WHAT DOES SYSTEM IDENTIFICATION HAVE TO OFFER?

M. Gevers & G. Bastin

Louvain University, Bâtiment Maxwell, B-1348 Louvain la Neuve, Belgium

Abstract. This contribution concerns the application of system identification to real-life problems. What system identification has to offer depends upon the objective the user has in mind and the prior knowledge he has about the system. We show how these two elements must be combined to choose a model structure; this is the most difficult part of an identification application.

Keywords. Identification; Model selection; Model structure; Parameter estimation.

1. INTRODUCTION

In the four companion papers of this session [1]-[4] a systematic methodology has been developed for the treatment of an identification problem. Technical answers have been given to the questions any user is faced with when he has to identify a model for a system within a given model set. But the first question a user faces is: "Does system identification have something to offer for my problem and, if yes, how much can I hope for?" The answer depends upon the objective the user pursues and the prior knowledge he has.

In this presentation we will consider System Identification from an applications viewpoint. The choice of an appropriate model set is the most difficult step of the identification procedure. We shall see that prior knowledge and the pursued objective are the main ingredients for this choice. For a given objective the amount of prior knowledge often enables the user to say a priori whether system identification has something to offer, and how much.

We present a number of applications to illustrate our arguments. These are all drawn from our own experience over the last 10 years. This is therefore a rather personal account, which does not claim to cover the considerable amount of applied work that has been accomplished by many other groups throughout the world.

2. WHAT CAN SYSTEM IDENTIFICATION BE USED FOR?

The possible objectives can be broadly classified as follows.

2.1. Estimation of parameters that have a physical interpretation.

This covers a broad class of applications, particularly in non industrial problems. Several examples are given to illustrate this.

Example 1: Groundwaterflow model
The flow of water in a groundwater reservoir is described by the following 2-dimensional partial differential equation:

$$S\frac{\partial h}{\partial t} = \text{div } T \text{ grad } h + q \qquad (2.1)$$

Here h is the water level (=the state), q is the input flow rate (=the input), S is the storage coefficient and T is the transmissivity. S and T are functions of the soil characteristics. The estimation of S and T is often the purpose of the identification, because they provide a physical insight into the behaviour of the reservoir. For example, a high value of T in a subregion means that a high pumping rate can be achieved in that area.

Example 2: Cardiovascular model.
The cardiovascular system can be represented by the following electrical model:

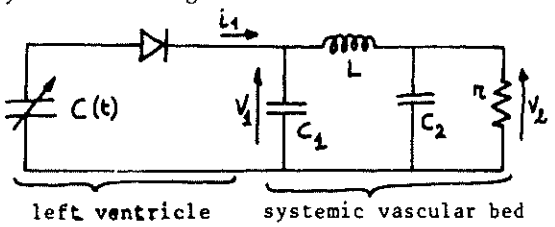

Fig. 1

Voltages represent blood pressures, while currents represent blood flows. The parameters C_1, C_2, L, r and C(t) can be estimated from measurements of aortic pressure p(t) (i.e. $v_1(t)$) and ejected blood flow q(t) (i.e. $i_1(t)$). These parameters can be given a physical interpretation: C_1 and C_2 are related to the distensibility of the vascular wall, L is related to the inertia of the blood mass, r represents the peripheral resistance, while 1/C(t) is related to the contractility of the left ventricle [5].

Example 3: A biomethanization process.
Biomethanization is an anaerobic fermentation process where the decomposition of biodegradable matters by bacterial populations is linked to methane gas production. The third stage of the biochemical reaction can be described by the following nonlinear state-space model [6]:

$$\frac{d}{dt}\begin{pmatrix} x \\ s \end{pmatrix} = \begin{pmatrix} -kx + \frac{\mu^* xs}{K+s} - gx \\ -\frac{\mu^* xs}{Y(K+s)} + pgu - gs \end{pmatrix} \qquad (2.2a)$$

$$q = Y^* \frac{\mu^* xs}{K+s} \qquad (2.2b)$$

x (concentration in methanigenic bacteria) and s (substrate concentration) are the state variables; q (gas production) is the output; g (dilution rate) and u (concentration of acids in the influent) are the inputs. The 6 parameters have the following interpretation:

- k : decay rate for the bacteria
- K : saturation of the bacterial growth
- p : unknown part of the introduced acids which are hydrolized during the reaction
- μ^*: specific growth rate for the bacteria
- Y and Y^*: yield coefficients.

It is easy to observe that these physical parameters cannot be uniquely identified from inputs and outputs. Defining

$\tilde{x} = Y^* x$, $\tilde{s} = s/K$, $\tilde{Y} = Y^* KY$, $\alpha = p/K$

leads to an equivalent state-space model that has only 4 unknown parameters: K, μ, $\tilde{Y}$, α. However the physical significance of the 2 parameters $\tilde{Y}$ and α is not so clear.

2.2. Obtaining a model for simulation studies.

In order to study the effect of different control strategies on a variable of interest, it is of course always best to work on the system itself. But in practice this is not always possible for the following reasons :

a) experiments may be too costly;
b) the system can be unstable : experimenting can then be dangerous;
c) the time constants of the system can be very large, which would make the experimenting much too slow.

For these reasons it is often required to design a model of the system which is then used for simulation studies.

Example 1: Groundwater flow model.
A groundwater flow model allows one to select by simulation the best pumping locations and to study the effect of new pumping stations on the base flow of the rivers that drain that reservoir.

Example 3: Biomethanization process.
The model allows us :

a) to look for optimal nominal steady state values
b) to study various starting procedures for the process.

These studies are almost impossible on the process itself, because

a) in its present working condition it is technically very difficult to maintain the process in steady state
b) the starting procedure for a biomethanization reactor takes several weeks as opposed to a few minutes on the simulated model.

2.3. Prediction.

In many applications, it is desired to predict in real time the future evolution of the output of a system, based on measurements made on the system. One can distinguish two cases : either the output is considered as a purely stochastic random process, or it is related to another measurable variable.

Example 4: Riverflow prediction.
This example will be extensively discussed in section 5. The two cases just mentioned will be considered, depending on whether or not rainfall and evapotranspiration inputs are used to predict future riverflows.

2.4. Control.

In many applications, particularly in industrial applications, the objective of the identification exercise is to design a controller.

Example 5: Temperature control in a glass-furnace.
The production of high-quality glass requires that the variations of the temperature in the glass melt be kept at a strict minimum. In [7] a rather simple lumped stochastic model was identified, which relates some temperatures in the glass melt to the fuel flow. This model was then used to design a predictor and a minimum variance feedback controller.

The classification in terms of the pursued objective is useful because what can be accomplished with system identification, and the related question of what prior knowledge is required, is very often a function of the objective. One should bear in mind, however, that several of these objectives will often be pursued at the same time. For example, in the groundwater flow application, objectives 1 and 3 are simultaneously pursued.

3. WHAT FORM CAN PRIOR KNOWLEDGE TAKE AND HOW TO OBTAIN IT ?

In this section we describe different forms of prior knowledge and the ways of obtaining it. We shall use the name "prior knowledge" to denote two different concepts :

- real knowledge or information about the model structure, the parameters or the data, that has been communicated to the user or that he has gathered from the data.
- assumptions that the user makes before he embarks on his identification exercise (such prior assumptions are often required to guarantee identifiability). Prior knowledge can be available about the model structure, the parameter vector θ or the data Z.

3.1. Knowledge about the model structure.

The structure of the model set can be obtained essentially in two different ways.

a) From physical principles or experimental findings.
   Examples: The structure of the groundwater flow model or the methane reactor model are based on physical principles, while the nonlinear part of the rainfall-riverflow model (see section 5) is based on experimental findings.
b) From data analysis.
   Example: Nonlinear relationships can often be detected by correlating the time-variation of a parameter with other measurable variables of the system. See [8] for examples.

3.2. Knowledge about the parameters.

This prior knowledge can take three forms :

a) $\theta \in D$ or $g_\alpha(\theta) = 0$; the parameters lie in a domain or satisfy a constraint which depends on a small number of parameters α. The estimation of θ(often a high-dimensional vector) is replaced by the estimation of α(usually of small dimension).
   Example: In the groundwater flow application, the transmissivities must be positive, for physical reasons. For identifiability reasons, the assumption is sometimes made that the transmissivity is a polynomial function of space : see section 6.
b) a prior estimate $\hat{\theta}(.)$ of θ is available : this gives an initial guess for θ in the optimization algorithm.
c) a prior density $f_\alpha(\theta)$ is given or is assumed. The estimation problem becomes Bayesian. This prior knowledge will improve the accuracy of the estimate of θ, and can be used in the computation of the covariance on the error of θ. The groundwater flow application in section 6 will illustrate this point.

The estimation of a prior statistical model for θ contains two steps :

1) the estimation of the structure of this model. It can be obtained from experimental findings; very simple Gauss-Markov models are usually adopted. (Example: a Wiener field in the ground-waterflow application of section 6; a Wiener process in many time-varying models).
2) the estimation of the parameters α of this statistical model. (If θ is a Wiener process, $\theta_k = \theta_{k-1} + \varepsilon_k$, then the variance $E\{\varepsilon_k^2\} = \alpha$ must be estimated). If θ is a measurable physical variable, θ can sometimes be estimated a priori from experimental findings. In most cases α will be adjusted a posteriori in such a way that the prediction error variance on the outputs of the model is of the same order of magnitude as the variance of the measurement errors. See section 6.

3.3. Knowledge about the data.

This knowledge can take the following two forms :

a) $Z \in D_Z$, i.e. the data are constrained.
   Example: riverflows can only be positive.
b) $f_\alpha(Z)$: a prior density is given.
   Example : Experimental findings tell us that rainfalls and riverflows obey a log-normal distribution. Therefore, in the rainfall-riverflow model the input-output data are replaced by their logarithms. This has two advantages :
   a) the data now have a gaussian distribution
   b) there is no risk that the model would predict a negative riverflow.

4. HOW TO CHOOSE A MODEL SET FROM PRIOR KNOWLEDGE ?

In many peoples' minds system identification suggests the idea that there is a "true system". In fact there does not exist something as a true system. But since this concept is useful and operational, we shall consider that the user postulates that there is an exact description of that part of reality that is of interest to him. This exact description (which is the user's approximation of reality) is called by him the true system. To fix the ideas, assume that the true system is as follows :

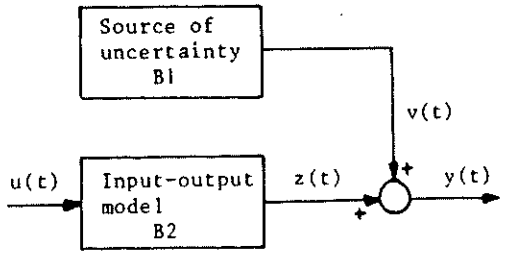

Fig. 2a: The true system

The user's idea about the true system will depend upon the prior knowledge he has about each of the boxes B1 and B2 of Fig.2a. A limiting case is where he has no prior knowledge at all.

The task of the user is then to choose a model set and to estimate its parameters. This choice will be a function of his objective and will be based on his prior knowledge. It is the most difficult part of the identification procedure. In order to clarify the discussion, we distinguish between two categories of models (both are of course predictors in the sense defined in [2]) :

a) prediction models :

$$\hat{y}(t+1|\theta) = f_p(U^t, Y^t; \theta)$$

   The prediction of $y(t+1)$ is based on all input-output data up to time t.
b) simulation models :

$$\hat{y}(t+1|\theta) = f_s(U^t; \theta)$$

   The prediction of $y(t+1)$ uses the inputs only.

Remarks: 1) both models can be identified using a prediction error identification method that minimizes some function of the errors $y(t+1) - \hat{y}(t+1|\theta)$.

2) simulation models are often called "output error models".

We can now formulate an answer to the question "What does system identification have to offer?" by considering each of the objectives separately.

4.1. Prediction.

The objective is to construct a model that is able to predict $\hat{y}(t+1)$ from measurements of input-output data taken on the system. Such a model can be identified and its performance can be evaluated without any prior knowledge about the "true system": the quality criterion coincides with the identification criterion. The model will converge to the best possible predictor in the chosen model set [2]. Of course any knowledge about what is in B1 and B2 will be helpful in selecting the model set and will improve the performance of the predictor (see the riverflow models of section 5).

4.2. Simulation

Here the objective is to construct a model that is able to simulate the behaviour of the true system even when the environment is different than that which prevailed during data collection. This is the case, e.g., in the methanization process. One could think of identifying a prediction model $\hat{y}(t+1|\theta) = f_p(Y^t,U^t;\theta)$, and of transforming it into a simulation model by replacing Y^t by $\hat{Y}^t$: $\hat{y}(t+1|\theta) = f_p(\hat{Y}^t,U^t;\theta)$. But this could yield disastrous results, because the best prediction model is not always the best simulation model as the following example indicates.

Example: Let the true system be

$$y(t) + a_o y(t-1) = b_o u(t-1) + c_o e(t-1)$$

where $\{u(t)\}$ and $\{e(t)\}$ are independent sequences of i.i.d. zero mean random variables of unit variance. See Fig. 2b.

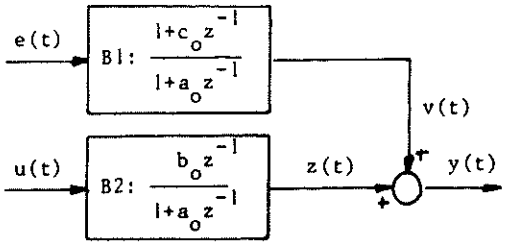

Fig. 2b

Consider first a prediction model:

$$\mathcal{M}_1: \hat{y}(t+1|a,b) = -ay(t) + bu(t)$$

Then the best predictor will yield $a = a_o - c_o/r_o$, $b = b_o$, where $r_o = E\{y^2(t)\}$. Consider now a simulation model:

$$\mathcal{M}_2: \hat{y}(t+1|a,b) = -a\hat{y}(t|a,b) + bu(t)$$
$$= \frac{bz^{-1}}{1+az^{-1}}\, u(t+1)$$

Within model set $\mathcal{M}_2$, the best predictor will yield $a^* = a_o$, $b^* = b_o$. Now suppose that the steady-state gain of the system must be computed. Then the simulation model will give the correct answer, $b_o/1+a_o$, while the prediction model will yield a completely erroneous answer $b_o/(1+a_o-c_o/r_o)$.

With a simulation model we want to model B2 only, in order to predict z(t), i.e. we want to filter out the effect of the noise. With a prediction model on the other hand we want to predict y(t); hence we want to model v(t). The important difference between prediction and simulation is the following:
- with prediction, the model selection criterion coincides with the identification criterion: minimize some function of the prediction error.
- with simulation, the identification criterion (min $E1(t,\theta,\varepsilon(t,\theta))$ does not coincide with the model selection criterion: in the above example the identification criterion would select $\mathcal{M}_1$, which is a poor simulation model (except in the hypothetical case where the model set contains the true system). The identification criterion is unable to discriminate between a good or a bad model; only prior knowledge or extensive simulation studies can do this.

4.3. Estimation of parameters that have a physical interpretation.

The model set must now be chosen to estimate physically meaningful parameters that cannot be directly measured. The quality criterion for the selection of the model set is that the estimated $\hat{\theta}$ must be as close as possible to the true θ_o. Here again this selection criterion does not coincide with the identification criterion. Therefore:
a) prior knowledge about the structure of the relationship between parameters and data is essential. The user must know or postulate a "true system" (or at least that part of the true system that contains the parameters of interest);
b) the model set must be chosen such that $\hat{\theta}_N$ converges to the true θ_o. It must therefore contain this true system and it must be uniquely identifiable. Roughly speaking this means that the model set $\mathcal{M}(\theta)$ must have the same input-output properties as the true system for a unique value of the vector θ. Here the inputs are understood to contain both the deterministic and the stochastic inputs. A more precise definition of uniquely identifiable model sets is given in [9].

Prediction error methods allow us to make statements about convergence and asymptotic covariances of parameter estimates (see [2]). This makes sense only in the context of parameter estimation when the structure of the true system is known a priori. If in addition prior statistical knowledge is available about θ, this can be used to increase the accuracy of the estimates. See section 6; see also [3].

4.4. Control.

For most control applications, particularly for minimum variance control, it is required to have a good predictor. Therefore the discussion of section 4.1 applies. It is even

less important here that the model set contains the true system, because feedback tends to reduce the sensitivity to model inaccuracies. However if the objective is to simulate control strategies (including control actions that vary greatly from those that were used during data collection), then a simulation model is necessary, and section 4.2 applies.

Conclusion : We have shown in this section that what system identification has to offer depends essentially on the pursued objective and on the available prior knowledge about the true system. The pursued objective determines which minimum amount of prior knowledge is required.

5. APPLICATION 1 : A RAINFALL-RIVERFLOW PROCESS.

In this application it was desired to obtain a riverflow prediction model for the Semois river (in Belgium). The river basin has a surface area of 1235 km2. The available data were daily values during 7 years (1967-1973) of the rainfall in 17 gauges, the mean evapotranspiration over the basin, and the riverflow at the outlet of the basin. From the 17 rainfall measurements, a daily mean rainfall over the whole basin can be computed using spatial interpolation techniques [10]. More details about this application can be found in [11]- [12].

The objectives of the identification were to obtain a short term riverflow prediction model allowing judicious actions in case of floods, and a long term simulation model allowing a forecast of the base flow (= low water) under various rainfall hypotheses. As we learned more about the system and its structure, we identified a succession of models. We first describe these models, and then compare their properties.

The following notations will be used :
L(k) = daily mean rainfall over the basin
ETP(k) = daily potential evapotranspiration
Q(k) = daily mean riverflow
L(k) and ETP(k) are the inputs of the process, while Q(k) is the output.

Model 1 : Linear AR (1)

$$Q(k) = aQ(k-1) + c + \varepsilon(k)$$

Model 2 : Linear ARMA (2,3)

$$Q(k) = \sum_{i=1}^{2} a_i Q(k-i) + \sum_{i=1}^{3} c_i \varepsilon(k-i) + c + \varepsilon(k)$$

Model 3 : Linear ARMAX (2,4,3)

$$Q(k) = \sum_{i=1}^{2} a_i Q(k-i) + \sum_{i=1}^{4} b_i L(k-i) + \sum_{i=1}^{3} c_i \varepsilon(k-i) + c + \varepsilon(k)$$

Model 4 : Nonlinear

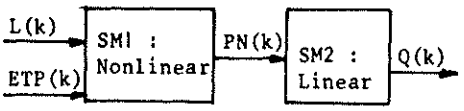

Fig. 3

Submodel 1
Submodel 1 is a nonlinear dynamical model whose structure has been obtained from physical considerations and from experimental findings. A new state variable was introduced, the storage S(k), which represents the amount of water stored on day k in the upper suface of the soil. The output of SM1 is the net rain PN(k) : it represents that part of the rain that runs on the basin surface and will eventually reach the river. The construction of this nonlinear model was based on the following considerations :
1) a rainfall will contribute to a riverflow only if there is a rainfall surplus, i.e. if $L(k) > ETP(k)$.
2) assuming that there is a surplus, then the proportion of this surplus that will run off to the river will be much higher if the surface of the soil is very wet (S(k) large), then if it is dry (S(k) small). One can also define a saturation level for S(k) : S_{max}.

Based on these ideas and on experimental data, a nonlinear model was obtained. It depends on 3 parameters that must be estimated.

Comment : SM1 as just described is a clear example of a case where the user postulates a description of reality. This description is of course based on experience and findings, but in fact it may have very little to do with reality. The important thing is that this model is identifiable and operational.

Submodel 2 : ARMAX (2,3,3)

$$Q(k) = \sum_{i=1}^{2} a_i Q(k-i) + \sum_{i=1}^{3} b_i PN(k-i) + \sum_{i=1}^{3} c_i \varepsilon(k-i) + c + \varepsilon(k)$$

The parameters of the nonlinear model and the parameters of the ARMAX model were identified by minimizing a global prediction error criterion.

Model 5 : Nonlinear

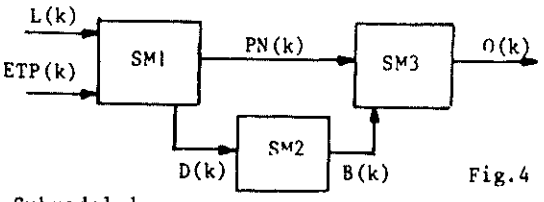

Fig.4

Submodel 1
This nonlinear submodel has an additonal output D(k), which represents the amount of water drained to the phreatic surface.

Submodel 2 : Linear
This is a simulation model for the base flow (i.e. that part of the riverflow that originates from groundwater). The following model was obtained :

$$B(k) = BR(k) + BL(k)$$

where

$$BR(k) = \alpha BR(k-1) + aD(k-d_r)$$

$$BL(k) = \beta BL(k-1) + bD(k-d_l)$$

BR and BL are the short and long term components of the base flow. The coefficient α, β, a, b, d_r, d_l are estimated during dry periods. Typically the following values were obtained : α and β correspond to time-constants of 15 and 244 days respectively, d_r = 5 days, d_l = 30 days. The structure of SM2 was again determined by prior knowledge based on experimental and physical considerations.

Submodel 3 : ARMAX (2,4,3)

$$Q(k) = B(k) + \sum_{i=1}^{2} a_i\{Q(k-i) - B(k-i)\} + \sum_{i=1}^{4} d_i PN(k-i) + \sum_{i=1}^{3} c_i \varepsilon(k-i) + c + \varepsilon(k)$$

Equivalently :

$$Q(k) = B(k) + \frac{D(z)}{A(z)} PN(k) + \frac{C(z)}{A(z)} \varepsilon(k)$$

Model 6 : Identical to model 5 but it uses only the best 3 rainfall gauges for the computation of L(k).

Table 1 compares the performances obtained with these various models.

	σ_ε	$\frac{\sigma_\varepsilon^2}{\sigma_Q^2}$
model 1	3.46	7.4 %
model 2	3.25	6.5 %
model 3	2.65	4.3 %
model 4	1.80	2.0 %
model 5	1.90	2.2 %
model 6	2.01	2.6 %

Table 1

Comments:

1) The ARMA(2,3) model performs only slightly better than a very simple AR(1) model.

2) Adding a measurable input L(k) improves the performance still further, but the really significant improvement occurs when a nonlinear model is introduced (model 4). This example clearly illustrates that system identification has much more to offer when prior knowledge (or prior assumptions) are injected by the user.

3) Adding a base flow model (models 5 and 6) slightly decreases the one-step predictive performance of the model. This is due to a decrease in the data-to-parameter ratio. It illustrates the fact that, when the objective is prediction, it is not always necessary (and sometimes harmful) to have a complex model that accurately describes the "true system". On the other hand, models 5 and 6 are essential when it comes to long term simulations of the base flow.

4) Notice finally that the model that uses only 3 pluviometers performs almost as well as the model that uses all 17. Recall however that these 3 pluviometers have been selected according to an optimal selection scheme [10].

6. APPLICATION 2 : THE GROUNDWATERFLOW PROCESS.

It follows from physical laws that the flow of groundwater is described by (2.1). In many reservoirs the fluctuations of h are so small and so slow that a steady-state model is adequate:

$$\nabla(T\nabla h) + q = 0 \qquad (6.1)$$

The objective is then to estimate T(x,y) at all points of a domain Ω from measurements of the water level h. We shall first show that this model is not identifiable without prior assumptions for T. We then introduce a deterministic model, and subsequently a stochastic model for T. Both will guarantee identifiability, but the stochastic model will, in addition, allow us to obtain a measure of the accuracy of the parameter estimates.

6.1. The problem of nonuniqueness.

For pedagogical reasons we consider first a one-dimensional flow model:

$$\frac{\partial}{\partial x} T(x) \frac{\partial h}{\partial x} + q(x) = 0 \qquad 0<x\leq L \qquad (6.2)$$

In practice the state h(x) is measured with error:

$$z(x) = h(x) + w(x) \qquad (6.3)$$

If (6.2) and (6.3) are discretized and if the boundary condition is unknown, we obtain:

$$T_k(h_{k+1}-h_k) + T_{k-1}(h_k-h_{k-1}) + q_k = 0, \quad k = 1,\ldots,n-1 \qquad (6.4)$$

$$z_k = h_k + w_k, \quad k = 0,\ldots,n \qquad (6.5)$$

The vector of unknown parameters becomes $\theta = (T_o,\ldots,T_{n-1},h_o,h_n)$. For given θ, $h_1(\theta)$, $\ldots,h_{n-1}(\theta)$ can be computed using (6.4). It appears clearly that we have n+2 parameters with only n+1 measures $z_o,\ldots,z_n$. Hence the model is not identifiable.

6.2. A deterministic approach.

We now mention two deterministic approaches to overcome this nonuniqueness problem:

a) A polynomial model is postulated for T(x):

$$T_k \triangleq T(x_k) = \sum_{l=1}^{M} \alpha_l g_l(x_k), \quad k = 0,\ldots,n$$

where M<< n, $g_l(x)$ are known polynomial functions and α_l are unkown coefficients. The following prediction error criterion is then minimized with respect to $\alpha_1,\ldots,\alpha_M,h_o,h_n$:

$$J_h = \sum_{k=o}^{n} \{z_k - \hat{h}_k(\theta)\}^2 \qquad (6.6)$$

The values of $\hat{h}_k(\theta)$ for k=1,...,n-1 in (6.6) are obtained from (6.4). Note that this solution corresponds to assuming a prior model $g_\alpha(\theta) = 0$, where α contains much fewer parameters than θ (see section 3).

b) An alternative approach, known as regularization, is to minimize a combined criterion:

$$J = J_h + \lambda J_T \qquad (6.7)$$

where J_h is as above, λ is a weighting coefficient, and J_T is a smoothness criterion:

$$J_T = \sum_{k=1}^{n-1} (T_k - T_{k-1})^2$$

The idea of these two methods is to incorporate prior structural constraints on the spatial variability of T in order to guarantee identifiability. However the choices of the prior models (polynomial structure, smoothness criterion) are rather arbitrary, and it is not always clear how to choose them. These two drawbacks disappear if a stochastic model is chosen for the spatial variability of T.

6.3. A stochastic approach.

It follows from experimental findings that the logtransmissivity can be represented by a Wiener model. Let $\theta_k = \log T_k$. Then the model becomes:

$$e^{\theta_k}(h_{k+1} - h_k) + e^{\theta_{k-1}}(h_k - h_{k-1}) + q_k = 0, \quad k = 1,\ldots,n-1 \qquad (6.8)$$

$$\theta_k = \theta_{k-1} + \varepsilon_k, \quad k = 1,\ldots,n-1 \qquad (6.9)$$

$$z_k = h_k + w_k, \quad k = 0,\ldots,n \qquad (6.10)$$

The w_k are either measurement errors, or (more often) interpolation errors; the latter is true when the water level must be interpolated at all grid nodes. We assume that $W=\{w_0,..,w_n\}$ and $\varepsilon=\{\varepsilon_1,\ldots,\varepsilon_{n-1}\}$ are two mutually independent i.i.d. Gaussian noises with variances σ_w^2 and σ_ε^2. Redefining $\theta=(\theta_1,..,\theta_{n-1},h_o,..,h_n)$ we can then reformulate the problem as follows: using the known measurements $z_o,\ldots,z_n$, find θ that maximizes the density $f_\theta(\varepsilon,W)$ subject to the constraints (6.8)-(6.9). It can be shown that this reduces to minimizing

$$J(\theta) = \frac{1}{\sigma_\varepsilon^2} J_\varepsilon(\theta) + \frac{1}{\sigma_w^2} J_W(\theta) \qquad (6.11)$$

where

$$J_\varepsilon(\theta) = \sum_{k=1}^{n-1} (\theta_k - \theta_{k-1})^2 \qquad (6.12)$$

$$J_W(\theta) = \sum_{k=0}^{n} (z_k - h_k)^2 \qquad (6.13)$$

Comments:

1) The criterion (6.11) is almost identical to (6.7). Thus we have not only given a justification for the use of the smoothness criterion, but (6.11) also gives a guide as to the choice of λ in (6.7): $\lambda = \sigma_w^2/\sigma_\varepsilon^2$.

2) In practice σ_w^2 is known (the 2-D interpolation technique used to compute estimates of h at each grid node also computes standard deviations; see [14]) but σ_ε^2 is unknown. σ_ε^2 is chosen such that the deviation between the water levels predicted by the model and the "measured" water levels (which usually result from spatial interpolation) is of the same order as the interpolation (or measurement) error (see 3.2,2). The deterministic and stochastic approaches have been successfully applied to an aquifer in Belgium [13]-[15]. In Figs. 6-7 we show the effect of the choice of σ_ε^2 on the estimated transmissivities and on the error between predicted and interpolated water levels. Figs. 6a and 7a represent the water levels simulated by the model (thick line) and the "measured" levels (fine line). The standard deviation of the measurement error is σ_w = 3m. Figs. 6b and 7b represent the corresponding transmissivities. We see from Fig. 7 that when T varies greatly in space, the fit between the simulated and the "measured" water levels is almost perfect: the root mean square (RMS) deviation is 3.5cm (compared to a measurement error σ_w of 3m). When T is very smooth, as in Fig.6, this RMS deviation is 77cm.

3) Fig. 8 shows estimated transmissivities with 2σ confidence intervals along a cross-section of the aquifer (see [15] for the error analysis). These standard deviations are Bayesian: they take into account both the prior Wiener model and the measurements of h.

4) This example illustrates several points that were mentioned in sections 2 to 4:
- the user has a clear description of her "true" system, namely eq. (6.1).
- however, without prior assumptions on T, this is not enough. If her objective is to estimate T or to do simulation studies, then system identification has nothing to offer at this point.
- if prior assumptions are made on T, the model becomes identifiable.
- the prior model on T injects new parameters ($\alpha_1,\ldots,\alpha_M$; or λ; or σ_ε^2) which must be estimated. However a high-dimensional problem has been replaced by a low-dimensional one.
- with the injection of a stochastic prior model $f_\alpha(T)$ for T, the global estimation problem becomes Bayesian. α (here σ_ε^2) must be estimated, but this increased prior knowledge improves the accuracy of the parameter estimates. $f_\alpha(T)$ is also used to compute standard deviations for $\hat{T}$.

7. CONCLUSION

Thinking, knowledge and insights are often crucial. Identification is not an automatic procedure. Choosing $\mathcal{M}$: a difficult choice.

8. REFERENCES

[1] A.K.HAJDASINSKI, P.EYKHOFF, A.A.H.DAMEN, A.J.W.VAN DEN BOOM, "The choice and use of different model sets for system identification", *6th IFAC Symp. Identification and System Parameter Estimation, Tutorial Workshop on Syst. Identif., Washington D.C., June 1982.*

[2] L.LJUNG, "Identification criteria and their asymptotic properties", *idem.*

[3] G.C.GOODWIN, "Experiment design", *idem.*

[4] L.LJUNG, "Model Validation", *idem.*

[5] B.A.DESWYSEN, A.A.CHARLIER, M.GEVERS, "Quantitative evaluation of the systemic

arterial bed by parameter estimation of a simple model", *Medical & Biological Engin. & Computing, Vol. 18, pp. 153-166, March 1980.*

[6] A.CHALON, G.BASTIN, M.INSTALLE, "Identification of a bio-methanization process: a case study", *6th IFAC Symp. Identification & System Parameter Estimation, Washington D.C., June 1982.*

[7] S.P.DABLEMONT, M.GEVERS, "Identification and control of a glass furnace", *Proc. 4th IFAC Symp. Identification & System Parameter Estimation, Tbilisi, USSR, pp. 446-456, Sept. 1976.*

[8] P.C.YOUNG, "The validity and credibility of models for badly defined systems", to appear in *Uncertainty and forecasting of water quality, M.B.Beck & G. Van Straten Eds., IIASA/Pergamon, Oxford, 1982.*

[9] M.GEVERS, G.BASTIN, "What does system identification have to offer?", extended version of this paper, *Internal Report, Louvain University, 1982.*

[10] B.LORENT, G.BASTIN, C.DUQUE, M.GEVERS, "Estimation of the mean rainfall over a region and optimal selection of the rainfall measurement locations", *6th IFAC Symp. Identif. & Syst. Parameter Estimation, Washington D.C., June 1982.*

[11] B.LORENT, M.GEVERS, "Identification of a rainfall-runoff process; Application to the Semois river (Belgium)", *4th IFAC Symp. Identif. & Syst. Parameter Estimation, Tbilisi, USSR, Sept. 1976.*

[12] B.LORENT, "River flow prediction and simulation model", *Workshop on "Recent developments in real-time forecasting/Control of water resource systems", IIASA, Austria, Oct. 1976.*

[13] G.BASTIN, M.GEVERS, "Joint use of space interpolation and optimization methods for steady-state aquifer modeling with scarce data", *Modeling, Identification and Control in Environmental Systems, Vansteenkiste Ed., North-Holland, pp. 427-441, 1978.*

[14] G.BASTIN, "Approches déterministe et stochastique pour l'identification de systèmes hydrogéologiques", *Ph.D. dissertation, Louvain University, 1979.*

[15] G.BASTIN, C.DUQUE, "Modeling of steady-state groundwaterflow systems: deterministic and stochastic approaches", *Proc. IASTED Symp. on Modeling, Identification and Control, Davos, pp. 125-129, Feb. 1981.*

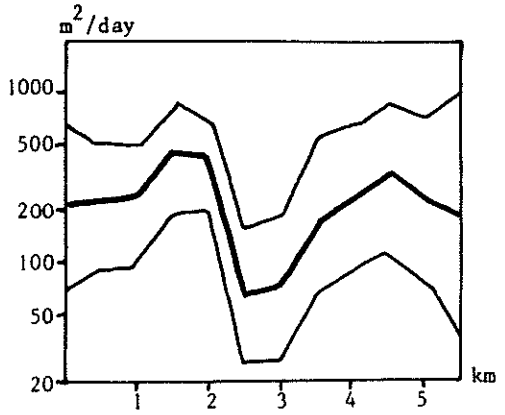

Fig. 8: estimated transmissivity (in m^2/day) and 2σ confidence intervals along a cross-section of the reservoir.

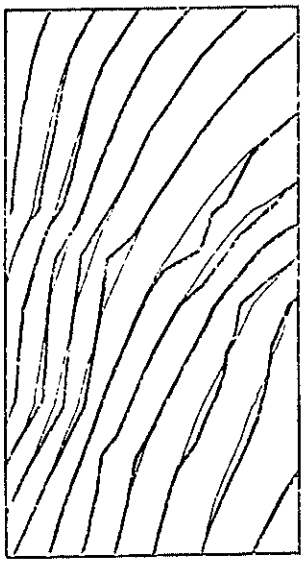

Fig.6a: simulated (——) and "measured" (▬) water levels

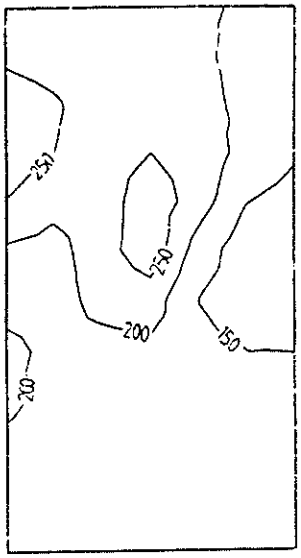

Fig.6b: estimated transmissivities (in m^2/day)

$\sigma_\varepsilon^2 = 2m^4/day$

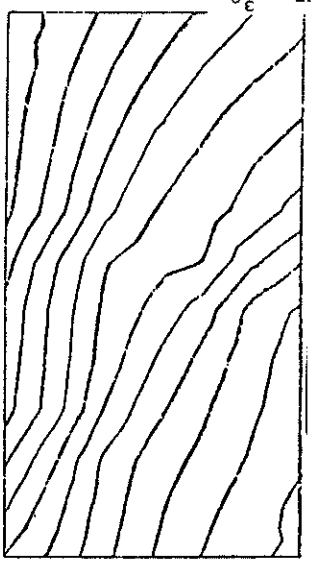

Fig.7a: simulated and "measured" water levels coincide with

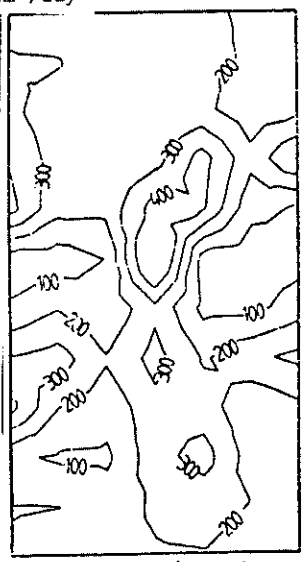

Fig.7b: estimated transmissivities (in m^2/day)

$\sigma_\varepsilon^2 = 2000m^4/day^2$