minima in non-linear least-squares estimation,
- time-varying parameters,
- limited amount of substrate/biomass data,
- trade-off between bias reduction and loss of physical insight in unconstrained optimization.

In environmental applications models become far more complex. For instance, in addition to biochemical processes, the effect of solar radiation, usually expressed in terms of an exponential function, may become important. Modeling effects of erosion most often introduces discontinuous functions, *etc.* Moreover, these system are usually described as compartmental models with catenary, circular, mammillary or full web structure with many parameters. In particular, environmental systems are driven by disturbances, as solar radiation, wind, precipitation and for identification, usually, prior knowledge, amount of data and possibilities for experiment design are limited.

### 3. IDENTIFICATION APPROACHES

In this section some suggestions to solve the first four key problems in the identification of biochemical processes will be presented. Identifiability analysis on models with many parameters is

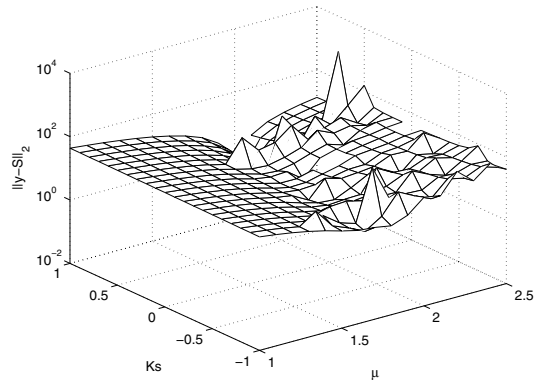


Fig. 1. Norm of the prediction error with  $x = S$  (substrate),  $K = 1$  and  $\mu = 2$ .

still an unsolved problem, but for smaller models as (1 - 2) it is known that both  $\mu$  and  $K$  are theoretically identifiable. However, Holmberg and Ranta (1982) showed that, given noise corrupted data, it is difficult to estimate the parameters accurately. Then, one option is to design experiments optimally such that *e.g.* the variance of the estimates is minimal (see *e.g.* Keesman and Stigter (2002); Stigter and Keesman (2004)).

Notice that (1 - 2) is a non-linear-in-the-parameters model. Fig. 1 shows the existence of local minima for this type of rational model (see Doeswijk and Keesman (2005) for details). However, for the discrete-time version of the model we obtain

$$x_{k-1}(x_k - x_{k-1}) = \begin{bmatrix} -(x_k - x_{k-1}) & -x_{k-1} \end{bmatrix} \begin{bmatrix} K \\ \mu \end{bmatrix} \quad (3)$$

which becomes a linear regression with all its nice properties. However, when  $x$  is corrupted with noise, the estimation problem becomes an *errors-in-variables* problem. Fortunately, total least-squares solutions exist to solve it. Instead of focussing on the estimation of the physical parameters, in this case  $\mu$  and  $K$ , (3) can, after re-arranging terms, be directly used for prediction. Consequently, for rational models this so-called *linear regressive re-parametrization* approach followed by a linear re-arrangement of terms towards a predictor structure circumvents the problem of local minima and high computational demands during estimation. This idea can be further expanded to non-rational models. For instance, let

$$\dot{x} = e^{\alpha x} \quad (4)$$

Differentiation of both sides gives

$$\begin{aligned} \ddot{x} &= \alpha e^{\alpha x} \dot{x} \\ &= \alpha (\dot{x})^2 \end{aligned} \quad (5)$$

where  $\alpha$  is the unknown parameter. Hence, in discrete-time

$$x_{k+1} - 2x_k - x_{k-1} = \alpha(x_k - x_{k-1})^2 \quad (6)$$

which again is in a linear regression form from which the linear predictor can be easily derived. If, however, the model becomes too complex to re-parametrize it in a linear regression form, prior knowledge about parameter correlations and initial parameter values is essential in solving non-linear estimation problems. van Straten (1985) used steady-state analysis to find parameter correlations. Keesman (2002) proposed a method to find initial estimates from chemical structure formulae and available experimental data. In addition to this, (graphical) analysis of the cost function surface may also help to understand the problem of local minima (see Abusam et al. (2001)).

Bio-processes, due to the living organisms present in the system, are often described by models like (1-2) with time-varying parameters. Hence, there is a need to simultaneously estimate state variables and model parameters, leading to augmented state vectors of large dimension. It has been demonstrated by Keesman (2002) how among other techniques singular perturbation analysis may help to reduce the dimension of the estimation problem. In *batch* bio-reactors, one has the opportunity to distinguish between different batch phases, like phases of unlimited growth or phases with only decay. Consequently, the estimation of states and parameters should preferably be limited to these specific phases in which only a limited amount of parameters are significant.

The limited availability of experimental data is another issue that impedes the estimation of parameters and evaluation of parameter estimation uncertainty in bio-reactor or environmental models. As an alternative to a stochastic approach, which fails for small data sets, a so-called *unknown-but-bounded* or *set-membership* approach has been proposed. For details we refer to Milanese et al. (1996) and, in particular, for non-linear estimation problems to Keesman (2003).

#### 4. CONCLUDING REMARKS

Key problems in the estimation of parameters in bio-reactor models have been indicated. In environmental problems we encounter the same problems, although in general the models are much more complex and thus the dimensions of the parameter space are much higher. This curse of dimensionality is an issue in many problems, such as non-linear optimization and identifiability analysis. Identification for control, but then within a dynamic optimization context for non-linear systems, the development of automatic identification procedures using some multi-model approach, the design of measurement networks for complex systems and the application of advanced particle

filtering techniques, are other issues that need (further) attention in the near future.

Nowadays, there is a trend to focus on specific model classes, as e.g. polynomial, Volterra, Wiener or Hammerstein models and to aim at further development of identification procedures using specific system's knowledge.

If we look back, we see a very active system identification community. Hence, it is expected that in the near future this community may be well able to tackle the above-mentioned problems.

#### References

- A. Abusam, K. J. Keesman, G. van Straten, H. Spanjers, and K. Meinema. Parameter estimation procedure for complex non-linear systems: calibration of ASM No.1 for N-removal in a full-scale oxidation ditch. *Wat. Sci. Tech.*, 43(7):367–376, 2001.
- G. Bastin and D. Dochain. *On-Line Estimation and Adaptive Control of Bioreactors*. Elsevier, Amsterdam, 1990.
- T. G. Doeswijk and K. J. Keesman. Parameter Estimation and Prediction of a Nonlinear Storage Model: an algebraic approach. In *IEEE Conference on Control and Automation, June 27-29, Budapest, Hungary, Paper MM-3.2*. 2005.
- A. Holmberg and J. Ranta. Procedures for Parameter and State Estimation of Microbial Growth Process Models. *Automatica*, 18(2): 181–193, 1982.
- K. J. Keesman. State and parameter estimation in biotechnical batch reactors. *Control Eng. Pract.*, 10:219–225, 2002.
- K. J. Keesman. Bound-based identification: Nonlinear-model case. In H. Unbehauen, editor, *Encyclopedia of Life Science Systems article 6.43.11.2*. UNESCO EOLSS, 2003.
- K. J. Keesman and J. D. Stigter. Optimal Parametric Sensitivity Control for the Estimation of Kinetic parameters in Bioreactors. *Math. BioSc.*, 179:95–111, 2002.
- M. Milanese, J.P. Norton, H. Piet-Lahanier, and E. Walter (Eds.). *Bounding Approaches to System Identification*. Plenum, London, 1996.
- J. D. Stigter and K. J. Keesman. Optimal parametric sensitivity control of a fed-batch reactor. *Automatica*, 40(8):1459–1464, 2004.
- G. van Straten. Analytical methods for parameter-space delimitation and application to shallow-lake phytoplankton-dynamics modeling. *Appl. Math. Comp.*, 17:459–482, 1985.