Model selection via worst-case criterion for nonlinear bounded-error estimation

S. Brahim-Belhouari, Michel Kieffer, G. Fleury, Luc Jaulin, Eric Walter

Model Selection via Worst-Case Criterion for Nonlinear Bounded-Error Estimation

Sofiane Brahim-Belhouari[1], Michel Kieffer[2], Gilles Fleury[1]*, Luc Jaulin[2,3] and Éric Walter[2]

[1] École Supérieure d'Électricité - Service des Mesures,
Plateau de Moulon, 91192 Gif-sur-Yvette, France
Phone: +33-[0]1-69-85-14-12; Fax: +33-[0]1-69-85-12-34
{Sofiane.Brahim, Gilles.Fleury}@supelec.fr

[2] Laboratoire des Signaux et Systèmes, CNRS-École Supérieure d'Électricité
Plateau de Moulon, 91192 Gif-sur-Yvette, France
{kieffer, jaulin, walter}@lss.supelec.fr

[3] On leave from Laboratoire d'Ingénierie des Systèmes Automatisés, Université d'Angers
2 bd Lavoisier, 49045 Angers, France
jaulin@babinet.univ-angers.fr

Abstract

In this paper the problem of model selection for measurement purpose is studied. A new selection procedure in a deterministic framework is proposed. The problem of nonlinear bounded-error estimation is viewed as a set inversion procedure. As each candidate model structure leads to a specific set of admissible values of the measurement vector, the worst-case criterion is used to select the optimal model. The selection procedure is applied to a real measurement problem, grooves dimensioning using Remote Field Eddy Current (RFEC) inspection.

1. Introduction

Data used to estimate the parameters of a model are always associated with some uncertainty. Evaluating the quality of a given estimator requires assessing quantitatively how this data uncertainty affects the estimates. It must be recognized that the results will be highly influenced by the assumptions explicitly or implicitly made about the error structure. Correct conclusions about the properties of the estimates require a correct characterization of the error structure.

The approach most commonly used is to express the errors in terms of stochastic uncertainty models. Due to incomplete information and the presence of structural model errors resulting from aggregation and obscurity of the process dynamics, a stochastic error approach is questionable, since many of these model errors are inherently deterministic [6]. In addition information about the reliability of the estimates can only be obtained in closed form if the error density is Gaussian and the model is linear [17] [18]. Because of these limitations, an attractive alternative to stochastic characterization has been proposed. The error in the data is no longer considered as a random variable with known or parameterized probability density function. Instead, the error is assumed to lie between some known upper and lower bounds. One is then looking for a suitable charaterization of the set of all parameter vectors consistent with the model structure, data and bounds on the error.

This paper deals with the problem of model selection for measurement purpose. A general formulation of measurement in a bounded-error context is presented in Section 2. Previous works [1] [4] [16] develop some criteria leading to the choice of a nonlinear parametric model in the context of inverse problem. None of them, however, employs a deterministic framework. In this paper a new selection procedure based on bounded-error estimation is proposed. Estimation is expressed in terms of set inversion. Using the SIVIA (Set Inversion Via Interval Analysis [5]) algorithm and a recently developed image evaluation algorithm using interval computations [8], it is possible to enclose in an approximate but guaranteed way all parameter vectors that are consistent with the data and noise assumptions. The principles of these algorithms will be briefly recalled in Section 3. Thus, a set of all admissible values of the measurement vector for each candidate model structure may be computed. In

*Corresponding author

0-7803-5276-9/99/$10.00 © 1999 IEEE

order to compare these sets, given a reference measurement, a worst-case design [7] presented in Section 4 will be used. In Section 5, the feasibility of the approach will be illustrated for a real measurement problem, grooves dimensioning using Remote Field Eddy Current (RFEC) inspection.

2. Problem formulation

A problem often encountered in various domains, such as nondestructive evaluation or so-called indirect measurements, is to estimate some unknown quantity $\boldsymbol{m}$ from a vector of observed values $\boldsymbol{y}$. This is due to the inability to use a transducer to measure $\boldsymbol{m}$ directly or for any other reason such as harsh environment [4] [14] [16]. The general problem can be described by the following equations [2] :

$$y_k = f(x_k, \boldsymbol{\theta}) + e_k \qquad k = 1, \ldots, N, \tag{1}$$

$$\boldsymbol{m} = \mathbf{g}(\boldsymbol{\theta}). \tag{2}$$

The first one (1) is the classical nonlinear regression model, where the observation variable $\boldsymbol{y} = [y_1, \ldots, y_N]^\top$ is related to $\boldsymbol{x} = [x_1, \ldots, x_N]^\top$ the vector of the experiment design (e.g., time, frequency or space coordinates) and to the observation error e_k. A vector $\boldsymbol{\theta}$ ($\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$) of p unknown parameters is then to be estimated from the N pairs of observations (x_k, y_k), where Θ is the prior feasible set for the parameters.
In the context of bounded-error estimation (e.g. [12], [13], [17] and the references therein), the feasible observation error will be expressed now in terms of a set , i.e., $e_k \in \mathbb{E}_k$, where :

$$\mathbb{E}_k = \{e_k \in \mathbb{R} : \quad \underline{e}_k \leq e_k \leq \overline{e}_k\}, \quad k = 1, \ldots, N. \tag{3}$$

In this equation $\underline{e}_k$ and $\overline{e}_k$ are the lower and upper bounds on the kth observation error, assumed to be known. Estimating parameters in this context amounts to looking for the set $\mathbb{S}$ of all admissible values of $\boldsymbol{\theta}$ that are consistent with (1) and (3). $\mathbb{S}$ is thus the intersection of Θ and the set of the solutions for $\boldsymbol{\theta}$ of the $2N$ inequalities :

$$y_k - \overline{e}_k \leq f(x_k, \boldsymbol{\theta}) \leq y_k - \underline{e}_k, \qquad k = 1, \ldots, N. \tag{4}$$

This bounded-error approach leads to set estimates, contrary to usual applications of the maximum-likelihood approach which yield point estimates. $\mathbb{S}$ has been called the posterior feasible parameter set.
The quantity to be measured, $\boldsymbol{m}$, is related to the parameters $\boldsymbol{\theta}$ via (2). It is usually defined by a functional of the parametric model $\boldsymbol{m} = \mathcal{G}(f)$ (i.e., involving derivation, integration, interpolation, extrapolation, ...). This relation is then transformed into a function of $\boldsymbol{\theta}$.
Thus, a second problem to be solved is to find the set $\mathbb{M}$ of all admissible values of $\boldsymbol{m}$ corresponding to the image of $\mathbb{S}$ using the function g.
The creation of candidate model structures remains mainly of a heuristic nature. So suppose that a set F of n_f structures of nonlinear parametric model is available. For the experimental conditions used, each candidate model f_j ($j = 1, \ldots, n_f$) generates a vector output :

$$\mathbf{f}_j(\boldsymbol{x}, \boldsymbol{\theta}) = [f_j(x_1, \boldsymbol{\theta}), \ldots, f_j(x_N, \boldsymbol{\theta})]^\top, \tag{5}$$

The choice of a given structure affects directly the estimate of $\boldsymbol{m}$. This is basically related to the sensitivity of the models f_j and function g to $\boldsymbol{\theta}$. This paper is concerned with the problem of model selection in such a deterministic framework.

3. Set inversion for finding parameter sets

The algorithms used for characterizing $\mathbb{S}$ depend on whether the model output is linear in the parameters. In the first case, it is possible to use three types of algorithm involving outerbounding ellipsoids or orthotopes, or an exact description [17].
When the model is nonlinear in the parameters, the problem is much more difficult since $\mathbb{S}$ may be nonconvex and even nonconnected. But it is still possible to get a precise description of $\mathbb{S}$ using interval analysis and set-inversion. The task of finding all $\boldsymbol{\theta} \in \Theta$ satisfying (4) may be expressed as that of characterizing the set :

$$\mathbb{S} = \{\boldsymbol{\theta} \in \Theta : \quad \mathbf{f}(\boldsymbol{x}, \boldsymbol{\theta}) \in \boldsymbol{y} - \mathbb{E}\}, \tag{6}$$

where $\mathbb{E} = \mathbb{E}_1 \times \ldots \times \mathbb{E}_N$. $\mathbb{S}$ may be equivalently expressed as :

$$\mathbb{S} = \mathbf{f}_\Theta^{-1}(\boldsymbol{y} - \mathbb{E}) = \mathbf{f}_\Theta^{-1}(\mathbb{Y}), \tag{7}$$

where $\mathbf{f}_\Theta^{-1}$ is the reciprocal function (in a set-theoretic sense) of $\mathbf{f}$ defined over Θ, and $\mathbb{Y} = \boldsymbol{y} - \mathbb{E}$ is the prior feasible set for the model outputs. For a given function $\mathbf{f}$, interval

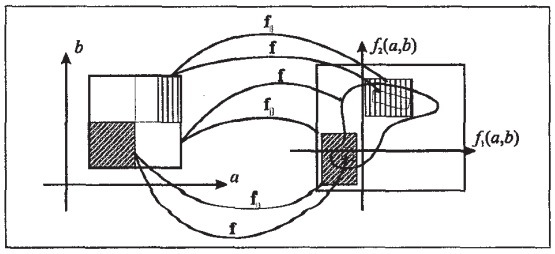

Figure 1. Inclusion function $\mathbf{f}_{[]}$ of a function $\mathbf{f}$.

analysis [11] provides an *inclusion function* $\mathbf{f}_{[]}$ (see Figure 1) that returns boxes (vector intervals) that are guaranteed to contain the image by $\mathbf{f}$ of any given box $[\boldsymbol{\theta}]$ included in

the domain of $\mathbf{f}$. The smaller the box $[\theta]$ is, the more accurately $\mathbf{f}_{\square}([\theta])$ describes $\mathbf{f}([\theta])$ [11] [15]. This property is used in the SIVIA [5]) algorithm. Suppose that Θ is a box; a box enclosing the range of $\mathbf{f}$ over Θ is obtained evaluating $\mathbf{f}_{\square}(\mathbf{x},\Theta)$. If $\mathbf{f}_{\square}(\mathbf{x},\Theta) \cap \mathbb{Y} = \emptyset$, then it is proved that no $\theta \in \Theta$ may satisfy (4) and $\Theta \cap \mathbb{S} = \emptyset$; if $\mathbf{f}_{\square}(\mathbf{x},\Theta) \subset \mathbb{Y}$, then all values of $\theta \in \Theta$ satisfy (4), and Θ is incorporated in the approximate solution set $\mathbb{S}^0$. If none of these condition is satisfied, Θ may be bisected into two sub-boxes on which the same tests are applied recursively. Boxes are bisected till their width is lower than some precision parameter ϵ. At this step, if nothing has been proved on the box, they are stored into an uncertain set $\mathbb{S}^+$.
One may thus enclose $\mathbb{S}$ into two *subpavings* (union of nonoverlapping boxes) :

$$\mathbb{S}^0 \subset \mathbb{S} \subset (\mathbb{S}^0 \cup \mathbb{S}^+) . \tag{8}$$

Obtaining the image $\mathbb{M}$ of the set $\mathbb{S}$ by the function g is also achieved using interval analysis by a recently developed IMAGESP (Image SubPaving evaluation) algorithm [8]. This algorithm is able to return a subpaving, i.e., a union of boxes, containing the image of a given subpaving (for example $\mathbb{S}^0$) by a function (for example g), with any desired precision.

4. Worst-case criterion

The theory of model building is still poor and one is usually confronted to the model selection problem, in which several mathematical functions compete to approximate empirical observations.
Consider a finite number of competing models $f_j (j = 1, \ldots, n_f)$, each of which yields a particular evaluation $\mathbb{M}_j$, of the set of all admissible values of the measurement quantity. In order to design a robust measurement estimator, we use the worst-case criterion to select the best model structure. This natural idea is used in several domains (e.g. robust control [18] and production quality [7]).
We define a worst case distance :

$$\Delta m_j = \max_{m \in \mathbb{M}_j} ||\boldsymbol{m}_r - \boldsymbol{m}|| \tag{9}$$

where $\boldsymbol{m}_r$ is the reference measurement, assumed available. This information is usually obtained from experimental data or from the simulation of finite element code. The dependency of Δm in the experimental conditions $\boldsymbol{x}$ need not be made explicit since these are assumed fixed.
To choose a model according to this criterion, the model with the smallest Δm is picked :

$$f^* = \arg \min_{1 \le j \le n_f} \max_{m \in \mathbb{M}_j} ||\boldsymbol{m}_r - \boldsymbol{m}|| \tag{10}$$

5. Grooves dimensioning using Remote Field Eddy Current (RFEC) inspection

Remote field eddy current inspection is used for dimensining grooves that may occur in ferromagnetic pipes. An efficient method has been proposed in [3] to estimate the depth and the length of corrosion grooves from measurement of a pick-up coil signal phase at different positions close to the defect.
We perform a finite element calculation to obtain a parametric model of the physical phenomenon. In this approach, the dimensions of the groove are linked to the model parameters through a polynomial function. So, an estimate of the size of the groove may be computed.
Figure (2) illustrates a typical experimental apparatus used for groove dimensioning; the sensor is pushed inside the pipe and along with the coil position $\boldsymbol{x}$, the phase of the detector voltage $\boldsymbol{y}$ is acquired. The distance between exciter and detector coils is chosen so that the remote field condition is satisfied. We need a previous knowledge of the relation-

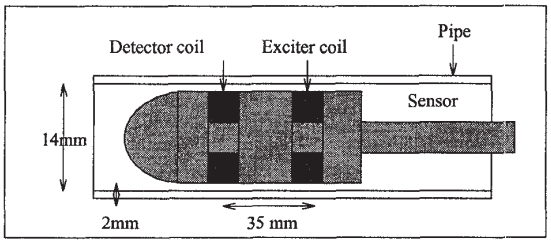

Figure 2. Schematic illustration of the sensor pipe.

ship between the groove parameters (length and depth) and the observed data (detector phase). A finite element modeling has been used to obtain such a relation.
Taking into account symmetry and the range of the data, several mathematical functions may be considered to approximate it.
Three competing nonlinear model structures f_1 ($p_1 = 2$), f_2 ($p_2 = 3$) and f_3 ($p_3 = 4$) are considered here :

$$\begin{aligned}
f_1(\boldsymbol{x},\boldsymbol{\theta}) &= \frac{\theta_1}{\theta_2 + (\boldsymbol{x} - L)^2} + \frac{\theta_1}{\theta_2 + (\boldsymbol{x} + L)^2} \\
f_2(\boldsymbol{x},\boldsymbol{\theta}) &= \theta_1(\arctan(\theta_2(\boldsymbol{x} + \theta_3 + L)) \\
&\quad - \arctan(\theta_2(\boldsymbol{x} - \theta_3 + L)) \\
&\quad + \arctan(\theta_2(\boldsymbol{x} + \theta_3 - L)) \\
&\quad - \arctan(\theta_2(\boldsymbol{x} - \theta_3 - L))) \\
f_3(\boldsymbol{x},\boldsymbol{\theta}) &= \theta_1(\exp(-\theta_2|\boldsymbol{x} - \theta_4 + L|^{\theta_3}) \\
&\quad + \exp(-\theta_2|\boldsymbol{x} + \theta_4 + L|^{\theta_3}))
\end{aligned}$$

where L is an experimental constant (half-distance between coils). Figure 3 shows an example of simulated observed

data computed by a finite element code and mean-square fitted curves for each models, for given groove parameters. These model structures have a different numbers of parameters. The Final Prediction Error (FPE) criterion developed by Akaike (1970), measures a compromise between model complexity and goodness of fit [10] :

$$J_{FPE}(f_j) = \frac{1+p_j/N}{1-p_j/N} \cdot \frac{1}{N} \sum_{k=1}^{N} \frac{1}{2}(f_j(x_k, \hat{\theta}) - y_k)^2$$

The obtained results are reported in Table 1. The model f_2 appears to be the best model in the FPE criterion sense (this does not still the same for another groove parameters). However, when estimating a measurement quantity, the structure classification order may change. The measurement goal is to estimate the size of the defect (length (l) and depth (d)) from the knowledge of detector phase. In a previous approach [4] [3], the measurement quantity is expressed as a function of the optimal parameters of the model. An algebraic form for g is chosen, a bilinear polynomial function of θ, which may be written as follows :

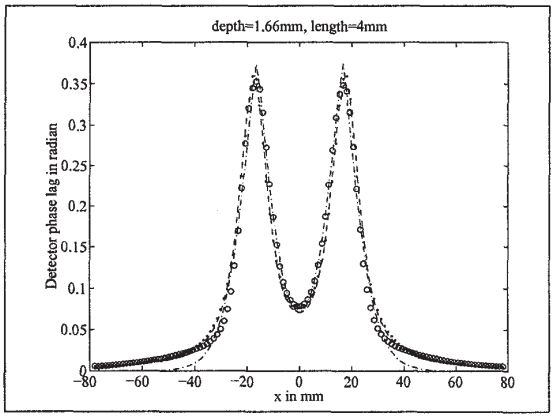

Figure 3. Observations (o) and mean-square fitted curves for the models 1 (..), 2 (- -) and 3 (.-.).

$$d = g_d(\theta) = \sum_{i,h} c^d_{ih}\theta_i\theta_h + \sum_i c^d_i\theta_i$$

$$l = g_l(\theta) = \sum_{i,h} c^l_{ih}\theta_i\theta_h + \sum_i c^l_i\theta_i$$

For several real defects dimensions, the coefficients (c^d_{ih}, c^d_i) and (c^l_{ih}, c^l_i) have been computed in the least-squares sense [3]. Thus for each candidate models an analytical form of $\boldsymbol{m}_j = [d_j, l_j]^\top (j = 1, 3)$ as a function of θ is obtained.
In deterministic framework the error is assumed to lie between some upper and lower bounds $(\underline{e}_k, \overline{e}_k)$. Thus each collected data from the detector belongs to an interval $y_k \in \mathbf{y}_k = [y_k + \underline{e}_k, y_k + \overline{e}_k]$. The prior feasible set $\mathbb{Y}$ for the model outputs is given by the cartesian product of these N intervals $\mathbb{Y} = \mathbf{y}_1 \times \ldots \times \mathbf{y}_N$ (figure 4).
For our application some decisions should be taken from

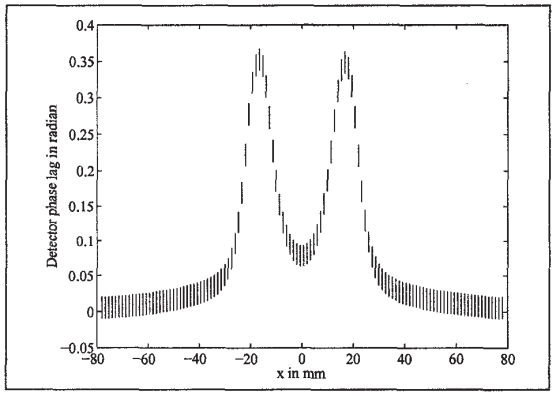

Figure 4. Data with error bars.

the numeric value of the defect size, we have to evaluate quantitatively how this data uncertainty affects the estimates. The error bounds are assumed unknown and the data uncertainty is an additional parameter to be identified. A simple bounds structure given by a constant bound e for all the data set is chosen $(\mathbf{y}_k = [y_k - e, y_k + e], k = 1 \ldots N)$. Estimation procedure should find the set of the parameters values and the additional parameter e. A recently developed MEBOE (Minimal error bound estimator) algorithm [9] estimates the smallest error bound so that the set $\mathbb{S}$ is not empty. Using the SIVIA algorithm the feasible set $\mathbb{S}$ for the parameters is determined. Figure 5 shows this uncertainty information on parameters for the first model. The set $\mathbb{M}$ of all admissible values of the measurement for each candidate models is computed using the image evaluation algorithm (Figure 6). In order to guarantee the quality of the measurement in the worst case, Δm for each candidate model is computed. f_2 is still the best structure in the second criterion sense. However, the selection criterion used here takes into account the final objective of modelling.

	f_1	f_2	f_3
J_{FPE}	$2.1\ 10^{-4}$	$1.9\ 10^{-4}$	0.0105
Δm	2.20	1.53	4.05

Table 1. Obtained results for both criterion.

6. Concluding remarks

The problem of nonlinear model selection has been considered for a measurement dedicated approach. In context

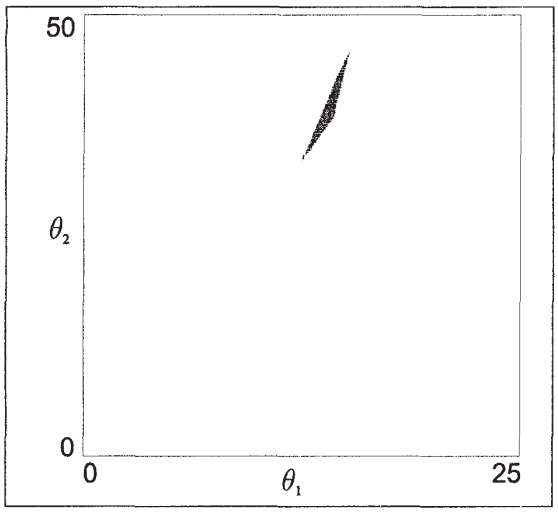

Figure 5. Set of feasible parameters for the first model structure.

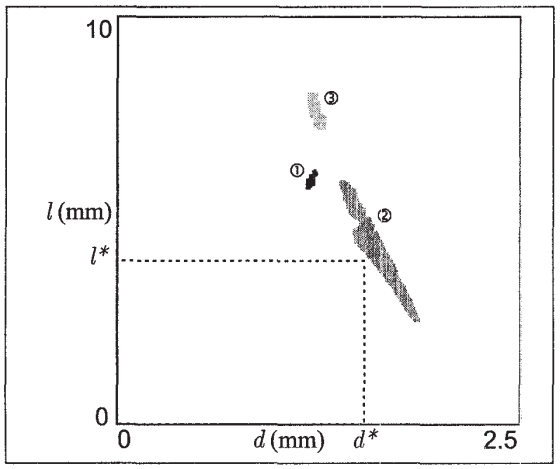

Figure 6. Image $\mathbb{M}$ of the set $\mathbb{S}$ for each candidate model.

of bounded-error estimation, a new selection procedure has been proposed. As each candidate structure yields a particular evaluation of the set of all admissible values of the measurement quantity, a worst case design is used to select the optimal model. The pessimism introduced by interval computations, as soon as there are many occurrences of the parameters in the formal expression of the model function, and the increasing of the complexity of the set description with the parameters number are the main limitations of such deterministic techniques. It might be of interest to use a different selection criterion and compare its results with the worst-case ones.

The choice between the stochastic approach presented in [1] [2] and the deterministic one treated in this paper, can be oriented by answering the two following questions :

- What kind of information is available ?

If the probability density function (PDF) of the additive random noise which corrupts the data is known, statistical methods are likely to be more appropriate. On the other hand, if noise specification are in terms of tolerance, the deterministic approach might be more suitable.

- What is the purpose of the measurement ?

If the goal is to guarantee the quality in the worst case, then minimax design should be preferred, which can be done in a deterministic framework such as that described here.

If we intend to get credible PDF for the measurement, the minimization of the PDFs distance with statistical tools seems more appropriate [1].

We thank M.E. Davoust and J. Oksman for letting us use their data.

References

[1] S. Brahim-Belhouari and G. Fleury. Choice of nonlinear models via Kullback Leibler information criterion - a measurement dedicated approach -. *Proc. Int. Symp. on Nonlinear Theory and its Application*, pp. 1141-1144, Swizterland, 1998.

[2] S. Brahim-Belhouari and G. Fleury. Probability distribustion in nonlinear estimation - a measurement dedicated approach -. *Proc. 9th IEEE Workshop on Statistical Signal and Array Processing*, pp. 395-398, Portland (USA), 1998.

[3] M.-E. Davoust, G. Fleury and J. Oksman. A parametric estimation approach for grooves dimensioning using remote field eddy current inspection. *To appear in Research in Nondestructive Evaluation*, 1999.

[4] G. Fleury. Non-intrusive time-of-flight flow-meter-parametric estimation and optimization. *Sensors and Actuators*, A(46-47):364-368, 1995.

[5] L. Jaulin and E. Walter. Set inversion via interval analysis for nonlinear bounded-error estimation. *Automatica*, 29(4):1053-1064, 1993.

[6] M. Milanese, J. P. Norton, H. Piet-Lahanier and E. Walter (ed.). Bounding approaches to system identification. *Plenum*, New York, 1996.

[7] M. Keramat. Analyse statistique et optimisation du rendement de fabrication des circuits électroniques.

Ph.D Dissertation, Université de Paris XI, Orsay, 1998.

[8] M. Kieffer, L. Jaulin and E. Walter. Guaranteed recursive nonlinear state estimation using interval analysis. *To appear in Proc. 37th IEEE CDC*, 1998.

[9] M. Kieffer. Estimation ensembliste par analyse par intervalles, application à la localisation d'un vehicule. *Ph.D Dissertation*, Université de Paris XI, Orsay, 1999.

[10] L. Ljung. System Identification, Theory of the User. *Prentice-Hall, Englewood cliffs*, 1987.

[11] R. Moore. Interval Analysis. *Prentice-Hall*, Englewood Cliffs, New Jersey, 1966.

[12] R. Moore. Parameter sets for bounded-error data. *Math. Comput. Simulation*, 34:113-119, 1992.

[13] J. P. Norton (ed.). Special issue on bounded-error estimation. *Int. J. of Adaptive Control and Signal Processing*; Issue 1 : 8(1), 1994; Issue 2 : 9(2), 1995.

[14] B. D. Olin, W. Q. Meeker. Application of statistical methods to Non-Destructive Evaluation. *Technometrics*, 38(2):95-112, 1996.

[15] H. Ratschek and J. Rokne. Computer Methods for the range of functions. *Halsted Press*, New York, 1984.

[16] L. Sandu, J. Oksman and G. Fleury. Information criteria for the choice of parametric functions for measurement. *Accepted in IEEE Trans. on Measurements*, 1999.

[17] E. Walter (ed.). Parameter Identification with error bound. *Math. Comput. Simulation*, 32(5&6), 1990.

[18] E. Walter, L. Pronzato. Identification of parametric models from experimental data. *Springer*, London, 1997.