MEMBERSHIP-SET ESTIMATION USING RANDOM SCANNING AND PRINCIPAL COMPONENT ANALYSIS

Karel KEESMAN

Department of Chemical Engineering, University of Twente, 7500 AE Enschede, Netherlands

A set-theoretic approach to parameter estimation based on the bounded-error concept is an appropriate choice when incomplete knowledge of observation error statistics and unavoidable structural model error invalidate the presuppositions of stochastic methods. Within this class the estimation of non-linear-in-the-parameters models is examined. This situation frequently occurs in modelling natural systems. The output error method proposed is based on overall random scanning with iterative reduction of the size of the scanned region. In order to overcome the problem of computational inefficiency, which is particularly serious when there is interaction between the parameter estimates, two modifications to the basic method are introduced. The first involves the use of principal component transformations to provide a rotated parameter space in the random scanning because large areas of the initial parameter space are thus excluded from further examination. The second improvement involves the standardization of the parameters so as to obtain an initial space with equal size extension in all directions. This proves to largely increase the computational robustness of the method. The modified algorithm is demonstrated by application to a simple three-parameter model of diurnal dissolved oxygen patterns in a lake.

1. Introduction

The problem of estimating model parameters is closely related to that of error structure characterization [1,2]. It must be recognized that our results will be highly influenced by our assumptions explicitly or implicitly made about the error structure. It is evident that correct conclusions about the properties of the estimates require a correct error structure characterization. Conventionally the errors are expressed in terms of stochastic uncertainty models. Due to incomplete information and the presence of structural model errors resulted from aggregation and obscurity of the process dynamics, a stochastic error approach is questionable, since many of these model errors are inherently not stochastic [2]. In addition information about the reliability of the estimates can only be deduced in an exact way if the error density function is Gaussian and the model is linear.

Because of these limitations a set-theoretic way of modelling uncertainty has been proposed using ellipsoidal-bounding [2–7] or orthotope-bounding algorithms [5,8–10]. The only assumption is that the uncertainty is bounded. The ellipsoidal-bounding algorithms are characterized by their (possibly recursive) solution of the intersection between a prior parameter set (Ω_p) and a conditional parameter set ($\Omega_{p|y}$), i.e.

$$\tilde{\Omega}_p = \Omega_p \cap \Omega_{p|y} \qquad (1)$$

where $\tilde{\Omega}_p$ is the posterior parameter set. The orthotope-bounding formulation on the contrary can be reduced to a linear programming problem. It is evident that estimation of the parameter

outer bounds by means of ellipsoids or orthotopes will give too pessimistic results from the point of view of parameter uncertainty. Recently, (exact) polytope-bounding algorithms have been proposed in order to represent the parametric uncertainty in a more reliable way [11–16]. However, the above-mentioned algorithms are generally restricted to models linear in their parameters, and an equation error model formulation. Only the algorithms presented in [8,10,15–16] can deal with an output error model formulation (non-linear optimization problem) if the signs of some unknown parameters are known in advance.

Within the class of set-theoretic estimation methods alternative (iterative) algorithms, based on an output error model formulation and discrete approximation of the parameter space, are also proposed (see [17–19]). These algorithms will result, by their use of random scanning techniques, in a finite set of feasible parameter vectors ("likelihood set"), which reflects parameter uncertainty in a consistent and discrete way.

In the majority of the literature random scanning algorithms have been used in optimization problems to avoid the problem of finding a local optimum, but their use is certainly not restricted to these applications. Fedra et al. [17] applied an overall random scanning method in order to find a set of feasible parameter vectors, on the basis of predefined system behaviour constraints. The main drawback of this method, however, is its computational inefficiency. In previous publications we proposed some modifications, such as incorporation of information from preceding data-analysis and fuzzy set-theoretic assumptions [19].

In this paper we discuss an additional and powerful method to improve the computational efficiency in the frequently observed situation that the elements of the estimated parameter vectors are correlated or fuzzily linearly interactive, when speaking in fuzzy set-theoretic terms (see [20]). Interaction of parameter estimates is a particularly threatening condition for the efficiency of overall random scanning methods, because large areas of the parameter space do not contribute to the likelihood set. Our purpose is to demonstrate that the incorporation of principal component analysis in iterative parameter estimation methods using random scanning and bounded-error approach largely improves the situation. The algorithm is illustrated on the basis of a simple water quality model with bounded noise.

2. Problem formulation

Consider the following class of models represented by a general first-order ordinary differential equation:

$$\dot{x}(p, t) = f[x(p, t), u(t)], \tag{2}$$

where $x(\cdot)$ = state vector, $u(\cdot)$ = input vector, p = parameter vector (dim(p) = n), f = vector function.

Output error methods are not restricted to this class of models. We adopted this class for the purpose of demonstration only. Extension to models of distributed systems is straightforward.

Because lack of information about detailed error structure characteristics and the presence of structural model errors (restrictions encountered in many practical problems of identification), the uncertainty is expressed in terms of bounds.

This uncertainty will affect each observation (y_k), so that the N observations are modelled as

$$y_k = g[x(p, t_k)] + e_k, \quad k = 1, \ldots, N, \tag{3}$$

where e_k is the observation error at the kth time instant (t_k) and g is a vector function.
The observation error will be expressed now in terms of a set, i.e. $e_k \in \Omega_e(k)$, where

$$\Omega_e(k) = \{ e_k \in \mathbb{R} : \underline{e}_k \leqslant e_k \leqslant \bar{e}_k \}, \quad k = 1, \ldots, N, \tag{4}$$

where $\underline{e}_k$ and $\bar{e}_k$ are the lower and upper bound on the kth observation error.
In this context also the observations belong to a set, the so-called "behaviour space" (Ω_y), which is defined as

$$\Omega_y = \{ y \in \mathbb{R}^N : y_k - \bar{e}_k \leqslant y_k \leqslant y_k - \underline{e}_k; \ k = 1, \ldots, N \}. \tag{5}$$

Let the prior parameter space be defined as

$$\Omega_p = \{ p \in \mathbb{R}^n : \underline{p} \leqslant p \leqslant \bar{p} \}, \tag{6}$$

where $\underline{p}$ and $\bar{p}$ are the vectors of lower and upper bounds on p.
The ultimate aim is then to identify the posterior parameter space ($\tilde{\Omega}_p$), where

$$\tilde{\Omega}_p = \{ p \in \mathbb{R}^n : y_k - g[x(p, t_k)] = e_k; \ p \in \Omega_p, e_k \in \Omega_e(k), k = 1, \ldots, N \} \tag{7}$$

such that *all* $p \in \tilde{\Omega}_p$ are possible estimates of the parameter vector, which together express the associated posterior parameter uncertainty.
If the estimation problem is solved satisfactorily then the model response space,

$$\tilde{\Omega}_y = \{ \tilde{y} \in \mathbb{R}^N : \tilde{y}_k = g[x(p, t_k)]; \ p \in \tilde{\Omega}_p, k = 1, \ldots, N \}, \tag{8}$$

i.e. a space containing all acceptable model outputs, is a subspace of the behaviour space.
The method that Fedra et al. [17] proposed to solve this problem can be summarized as follows:
  (1) definition of system behaviour in terms of constraints in the observation space (Ω_y),
  (2) definition of a prior parameter space (Ω_p),
  (3) Monte Carlo simulation of the model on the basis of uniformly sampled parameter vectors resulting in a set of model responses,
  (4) classification of the parameter vectors as feasible or non-feasible, which will result in the discrete parameter space $\tilde{\Omega}_p$.
However, the main drawback of this method is its computational inefficiency, especially when parameter estimate interactions are present. The problem then is how to identify $\tilde{\Omega}_p$ in a more effective way without loosing its nice property of computational robustness.

3. Parameter space adjustment algorithm

A method to identify $\tilde{\Omega}_p$ in a more effective way will make use of sequential updates of the parameter space in the space domain. Herein two kinds of transformations can be distinghuished, namely an affine transformation along the axes (expansion and/or translation), and a linear transformation by means of a parameter space rotation.

Analyses of the parameter space are performed in a standardized space (Ω_θ) on the basis of a priori parametric transformations of the original parameters, i.e.

$$\theta_j = 2 * (p_j - \underline{p}_j)/(\bar{p}_j - \underline{p}_j) - 1, \quad j = 1, \ldots, n. \tag{9}$$

This particular transformation is selected, because it has the attractive property that the centre of the prior parameter set Ω_θ equals $\mathbf{0}$ and its dispersion matrix is proportional to the unit matrix, if all possibilistic (instead of probabilistic) variables p_j are uniformly distributed and non-interactive. The dispersion matrix Σ is defined as,

$$\Sigma = \frac{1}{M} \sum_{i=1}^{M} \left(\theta_i \theta_i^{\mathrm{T}} \right), \tag{10}$$

where θ_i is a centralized parameter vector, and M is the cardinality of the finite set $\tilde{\Omega}_{\mathrm{p}}$. After a predefined number of simulations the set of feasible parameter vectors, containing information about the parameter space $\tilde{\Omega}_{\mathrm{p}}$, can be evaluated. First, information can be obtained about parameter space expansion and/or translation for a new initial parameter space [21]. But it is also clear that the position of the posterior parameter space may have been rotated with regard to the initial space (see [17–19]) due to multiple "collinearity". From eigenvalue decomposition of the dispersion matrix of the likelihood set (10) it is possible to deduce information about the rotation and to use it for the definition of a subsequent initial parameter space.

It is well known that principal component transformation is simply a rotation of the original axes of the coordinate system. So by determination of the principal axes a better initial parameter space can be specified. The principal axes $\Phi_1, \ldots, \Phi_n$, or eigenvectors of matrix Σ, are found by solving the characteristic equation

$$|\Sigma - \lambda_j I | = 0, \quad j = 1, \ldots, n, \tag{11}$$

for the eigenvalues λ_j, and then

$$(\Sigma - \lambda_j I)\Phi_j = \mathbf{0}. \tag{12}$$

Since the dispersion matrix is symmetrical these principal axes are orthogonal to each other or, in other words, they correspond to linearly independent directions which simplifies further analysis.

The parameter space rotation is performed according to

$$\tau^{\mathrm{T}} = \theta^{\mathrm{T}}\Phi, \tag{13}$$

where τ is the parameter vector in rotated axes system, and Φ is the matrix containing the normalized eigenvectors of Σ (i.e. $\Phi = [\Phi_1, \ldots, \Phi_n]$, where $\| \Phi_j \|_2 = 1$ for $j = 1, \ldots, n$).

The normalized components of the columns of Φ are also the direction cosines of the transformed axes relative to the original axes, so that it is clear that the principal components transformation causes only a rotation.

In this rotated coordinate system new parameter ranges must be specified on the basis of available information from the preceding simulations. Because of the random sampling, the set of feasible parameter vectors will not reflect the potential parameter ranges. With this in mind the outer bounds of the new parameter space after rotation are expanded according to

$$\begin{aligned} \bar{\tau}^*_{i+1} &= \bar{\tau}_i + \beta(\bar{\tau}_i - \underline{\tau}_i), \\ \underline{\tau}^*_{i+1} &= \underline{\tau}_i - \beta(\bar{\tau}_i - \underline{\tau}_i), \end{aligned} \tag{14}$$
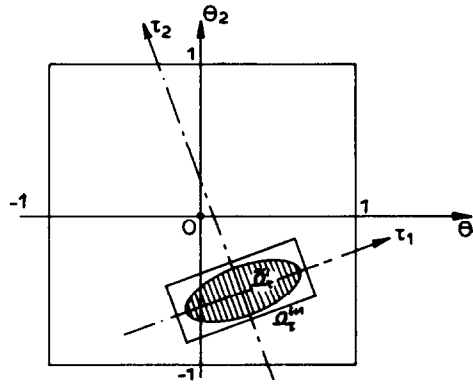
Fig. 1. Geometrical representation of the transformed parameter spaces.

where $\bar{\tau}_i = \max\{[\boldsymbol{\theta}^T\boldsymbol{\Phi}]^T\}$, and $\underline{\tau}_i = \min\{[\boldsymbol{\theta}^T\boldsymbol{\Phi}]^T\}$ are the upper and lower values of the projections on the principal axes of all available feasible parameter vectors at the ith iteration, and $\underline{\tau}_{i+1*}$, $\bar{\tau}_{i+1*}$ contain the new bounds of Ω_τ^{i+1} for the next iteration. At this point we only have to choose a proper expression for the expansion factor β. From experiences we suggest

$$\beta = 1/M, \quad M \geqslant n + 1. \tag{15}$$

Note that the second iteration can only start when at least $(n + 1)$ feasible parameter vectors have been found. It is recommended to specify the number of simulations for each iteration larger than say 10 times the number of parameters.

To increase the computational efficiency subsequent realizations of the parameter vector will be sampled from Ω_τ^{i+1}. The initial physically interpretable parameter ranges remain unchanged, i.e.

$$\underline{p}_{i+1}^* = \underline{p}_i^*, \qquad \bar{p}_{i+1}^* = \bar{p}_i^*, \tag{16}$$

where $\underline{p}^*$ and $\bar{p}^*$ are vectors containing the potential lower and upper parameter bounds in the original space. The aforementioned procedure can be represented geometrically in the following way (see in *Fig. 1*).

Formally the parameter space adjustment algorithm proposed can be summarized in the following steps:

(1) center and scale the original parameter ranges obtained from literature, experimentation etc., which will result in Ω_θ (see (9)).

(2) identify feasible parameter vectors on the basis of a predefined number of simulations and a priori information from preceding iteration. In the first iteration, the likelihood set can be found using space delimitation [22] and fuzzy set-theoretic information [19].

(3) rotate the standardized coordinate system on the basis of principal components analysis of the feasible parameter vectors. The posterior parameter space in the rotated coordinate system ($\tilde{\Omega}_\tau^i$) is then aligned to the principal axes.

(4) extend the appearing parameter bounds in the rotated coordinate system, according to (14), to avoid losses of information.

(5) if M is sufficiently large, then stop else continue with step (2).

Parameter vectors for subsequent simulations are sampled now from Ω_τ^{i+1}, followed by a reverse transformation using the inverse form of (13) and (9), to produce original parameter vectors. In order to obtain a more uniform sampling pattern, the Monte Carlo sampling scheme has been replaced by the so-called Latin hypercube sampling scheme [23].

4. Application

The algorithm proposed will be applied to the estimation of parameters of a simple water quality model, that describes the diurnal dissolved oxygen variation in a well mixed lake. The data (*Fig. 2*) have been obtained from lake " De Poel en 't Zwet", a lake situated in the western part of the Netherlands, for the period 21–30 April 1983.

The rate of change of dissolved oxygen is determined by re-aeration exchange with the atmosphere, by photosynthetic production from algae and water plants, and by consumption due
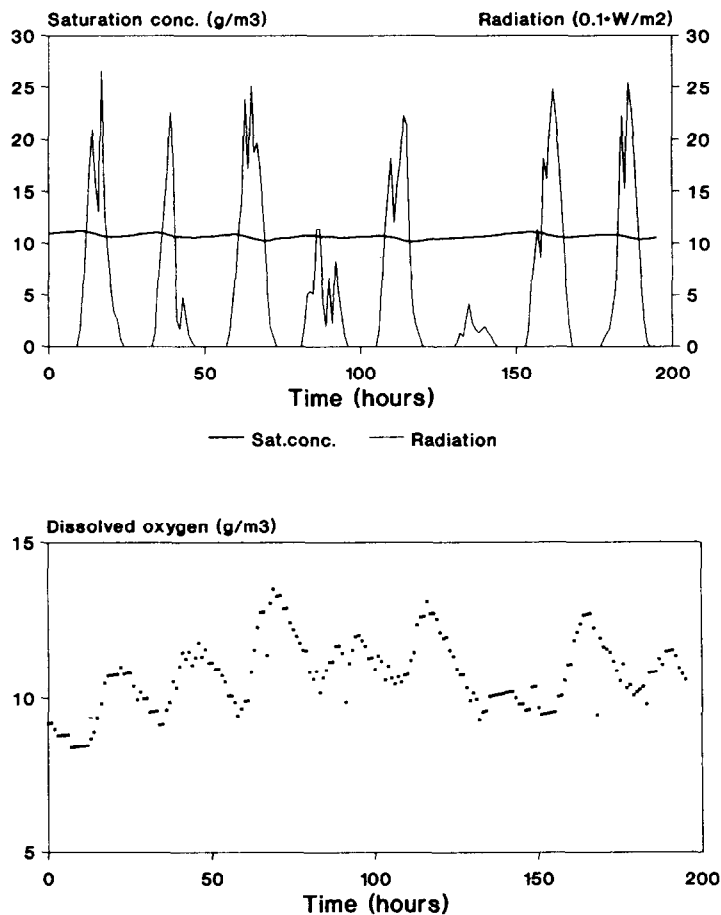

Fig. 2. Input and output data for Lake "De Poel en 't Zwet" (21–30 April 1983).

to respiration, biodegradation and sediment processes, i.e.

$$\dot{c}(t) = K_r[\text{Cs}(t) - c(t)] + aI(t) - R, \tag{17}$$

where $c(\cdot)$ = dissolved oxygen concentration (g/m^3), $\text{Cs}(\cdot)$ = saturation concentration (g/m^3), $I(\cdot)$ = radiation (W/m^2), K_r = re-aeration coefficient (1/d), a = photosynthetic rate coefficient (g/mdW), R = sink term (g/m^{3}d).

From preceding investigations of this model and associated data (see [19]) the following initial information is available:

(1) observation error is assumed to belong to a set (4), where $\underline{e}_k = -1.5$ g/m^3; $\bar{e}_k = 1.5$ g/m^3 for $k = 1, \ldots, 196$.

(2) observation at $k = 169$ is unreliable, i.e. $\underline{e}_k = -\infty$; $\bar{e}_k = \infty$ for $k = 169$.

(3) initial parameter space:

$$\Omega_p = \begin{cases} K_r \in [0.5, 1.8], \\ a \in [0.03, 0.07], \\ R \in [0.9, 3.5]. \end{cases} \tag{18}$$

On the basis of this information three successive iterations (125 simulations) are performed. For each iteration the parameter vector are sampled randomly from uniform membership functions in the stepwise adjustable parameter space. The result is that from the first and second iteration only 5 and 19 parameter vectors out of 50 simulations are accepted, while in the last run 21 vectors are accepted.

In *Fig. 3* the positions of all feasible parameter vectors with respect to the initial parameter space (18) are presented graphically. Each separate plot represents a projection of these parameter vectors onto one of the faces of the three-dimensional box. In these plots the
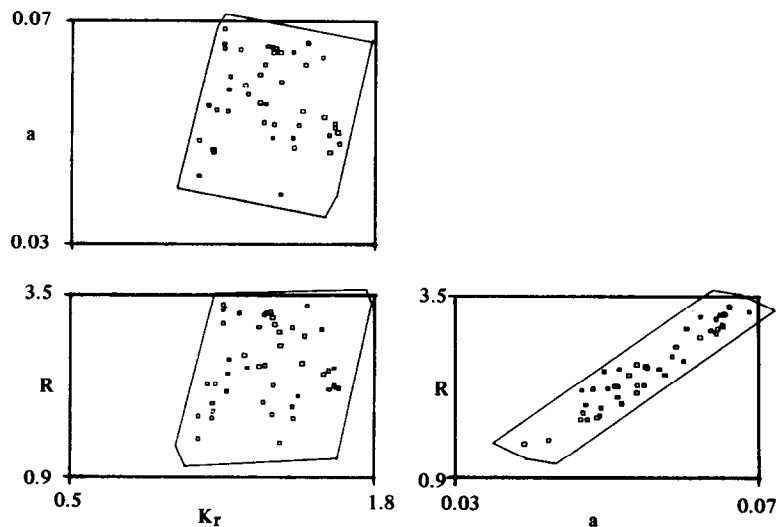

Fig. 3. Geometrical projection of the 45 feasible parameter vectors and the initial space for the last iteration onto the faces of the initial space.

deformation of the predefined parameter space induced by the procedure, is also reflected. Without application of parameter space translation and/or rotation the whole set of sampled parameter vectors would be situated randomly in the initial parameter space, which is evidently not the case here. It is therefore worth noting that the percentage of feasible vectors increased from 10% for the first iteration to 84% for the last iteration, reflecting the improved computational efficiency. Note that in the K_r–a and K_r–R planes the translation played a major role in the sampling of parameter vectors, while in the a–R plane the effect of space rotation is most visible. It appears that the a posteriori identified interaction between the parameter estimates (see Fig. 3) is consistent with the linearly dependency of the mutual parameters from preceding data-analysis i.e. $R = -0.07K_r + 51.5a - 0.17$ (see [19]).

Although the principal axes in the standardized and rotated coordinate system are orthogonal, this orthogonality is lost in the original coordinate system after inverse transformation according to (9). So this transformation causes deformation of the orthotopes in the original coordinate system (see Fig. 3), which also means that physical interpretation of the parameter combinations in this coordinate system determining the principal directions is not straightforward.

To emphasize the advantage of a set-theoretic way of modelling uncertainty, it must be noted that the number of available observations in the example could have been reduced without violating assumptions with respect to the uncertainties. Therefore, a membership-set estimation method, unlike a least-squares or maximum likelihood method, allows still the evaluation of the posterior parametric uncertainty in a reliable way from a small number of observations.

5. Conclusions

Within the context of set-theoretic uncertainty modelling the iterative method using random scanning appears to be simple and efficient. Incorporation of information from principal component analysis improves the computational efficiency significantly if parameter interactions are present. Inclusion of a transformation to scale the original parameter space avoids numerical problems when dealing with parameter ranges of different magnitudes.

Acknowledgment

References

[1] P. Eykhoff, System Identification — Parameter and State Estimation (Wiley, Chichester, 1974).
[2] F.C. Schweppe, Uncertain Dynamic Systems (Prentice-Hall, London, 1973).
[3] D.P. Bertsekas and I.B. Rhodes, Recursive state estimation for a set-membership description of uncertainty, IEEE Trans. Automat. Control AC-16 (2) (1971) 117–128.
[4] E. Fogel and Y.-F. Huang, On the value of information in system identification—bounded noise case, Automatica 18 (2) (1982) 229–238.
[5] G. Belforte and B. Bona, An improved parameter identification algorithm for signals with unknown-but-bounded errors, in: Proc. of 7th IFAC/IFORS Symp. Identification and System Parameter Estimation (Pergamon Press, Oxford, 1985) 1507–1512.

[6] J.P. Norton, Identification and application of bounded-parameter models, *Automatica* 23 (4) (1987) 497–507.

[7] J.P. Norton, Identification of parameter bounds for ARMAX models from records with bounded noise, *Internat. J. Control* 45 (2) (1987) 375–390.

[8] G. Belforte and M. Milanese, Uncertainty intervals evaluation in presence of unknown but bounded errors. Nonlinear families of models, in: *Proc. IASTED Int. Symp. Modelling, Identification and Control*, Davos (1981) 75–79.

[9] M. Milanese and G. Belforte, Estimation theory and uncertainty intervals evaluation in presence of unknown-but-bounded errors, linear families of models and estimators, *IEEE Trans. Automat. Control* AC-27 (2) (1982) 408–414.

[10] E. Walter, H. Piet-Lahanier and J. Happel, Estimation of non-uniquely identifiable parameters via exhaustive modeling and membership set theory, *Math. Comput. Simulation* 28 (1986) 479–490.

[11] V.V. Kapitonenko, Minimax identification in the case of uncertain original information, *Automat. Remote Control* (1982) 166–169.

[12] H. Piet-Lahanier and E. Walter, Exact recursive characterization of feasible parameter sets in the linear case, *Math. Comput. Simulation* 32 (5&6) (1990) 495–504 (this issue).

[13] V. Broman and M.J. Shensa, A compact algorithm for the intersection and approximation of N-dimensional polytopes, *Math. Comput. Simulation* 32 (5&6) (1990) 469–480 (this issue).

[14] S.H. Mo and J.P. Norton, Fast and robust algorithm to compute exact polytope parameter bounds, *Math. Comput. Simulation* 32 (5&6) (1990) 481–493 (this issue).

[15] T. Clément and S. Gentil, Reformulation of parameter identification with unknown-but-bounded errors, *Math. Comput. Simulation* 30 (1988) 257–270.

[16] T. Clément and S. Gentil, Recursive membership set estimation for output-error models, *Math. Comput. Simulation* 32 (5&6) (1990) 505–513 (this issue).

[17] K. Fedra, G. van Straten and M.B. Beck, Uncertainty and arbitrariness in ecosystems modelling: a lake modelling example, *Ecol. Model.* 13 (1981) 87–110.

[18] K.J. Keesman, *A set-membership approach to the identification and prediction of ill-defined systems: application to a water quality system*, Ph.D. Thesis, University of Twente, The Netherlands (1989).

[19] K.J. Keesman and G. van Straten, Modified set theoretic identification of an ill-defined water quality system from poor data, in: *Proc. IAWPRC Symp. Systems Analysis in Water Quality Management* (Pergamon Press, Oxford, 1987) 297–308.

[20] D. Dubois and H. Prade, Additions of interactive fuzzy numbers, *IEEE Trans. Automat. Control* AC-26 (4) (1981) 926–936.

[21] P.G. Whitehead and G.M. Hornberger, Modeling algal behaviour in the river Thames, *Water Res.* 18 (8) (1984) 945–953.

[22] G. van Straten, Analytical methods for parameter-space delimitation and application to shallow-lake phytoplankton-dynamics modeling, *Appl. Math. and Comput.* 17 (1985) 459–482.

[23] M.D. McKay, R.J. Beckman and W.J. Conover, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics* 21 (1979) 239–245.