INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Identifying intrinsic variability in multivariate systems through linearised inverse methods

Gilles Celeux — Agnès Grimaud — Yannick Lefèbvre — Étienne de Rocquigny

N° 6400

December 2007

Thème COG

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--6400--FR+ENG

Identifying intrinsic variability in multivariate systems through linearised inverse methods

Gilles Celeux* , Agnès Grimaud* , Yannick Lefèbvre† , Étienne de Rocquigny †

Thème COG — Systèmes cognitifs
Projets SELECT

Rapport de recherche n° 6400 — December 2007 — 23 pages

Abstract: A growing number of industrial risk studies include some form of treatment of the numerous sources of uncertainties affecting the conclusions; in the uncertainty treatment framework considered in this paper, the intrinsic variability of the uncertainty sources is modelled by a multivariate probability distribution. A key difficulty traditionally encountered at this stage is linked to the highly-limited sampling information directly available on uncertain input variables. A possible solution lies in the integration of indirect information, such as data on other more easily observable parameters linked to the parameters of interest through a well-known physical model. This leads to a probabilistic inverse problem: The objective is to identify a probability distribution, the dispersion of which is independent of the sample size since intrinsic variability is at stake. To limit to a reasonable level the number of (usually large CPU-time consuming) physical model runs inside the inverse algorithms, a linear approximation in a Gaussian framework are investigated in this paper. First a simple criterion is exhibited to ensure the identifiability of the model (i.e. the existence and unicity of a solution to the inverse problem). Then, the solution is computed via EM-type algorithms taking profit of the missing data structure of the estimation problem. The presentation includes a so-called ECME algorithm that can be used to overcome the possible pathology of slow convergence which affects the standard EM algorithm. Numerical experiments on simulated and real data sets highlight the good performances of these algorithms, as well as some precautions to be taken when using this approach.

Key-words: Uncertainty Modelling, Intrinsic Variability, Linear Approximation, Identifiability, EM and ECME algorithms

* INRIA Futurs, Projet SELECT, Université Paris-Sud
† EDF, R&D

Unité de recherche INRIA Futurs
Parc Club Orsay Université, ZAC des Vignes,
4, rue Jacques Monod, 91893 ORSAY Cedex (France)
Téléphone : +33 1 72 92 59 00 —Télécopie : +33 1 60 19 66 08

Idendification de la variabilité intrinsèque de systèmes multivariés par des méthodes inverses linéarisées

Résumé : Un nombre croissant d'études de risques industriels prend en compte différentes manières de traiter les nombreuses sources d'incertitudes affectant les conclusions. Dans cet article, la variabilité intrinsèque des sources d'incertitude est modelisée par une loi de probabilité multivariée. Une des principales difficultés souvent rencontrée est la disponibilité limitée d'informations sur les lois de probabilités des variables d'entrée incertaines. Une solution possible est l'intégration de renseignements indirects, comme des données sur d'autres paramètres plus facilement observables reliés aux paramètres d'intérêt par un modèle physique connu. Cela conduit à un problème probabiliste inverse : lorsque la variabilité intrinsèque est en jeu, l'objectif est d'identifier une loi de probabilité dont la dispersion est indépendante de la taille de l'échantillon. Pour limiter de façon raisonnable le nombre d'appels au modèle physique (souvent exigeants en temps de calcul CPU) lors de l'utilisation des algorithmes inverses, nous considérons dans cet article une approximation linéaire dans un cadre gaussien. Dans un premier temps, nous proposons un critère simple pour garantir l'identifiabilité du modèle (c'est-à-dire l'existence et l'unicité d'une solution du problème inverse). Ensuite, cette solution est calculée en utilisant des algorithmes de type EM, permettant de prendre en compte la structure de données manquantes du problème d'estimation. Pour accélérer l'algorithme EM, souvent lent, nous proposons une variante de la famille des algorithmes ECME. Les résultats numériques sur des données simulées et réelles montrent les bonnes performances de ces algorithmes, et nous permettent la mise en évidence de quelques précautions à prendre.

Mots-clés : Modélisation d'incertitudes, variabilité intrinsèque, approximation linéaire, identifiabilité, algorithmes EM et ECME

1 Introduction

A growing number of industrial risk studies include some form of treatment of the numerous sources of uncertainties affecting the conclusions. In the energy sector, such uncertainty analyses are for instance carried out in environmental studies (flood protection, effluent control, etc.), or in nuclear safety studies involving large scientific computing (thermo-hydraulics, mechanics, neutronics etc.). For most of those applications, at least partially probabilistic modelling of the uncertainties is considered. Then, apart from the important uncertainty propagation issues in the context of often complex and high CPU-time demanding scientific computing, one of the key issues regards the quantification of the sources of uncertainties. The problem is to choose reliable statistical models for the input variables such as uncertain physical properties of the materials or industrial process or natural random phenomena (wind, flood, temperature, etc.).

A key difficulty, traditionally encountered at this stage, is linked to the highly-limited sampling information directly available on uncertain input variables. An industrial case-study can largely benefit from two strategies: (a) integrate expert judgment, such as likely bounds on physical intervals or more elaborate probabilistic information, or (b) integrate indirect information, such as data on other, more easily observable, parameters that are linkable to the uncertain variable of interest by a physical model. Methods for (b) demand using of probabilistic inverse methods since the recovering of indirect information involves generally the inversion of a physical model. Roughly speaking, this inversion transforms the information into a virtual sample of the variable of interest, before applying to it standard statistical estimation. This field is acknowledged to have quite a large industrial potential in the context of rapid growth of industrial monitoring and data acquisition systems. One of the big issues in practice is to limit to a reasonable level the number of (usually large CPU-time consuming) physical model runs inside the inverse algorithms.

This paper concentrates on the situation where there is an *irreducible* uncertainty or variability in the input parameters of a physical model. Inverse probabilistic techniques in data assimilation or parameter identification are not new (e.g. Beck 1977 or Tarantola 1987). It may not be until quite recently that full probabilistic inversion was considered: The distribution of intrinsic (or irreducible, aleatory) input uncertainty is searched. Classical data assimilation or parameter identification techniques involve the estimation of input parameters (or initial conditions, for example in meteorology) that are unknown but physically fixed; this is naturally accompanied by estimation uncertainty for which variance may be computed. However such an estimation uncertainty happens to be purely epistemic or reducible, in the sense that it will decrease with the injection of larger observation samples. This is not satisfactory in the cases which motivated the present research, whereby fluid mechanical systems do evidence intrinsically variable input. Input values do not only suffer from lack of knowledge, but they also physically vary from one experience to another in the cases of flow observations taken at various times of the year on a fluctuating riverbed, or turbulent fluid flow measurements in a nuclear reactor with changing heat exchange properties. In those cases, the size of observation samples should impact the estimation accuracy, but it should not reduce the variance of the physical fluctuations.

Mathematically, observations are modelled with a vector of physical variables y that are connected to uncertain inputs x through a deterministic (and supposedly quite well-known) physical model $y = H(x, d)$. As a clear difference to classical parameter identification x is not supposed to have a fixed, albeit unknown physical value: It will be modelled a random variables taking different realisations for each observations. The purpose of the algorithm will be to estimate its probability distribution function instead of its point value. On the other hand, d stands for fixed inputs that may represent (i) variables under full control (e.g. experimental conditions), or (ii) uncertainties affecting some model inputs that are considered to be negligible or of secondary importance.

More specifically the following model is considered

$$Y_i = H(X_i, d_i) + U_i, \quad 1 \le i \le n \tag{1}$$

where

- (Y_i) in $\mathbb{R}^p$ denotes the data vectors,
- H denotes a known function from $\mathbb{R}^q$ to $\mathbb{R}^p$,
- (X_i) in $\mathbb{R}^{q_1}$ denotes non observed random data, assumed independent and identically distributed (i.i.d.) and with distribution $\mathcal{N}(m, C)$,
- (d_i) denotes fixed observed variables, with dimension q_2,
- (U_i) denotes measurement-model errors, assumed i.i.d. with distribution $\mathcal{N}(0, R)$, R being known.Variables (X_i) and (U_i) are assumed to be independent.

The aim is to estimate the parameter $\theta = (m, C)$. Before entering the technical part of this problem, a few remarks are to be highlighted.

Remark 1 *Two different interpretations follow according to the status given to the random vector. In a first interpretation – which is the one adopted in this paper – the distribution of U_i is supposed to be known, representing for instance a sensor measurement error deviating from an assumingly reliable model. In a second interpretation, the distribution of U_i has to be calibrated to represent the unknown measurement model error, for instance because the model is less reliable, and may deviate from data for other reason than sensor fluctuations. Hence, R has to be estimated as well.*

Remark 2 *As already mentioned, the present framework differs from data assimilation, as presented for instance in Talagrand (1997), in the fact that the uncertain parameters are intrinsically considered as stochastic. In data assimilation algorithms (BLUE, 3dVar, 4dVar, Kalman-filters, etc.), the dispersion of X tends to zero when the sample size n tends to infinity, as discussed in Mahé and de Rocquigny (2005): The probability distribution obtained via these algorithms quantifies the lack of knowledge, and not an intrinsically random behaviour.*

Remark 3 *The framework considered here addresses intrinsic variability in a somewhat restricted case since the dispersion is assumed to be Gaussian.*

Several difficulties may appear. One issue is the identifiability of the parameters of the random variables (X_i), which can be out of reach if the available information is not rich enough (for example because the non observable variables dimension is larger than the dimension of the observed outputs). Another issue is the time needed to compute the physical function H, since H is often the result of a complex code. A possible answer, used in the following, is to linearise the model around a fixed value x_0. The considered model becomes

$$Y_i = H(x_0, d_i) + J_H(x_0, d_i)(X_i - x_0) + U_i, \quad 1 \leq i \leq n \tag{2}$$

where $J_H(x_0, d_i)$ is the Jacobian matrix of the function H in x_0, with dimension $p \times q_1$.

This linearisation method has some drawbacks (for example how to choose the linearisation point, approximation errors, etc.) but it does not involve computational or numerical difficulties (other than the initial computation of the Jacobian matrix) and give reasonable results in many cases.

The data (X_i) being non observed, the estimation problem is a missing data structure problem that can be solved with an EM- type algorithm. The EM algorithm was developed by Dempster *et al.* (1977) and is an iterative algorithm with two steps: E Step (for Expectation) and M Step (for Maximisation). An extension to this algorithm, called ECME algorithm, for Expectation-Conditional Maximisation Either, was proposed by Liu and Rubin (1994). This algorithm which can be expected to converge more rapidly than EM has also been employed in the present study. For model (2), the ECME algorithm has been proposed independently by De Crecy (1996) under the name "Circe Method".

The paper is organised as follows. In Section 2, a criterion is proposed to ensure the identifiability of the linear model (2). In Section 3, the EM and an ECME algorithms are described and compared to estimate the linear model at hand. In Section 4, the two algorithms are applied to two examples: The first example consists of simulated data obtained from a quite simplified flooding model; the second example is a real-case thermo-hydraulics study carried out in Électricité De France (EDF). A brief discussion ends the paper.

2 Identifiability of the model

In this section a criterion is proposed to ensure the identifiability of the linear model (2), rewritten here in a simplified form:

$$Y_i = H_i X_i + U_i, \quad \text{for } i = 1, ..., n \tag{3}$$

with

- Y_i an observed vector in $\mathbb{R}^p$,

- H_i a known matrix with dimension $p \times q$,
- X_i a random vector with dimension q, the X_is are assumed i.i.d. with distribution $\mathcal{N}(\mu, \Sigma)$, (μ, Σ) being unknown.
- U_i a vector with dimension p, and the U_is are assumed to be i.i.d. errors with distribution $\mathcal{N}(0, R)$, R being known.

Then, Y_i is a Gaussian vector with $E(Y_i) = H_i\mu$ and $\mathrm{Var}(Y_i) = H_i \Sigma H_i^T + R$. Further notation is introduced:

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \mathbf{H} = \begin{pmatrix} H_1 \\ \vdots \\ H_n \end{pmatrix}, \mathbf{\Gamma} = \begin{pmatrix} H_1 & & \\ & \ddots & \\ & & H_n \end{pmatrix}, \mathbf{\Sigma} = \begin{pmatrix} \Sigma & & \\ & \ddots & \\ & & \Sigma \end{pmatrix} \text{ and}$$
$$\mathbf{R} = \begin{pmatrix} R & & \\ & \ddots & \\ & & R \end{pmatrix}.$$

Hence, Y has dimension np, while H, Γ, $\mathbf{\Sigma}$ and $\mathbf{R}$ are matrices with respective dimensions $np \times q$, $np \times nq$, $nq \times nq$ and $np \times np$. Thus, Y is a Gaussian vector with $E(Y) = H\mu$ and $\mathrm{Var}(Y) = \Gamma \mathbf{\Sigma} \Gamma^T + \mathbf{R}$.

The model is said to be identifiable if and only if for all (μ_1, μ_2) and (Σ_1, Σ_2)

$$\begin{cases} H_i(\mu_1 - \mu_2) = 0 \,\forall\, i \\ H_i(\Sigma_1 - \Sigma_2)H_i^T = 0 \,\forall\, i \end{cases} \Rightarrow \begin{cases} \mu_1 = \mu_2 \\ \Sigma_1 = \Sigma_2, \end{cases}$$

or equivalently if and only if for all (μ_1, μ_2) and (Σ_1, Σ_2)

$$\begin{cases} H(\mu_1 - \mu_2) = 0 \\ \Gamma(\Sigma_1 - \Sigma_2)\Gamma^T = 0 \end{cases} \Rightarrow \begin{cases} \mu_1 = \mu_2 \\ \Sigma_1 = \Sigma_2. \end{cases}$$

Proposition 1 *Assuming $q \leq np$, Model (3) is identifiable if and only if rank$(H) = q$.*

Proof:

- First, let assume that $\mathrm{rank}(H) = q$.
  1. In this case, H is injective and $H(\mu_1 - \mu_2) = 0 \Rightarrow \mu_1 = \mu_2$.
  2. It is now proved that Γ is full rank $\Leftrightarrow \Gamma^T\Gamma$ is inversible. Since $\Gamma^T\Gamma$ is a $q \times q$ matrix, $\Gamma^T\Gamma$ is inversible if and only if $\Gamma^T\Gamma$ is injective. So, let v be in $\mathrm{Ker}(\Gamma^T\Gamma)$. Then $\Gamma^T\Gamma v = 0 \Rightarrow v^T\Gamma^T\Gamma v = 0 \Rightarrow ||\Gamma v|| = 0 \Rightarrow \Gamma v = 0 \Rightarrow v = 0$ since Γ is injective. Hence, $\mathrm{Ker}(\Gamma^T\Gamma) = \{0\}$ and thus $\Gamma^T\Gamma$ is inversible.
     Conversely, let assume that $\Gamma^T\Gamma$ is inversible. Let v be in $\mathrm{Ker}(\Gamma)$. $\Gamma v = 0 \Rightarrow \Gamma^T\Gamma v = 0 \Rightarrow v = 0$ since $\Gamma^T\Gamma$ is injective. Hence $\mathrm{Ker}(\Gamma) = \{0\}$ and Γ is full rank.

3. Hence $\Gamma(\Sigma_1 - \Sigma_2)\Gamma^T = 0 \Rightarrow \Gamma^T\Gamma(\Sigma_1 - \Sigma_2)\Gamma^T\Gamma = 0$. Since $\Gamma^T\Gamma$ is inversible, it implies that $\Sigma_1 - \Sigma_2 = 0$.

- Conversely, if the model is identifiable, for all (μ_1, μ_2), $H(\mu_1 - \mu_2) = 0$ implies $\mu_1 = \mu_2$. Hence $\mathrm{Ker}(H) = \{0\}$ and $\mathrm{rank}(H)$ is equal to q.

This proposition provides an easy criterion to check at the very beginning of the study if the inverse problem at hand has a solution. Practically, if the model is not identifiable, several solutions can be envisaged, all based on the use of expert or engineering judgement. If expert knowledge is rich enough, a first solution – that will be illustrated in the flooding example – consists of fixing some of the parameters of the probability distribution of the random variables X; this may be tedious, especially when the parameter to be fixed are standard deviations, generally far more difficult to evaluate a priori than mean values. A simpler approach – illustrated in the real-case thermo-hydraulics example – is to assume some of the Xs are fixed (equal to a mean value determined by experts), thus neglecting their stochastic nature. But as we will see, this simplification has a price: It can lead to overestimate standard deviations for the remaining parameters.

Remark 4 *(Empirical Identifiability) In practice the condition $q \leq np$ mentioned in Proposition 1, is not sufficient to ensure that enough data is available for estimation. Consider that $q = p = 1$; at least $np = 2$ observations are necessary in order to be able to estimate m and C. With $np = 1$ observation, only the mean can be estimated. Hence, in the Gaussian case pratically, a supplementary condition could be added to ensure that the estimation is feasible. For example $n_0 q \leq np$, with n_0 greater than 2.*

3 EM and ECME algorithms

In the context of missing data, an estimation method largely acknowledged as relevant is the EM algorithm (Dempster *et al.*, 1977), which is an iterative algorithm with two steps: E Step (Expectation) and M Step (Maximisation). An extension devoted to accelerate the EM algorithm, which is known to often encounter slow convergence situations, is the ECME (Expectation-Conditional Maximisation Either) algorithm of Liu and Rubin (1994).

In this paper, the focus is placed on the linearised model (2)

$$Y_i = H(x_0, d_i) + J_H(x_0, d_i)(X_i - x_0) + U_i, \quad 1 \leq i \leq n$$

where $J_H(x_0, d_i)$ is the Jacobian matrix of H in x_0, with dimension $p \times q_1$.

The following notation is used: $Z_i = (Y_i^T, X_i^T)^T$, $i = 1, \ldots, n$ is denoting the completed data. We denote $h_i = H(x_0, d_i)$; $J_i = J_H(x_0, d_i)$ and consequently $Y_i = h_i + J_i(X_i - x_0) + U_i$. Moreover, we denote $A_i = Y_i - h_i - J_i(m - x_0)$; $B_i = CJ_i^T$; $V_i = J_i C J_i^T + R$.

Some preliminary results are useful to define the EM and ECME algorithms for Model (2).

- The variable Y_i is a linear combination of Gaussian variables with mean and variance

$$\mathrm{E}(Y_i) = h_i + J_i(m - x_0), \quad \mathrm{Var}(Y_i) = J_i C J_i^T + R. \tag{4}$$

- Z_i is a Gaussian vector with

$$\mathrm{E}(Z_i) = \begin{pmatrix} h_i + J_i(m - x_0) \\ m \end{pmatrix} \quad \text{and } \mathrm{Var}(Z_i) = \begin{pmatrix} J_i C J_i^T + R & J_i C \\ C J_i^T & C \end{pmatrix}. \tag{5}$$

- $X_i|Y_i$ is also a Gaussian vector (using a theorem on conditional distributions of a Gaussian vector, see for instance Saporta, 1990 p. 88) with

$$\begin{aligned} \mathrm{E}(X_i|Y_i = y_i) &= m + C J_i^T (J_i C J_i^T + R)^{-1}(y_i - h_i - J_i(m - x_0)) \\ &= m + B_i V_i^{-1} A_i & (6) \\ \mathrm{Var}(X_i|Y_i = y_i) &= C - C J_i^T (J_i C J_i^T + R)^{-1} J_i C \\ &= C - B_i V_i^{-1} B_i^T. & (7) \end{aligned}$$

3.1 EM algorithm

At iteration $k+1$, the EM algorithm is composed of two steps:

- The E Step (Expectation): It consists of computing $Q(\theta, \theta^{(k)}) = E[L(\theta, Z)|Y, \theta^{(k)}]$ where L is the completed loglikelihood.
- The M Step (Maximisation): $\theta^{(k+1)}$ is obtained by maximising the function Q, $\theta^{(k+1)} = \arg\max_{\theta \in \Theta} Q(\theta, \theta^{(k)})$.

The following equations describe more precisely the EM algorithm in the particular case of model (2). Further notation is used:

$$A_i^{(k)} = Y_i - h_i - J_i(m^{(k)} - x_0), B_i^{(k)} = C^{(k)} J_i^T \text{ and } V_i^{(k)} = J_i C^{(k)} J_i^T + R.$$

where $m^{(k)}$ and $C^{(k)}$ are the value of m and C at iteration k of the EM algorithm.

Then, analytical equations can be derived to compute $m^{(k+1)}$ and $C^{(k+1)}$ with respect to $m^{(k)}$ and $C^{(k)}$.

Proposition 2 *The updating equations for the parameter (m, C) of Model (2) at the M step of the EM algorithm are*

$$m^{(k+1)} = m^{(k)} + \frac{1}{n} \sum_{i=1}^{n} B_i^{(k)} (V_i^{(k)})^{-1} A_i^{(k)}$$

$$C^{(k+1)} = C^{(k)} + \frac{1}{n}\sum_{i=1}^{n}\left((B_i^{(k)}(V_i^{(k)})^{-1}A_i^{(k)})(B_i^{(k)}(V_i^{(k)})^{-1}A_i^{(k)})^T - B_i^{(k)}(V_i^{(k)})^{-1}(B_i^{(k)})^T\right)$$
$$-\frac{1}{n^2}\left(\sum_{i=1}^{n}B_i^{(k)}(V_i^{(k)})^{-1}A_i^{(k)}\right)\left(\sum_{i=1}^{n}B_i^{(k)}(V_i^{(k)})^{-1}A_i^{(k)}\right)^T.$$

Proof:

The density of the complete data can be written

$$p(Z|m,C) = p(Y|X,m,C)p(X|m,C) \propto p(X|m,C).$$

And

$$\begin{aligned} p(X|m,C) &= (2\pi)^{-\frac{nq_1}{2}}|C|^{-\frac{n}{2}}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}(X_i-m)^T C^{-1}(X_i-m)\right) \\ &\propto |C|^{-\frac{n}{2}}\exp\left(-\frac{n}{2}m^T C^{-1}m\right)\exp\left(-\frac{1}{2}\sum_{i=1}^{n}(X_i^T C^{-1}X_i - 2X_i^T C^{-1}m)\right). \end{aligned}$$

This distribution belongs to the exponential familly of distributions and the EM algorithm described in the Appendix for the exponential familly can be used. Considering the canonical parameter $\lambda^T = (\lambda_1, \lambda_2) = (C^{-1}m, C^{-1})$ leads to

$$p(X|m,C) = (2\pi)^{-\frac{nq_1}{2}}|\lambda_2^{-1}|^{-\frac{n}{2}}\exp\left(-\frac{n}{2}\lambda_1^T\lambda_2^{-1}\lambda_1\right)\exp\left(-\frac{1}{2}\sum_{i=1}^{n}(X_i^T\lambda_2 X_i - 2X_i^T\lambda_1)\right).$$

The sufficient statistic for λ is $t(X) = \begin{pmatrix} \sum_{i=1}^n X_i \\ -\frac{1}{2}\sum_{i=1}^n X_i X_i^T \end{pmatrix}$.

At $(k+1)$ iteration, the E-step of EM consists of computing $\mathrm{E}[t(X)|Y,\theta^{(k)}]$ where $\theta^{(k)} = (m^{(k)}, C^{(k)})$.

We have $\mathrm{E}\left(\sum_{i=1}^{n} X_i|Y,\theta^{(k)}\right) = nm^{(k)} + \sum_{i=1}^{n} B_i^{(k)}(V_i^{(k)})^{-1}A_i^{(k)}$.

$$\mathrm{E}\left(-\frac{1}{2}\sum_{i=1}^{n}X_iX_i^T|Y,\theta^{(k)}\right) = -\frac{1}{2}\sum_{i=1}^{n}\left[\mathrm{E}(X_i|Y,\theta^{(k)})\mathrm{E}(X_i^T|Y,\theta^{(k)}) + \mathrm{Var}(X_i|Y,\theta^{(k)})\right].$$

According to (6), it leads to

$\mathrm{E}\left(-\frac{1}{2}\sum_{i=1}^{n}X_iX_i^T|Y,\theta^{(k)}\right) =$

$$-\frac{1}{2}\left[nC^{(k)} + \sum_{i=1}^{n}\left[(m^{(k)} + B_i^{(k)}(V_i^{(k)})^{-1}A_i^{(k)})(m^{(k)} + B_i^{(k)}(V_i^{(k)})^{-1}A_i^{(k)})^T - B_i^{(k)}(V_i^{(k)})^{-1}(B_i^{(k)})^T\right]\right].$$

The M-step of the EM algorithm consists of updating the sufficient statistic $t(X)$. From equation (10) in the Appendix, it results in

$$\mathrm{E}(t(X)|Y, \lambda = \lambda^{(k)}) = \frac{\partial \ln(a)}{\partial \lambda}(\lambda^{(k+1)}),$$

where $\ln(a(\lambda)) = -\frac{n}{2}\left[\ln|\lambda_2| - \lambda_1^T \lambda_2^{-1} \lambda_1\right]$.

Using derivation of matrix and vector formulas, we get

$$\frac{\partial \ln(a)}{\partial \lambda_1}(\lambda) = n\lambda_2^{-1}\lambda_1 = nm$$

$$\frac{\partial \ln(a)}{\partial \lambda_2}(\lambda) = -\frac{n}{2}(\lambda_2^{-1} + \lambda_2^{-1}\lambda_1\lambda_1^T\lambda_2^{-1}) = -\frac{n}{2}(C + mm^T).$$

The iteration $(k+1)$ of the EM algorithm is thus

$$m^{(k+1)} = m^{(k)} + \frac{1}{n}\sum_{i=1}^{n} B_i^{(k)}(V_i^{(k)})^{-1}A_i^{(k)}$$

$$C^{(k+1)} = C^{(k)} + \frac{1}{n}\sum_{i=1}^{n}\left((B_i^{(k)}(V_i^{(k)})^{-1}A_i^{(k)})(B_i^{(k)}(V_i^{(k)})^{-1}A_i^{(k)})^T - B_i^{(k)}(V_i^{(k)})^{-1}(B_i^{(k)})^T\right)$$
$$-\frac{1}{n^2}\left(\sum_{i=1}^{n} B_i^{(k)}(V_i^{(k)})^{-1}A_i^{(k)}\right)\left(\sum_{i=1}^{n} B_i^{(k)}(V_i^{(k)})^{-1}A_i^{(k)}\right)^T.$$

3.2 ECME algorithm

The ECME (Expectation-Conditional Maximisation Either) of Liu and Rubin (1994) is an extension of the ECM algorithm (Meng and Rubin, 1993). The ECM algorithm decomposes the M-step of EM in several CM-steps (Conditional Maximisation) maximising the conditional expectation of the complete-data loglikelihood (conditional on some of the parameters). In the ECME algorithm, some of the CM- steps are replaced by steps maximising the corresponding constrained actual (incomplete-data) loglikelihood, $\ln L(\theta)$, instead of the Q-function. This algorithm is expected to be dramatically faster than the EM and ECM algorithms in terms of the number of iterations required for convergence when EM is jeopardised with slow convergence.

For model (2), an ECME algorithm has been proposed by De Crecy (1996) under the name Circe Method. More precisely, in this case, the iteration $(k+1)$ is expressed as follows: the E Step is the same as in EM and the CM Step of ECM algorithm is replaced by two steps. The first CM Step is as the M-step of EM with m fixed to $m^{(k)}$. The second CM Step maximises the incomplete-data loglikelihood over m, assuming $C = C^{(k+1)}$. According to Liu and Rubin (1994), the ECME algorithm monotonically increases the likelihood function, as the EM algorithm does, $L(\theta)$: $L(\theta^{(k+1)}) \geq L(\theta^{(k)})$; and under some assumptions it reliably converges to a maximum likelihood estimate.

Proposition 3 *The ECME algorithm for model (2) leads to the following updating equations:*

$$C^{(k+1)} = C^{(k)} - \frac{1}{n}\sum_{i=1}^{n}\left[(B_i^{(k)}(V_i^{(k)})^{-1}A_i^{(k)})(B_i^{(k)}(V_i^{(k)})^{-1}A_i^{(k)})^T - B_i^{(k)}(V_i^{(k)})^{-1}(B_i^{(k)})^T\right],$$

$$m^{(k+1)} - x_0 = \left(\sum_{i=1}^{n} J_i^T(V_i^{(k+1)})^{-1}J_i\right)^{-1}\left(\sum_{i=1}^{n} J_i^T(V_i^{(k+1)})^{-1}(Y_i - h_i)\right).$$

Proof:

The *E Step* is the same as in EM and consists of computing the expected completed data sufficient statistic. Denoting $\delta_i = X_i - m$, we have $Y_i = h_i + J_i(\delta_i + m - x_0) + U_i$ and

$$\begin{pmatrix} Y_i \\ \delta_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} h_i + J_i(m - x_0) \\ 0 \end{pmatrix}, \begin{pmatrix} J_iCJ_i^T + R & J_iC \\ CJ_i^T & C \end{pmatrix}\right).$$

The matrix $t(\delta) = -\frac{1}{2}\sum_{i=1}^{n}\delta_i\delta_i^T$ is the sufficient statistic for the canonical parameter

$\lambda = C^{-1}$. In this case, the function a of the exponential family distribution is $a(\lambda) = |\lambda^{-1}|^{\frac{n}{2}}$. Thus, at iteration $(k+1)$, the E-step consists of computing the term $\mathrm{E}[t(\delta)|Y, m^{(k)}, C^{(k)}]$.

$$\begin{aligned}
\mathrm{E}[t(\delta)|Y, m^{(k)}, C^{(k)}] &= -\frac{1}{2}\mathrm{E}\left(\sum_{i=1}^{n}\delta_i\delta_i^T|Y, m^{(k)}, C^{(k)}\right) \\
&= -\frac{1}{2}\sum_{i=1}^{n}\left[\mathrm{E}(\delta_i|Y, m^{(k)}, C^{(k)})\mathrm{E}(\delta_i^T|Y, m^{(k)}, C^{(k)}) + \mathrm{Var}(\delta_i|Y, m^{(k)}, C^{(k)})\right].
\end{aligned}$$

According to (6), it leads to

$\mathrm{E}(\delta_i|Y, m^{(k)}, C^{(k)}) = B_i^{(k)}(V_i^{(k)})^{-1}A_i^{(k)}$ and $\mathrm{Var}(\delta_i|Y, m^{(k)}, C^{(k)}) = C^{(k)} - B_i^{(k)}(V_i^{(k)})^{-1}(B_i^{(k)})^T$,

from which it is deduced

$$\mathrm{E}[t(\delta)|Y, m^{(k)}, C^{(k)}] = -\frac{1}{2}\left[nC^{(k)} + \sum_{i=1}^{n}\left((B_i^{(k)}(V_i^{(k)})^{-1}A_i^{(k)})(B_i^{(k)}(V_i^{(k)})^{-1}A_i^{(k)})^T - B_i^{(k)}(V_i^{(k)})^{-1}(B_i^{(k)})^T\right)\right].$$

CM-step 1 for ECME: (Estimation of the variance C)

Fixing $m = m^{(k)}$, $C^{(k+1)}$ is computed to maximise the expected complete-data likelihood for the exponential family distribution at hand (see the Appendix).

From $\frac{\partial \ln a}{\partial \lambda}(\lambda) = -\frac{n}{2}\frac{1}{\lambda} = -\frac{n}{2}C$, according to the equality (10), it leads to

$$C^{(k+1)} = C^{(k)} - \frac{1}{n}\sum_{i=1}^{n}\left[(B_i^{(k)}(V_i^{(k)})^{-1}A_i^{(k)})(B_i^{(k)}(V_i^{(k)})^{-1}A_i^{(k)})^T - B_i^{(k)}(V_i^{(k)})^{-1}(B_i^{(k)})^T\right].$$

CM-step 2 for ECME: (Estimation of the mean m)

The mean m is updated by maximising the constraint actual (incomplete-data) loglikelihood with C fixed to $C^{(k+1)}$. Since Y_i has a Gaussian distribution $\mathcal{N}(h_i+J_i(m-x_0), V_i^{(k+1)})$, the associated loglikelihood is

$$\mathcal{L}(Y,m) = -\frac{1}{2}\sum_{i=1}^{n}\left(\ln(|V_i^{(k+1)}|) + (Y_i - h_i - J_i(m-x_0))^T(V_i^{(k+1)})^{-1}(Y_i - h_i - J_i(m-x_0))\right) + Cste$$

Solving the equation $\frac{\partial \mathcal{L}}{\partial m}(Y,m) = 0$ leads to

$$m^{(k+1)} - x_0 = \left(\sum_{i=1}^{n} J_i^T(V_i^{(k+1)})^{-1}J_i\right)^{-1}\left(\sum_{i=1}^{n} J_i^T(V_i^{(k+1)})^{-1}(Y_i - h_i)\right).$$

Remark 5 *Identifiability problems can lead to a singular* $\sum_{i=1}^{n} J_i^T(V_i^{(k)})^{-1}J_i$ *matrix (see the flooding model example of section 4.1).*

Remark 6 *Since EM and sometimes ECME algorithms can encounter slow convergence situations, we avoided to use threshold on the likelihood function to assess convergence. On the studied examples, the EM and ECME algorithms are stopped when a large iterations number,* nb_{iter}*, is reached. The values of* nb_{iter} *is chosen on an empirical ground for each example.*

3.3 Relation between estimation quality and H partial derivatives

In this section, it is analysed how a sensitivity index could be related to the estimate variances.

Consider the linear model described in (3) and $\mathcal{L}_n(Y,\theta)$ the associated loglikelihood function. We have

$$\mathcal{L}_n(Y,\theta) = -\frac{np}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\ln(\det(\mathrm{Var}(Y_i))) - \frac{1}{2}\sum_{i=1}^{n}(Y_i - \mathrm{E}(Y_i))^T(\mathrm{Var}(Y_i))^{-1}(Y_i - \mathrm{E}(Y_i)).$$

The coefficients of the function $\dfrac{\partial^2 \mathcal{L}_n}{\partial\theta\partial\theta^T}$ may be analytically computed. And the classical estimation goes as follows:

$$\mathrm{Var}(\hat{\theta}) \cong \left(-\frac{\partial^2 \mathcal{L}_n}{\partial\theta\partial\theta^T}(Y,\hat{\theta})\right)^{-1}$$

Under the assumption that non-diagonal terms are dominated in the information matrix, the estimation variance should be close to the inverse of the diagonal terms:

$$\mathrm{Var}(\hat{m}_j) \cong \left(-\frac{\partial^2 \mathcal{L}_n}{\partial\theta\partial\theta^T}(Y,\hat{\theta})\right)^{-1}|_{j,j} \cong \frac{1}{-\frac{\partial^2 \mathcal{L}_n}{\partial m_j^2}(Y,\hat{\theta})} \tag{8}$$

For model (3), the partial derivatives $\left(\dfrac{\partial^2 \mathcal{L}_n}{\partial m_j^2}\right)_{j=1,\dots,q}$ are obtained by using derivation of vector formulas:

$$-\frac{1}{n}\frac{\partial^2 \mathcal{L}_n}{\partial m_j^2}(Y,\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial H}{\partial x_j}(x_0,d_i)\right)^T (\mathrm{Var}(Y_i))^{-1}\left(\frac{\partial H}{\partial x_j}(x_0,d_i)\right) \tag{9}$$

They may be interpreted as the ratio of squared partial derivative to overall variance, averaged over the $i = 1,\dots,n$ trials.

As a matter of fact, equation (9) could be further interpreted in terms of first order sensitivity indices. Formula (8) and (9) lead to

$$\mathrm{Var}(\hat{m}_j) \cong \frac{1}{n}\frac{\mathrm{Var}(X^{(j)})}{\bar{S}^{(j)}}$$

where $\bar{S}^{(j)} = \dfrac{1}{n}S_i^{(j)}$ and $S_i^{(j)} = \left[(\dfrac{\partial H}{\partial x_j}(x_0,d_i))^T(\mathrm{Var}(Y_i))^{-1}(\dfrac{\partial H}{\partial x_j}(x_0,d_i))\right]\mathrm{Var}(X^{(j)})$

denotes the classical first-order sensitivity index of the function $Y_i = H(X_i,d_i)$, measuring an approximate percentage of importance of the variable $X_i^{(j)}$ in explaining Y_i.

Hence $\dfrac{\mathrm{Var}(\hat{m}_j)}{\mathrm{Var}(X^{(j)})} = \dfrac{1}{n}(\bar{S}^{(j)})^{-1}$.

This means that the "coefficient of variation" of the estimator of the mean (as normalised by the input uncertainty variance) is small as the sensitivity index is large, which is a rather intuitive result. Note also that if, within $\mathrm{Var}(Y_i)$, measurement noise R is high compared to $X^{(j)}$-induced variance, then sensitivity index is lower and estimation variance increases.

However in an estimation process, $\mathrm{Var}(X^{(j)})$ and $\mathrm{Var}(Y_i)$ are not known a priori : only the derivative can be computed, supposing the linearisation point can be inferred without too much error. So that the sensitivity index may not be accessible. Expertise may however infer orders of magnitude of $\mathrm{Var}(X^{(j)})$. In section 4.1, $\mathrm{Var}(\hat{m}_j)$ is approximated by replacing $\mathrm{Var}(X^{(j)})$ with its estimate.

4 Applications

In this section, the EM and ECME algorithms are applied on two examples: The first example concerns simulated data from a flooding model and the second example involves thermal hydraulic flow within a pressurized water nuclear reactor, with experimental flow measurements.

4.1 Illustration on a flooding model

The model is related to the risk of dyke overflow during a flooding. The available model computes the water level at the dyke position (Z_c) and the speed of the river V with respect to the observed flow of the river upstream of the dyke (Q), and several non observed quantities: The river bed level beyond upstream (Z_m) and at the dyke position (Z_v), and the value of Strickler coefficient K_s measuring the friction of the river bed, which is assumed to be homogeneous in this quite simplified model. Thus

$$\begin{pmatrix} Z_c \\ V \end{pmatrix} = H(Z_m, Z_v, K_s; Q) + U$$

with $H(Z_m, Z_v, K_s; Q) = \begin{pmatrix} Z_v + \left(\frac{\sqrt{L}}{B}\right)^{3/5} Q^{3/5} K_s^{-3/5} (Z_m - Z_v)^{-3/10} \\ B^{-2/5} L^{-3/10} Q^{2/5} K_s^{3/5} (Z_m - Z_v)^{3/10} \end{pmatrix}$

where the values of the section length L and its width B are given and assumed to be fixed ($L = 5000, B = 300$). U is a centered Gaussian variable with a diagonal variance matrix and standard deviation 0.2 m and 0.2 m.s^{-1}.

The data have been simulated according to a protocol described in de Rocquigny (2006), assuming that (Z_m, Z_v, K_s) is a Gaussian vector such that

$$\begin{pmatrix} Z_m \\ Z_v \\ K_s \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} m_m \\ m_v \\ m_s \end{pmatrix}, \begin{pmatrix} \sigma_m^2 & 0 & 0 \\ 0 & \sigma_v^2 & 0 \\ 0 & 0 & \sigma_s^2 \end{pmatrix} \right).$$

In this example, there is clearly an identifiability issue since columns 1 and 3 of the Jacobian of H, J_H, are linearly dependent (cf. Section 2).But even this identifiabiliy issue would have been easily detected beforehand thanks to the criteria given in Proposition 1, it is still instructive to study the behaviour of EM and ECME algorithms in such a situation.

For the ECME algorithm, the matrix $\sum_{i=1}^{n} J_i^T V_{i,k}^{-1} J_i$ appears to be singular. Hence the algorithm does not converge. For the EM algorithm, Table 1 summarises the estimation obtained from 25 random runs initiated with Gaussian distributions centered to the true values of the parameters, with $nb_{iter} = 100$. This table provides the true parameter values, the maximum likelihood estimates, the mean of the 25 estimated values from the EM algorithm, its standard deviation and an associated Normalised Error Coefficient (NEC) which is for a parameter θ the ratio (standard deviation of $\hat{\theta}$)/$\hat{\sigma}_\theta$.

Remark 7 *Observing large values of NECs simultaneously on several parameters that are connected to different physical variables is usually symptomatic of non-identifiability. Roughly speaking, it means that the estimation algorithm is unable to differenciate the role of these physical parameters, and that the response of the algorithm –among the infinite set of equivalent solutions– is mainly driven by the starting point. Nonetheless, it is to be reminded that this diagnostic is not as reliable as that provided by Proposition 1: there is no conventional threshold that clearly indicates what should be considered as a "large" NEC value.*

It is to be noticed in Table 1 that the NECs are high for variance parameters σ_{Z_m} and σ_{K_s} and also for m_{K_s}, showing the identifiability problem occurring between variables Z_m and K_s.

Parameters	True values	Max Likelihood Estimates	Mean Estimates	Standard deviation	Norm. Error
m_{Z_m}	55	54.4606	55.2145	1.0436	2.2506
m_{Z_v}	50	50.1088	50.1074	0.0013	0.0024
m_{K_s}	30	32.1343	29.6814	3.1326	0.3939
σ_{Z_m}	$\sqrt{1/6} \simeq 0.41$	0.3733	0.4637	0.1508	0.3252
σ_{Z_v}	$\sqrt{1/6}$	0.5371	0.5356	0.0012	0.0022
σ_{K_s}	7.5	7.5810	7.9525	3.0445	0.3828

Table 1: Parameters estimates for the flooding model with EM algorithm from 25 random initial positions.

To try to solve the identifiability problem, in a first time, the mean parameter m_{K_s} is fixed equal to 30 according to expert judgement, and the other parameters are estimated with the EM and ECME algorithms, with $nb_{iter} = 100$, from 25 random initial positions. Results are summarised in Tables 2 and 3. For both algorithms, the NECs are still large for variance parameters σ_{Z_m} and σ_{K_s}. It means that there is always an identifiability problem between variables Z_m and K_s: determining only partial characterisics of the probability distribution through expert judgement is not sufficient to overcome this issue.

Finally, the parameters relative to K_s, m_{K_s} and σ_{K_s}, are fixed at the true values –which would require in practice a deep and tedious analysis by the experts– while the parameters relative to Z_m and Z_v are estimated through the EM and ECME algorithms with $nb_{iter} = 100$. Results, from 25 random initial positions, are summarised in Tables 4 and 5. NECs values indicate that identifiability is not an issue anymore; as a side effect, it can be seen that the parameter σ_{Z_v} is now better estimated (0.45 instead of 0.53 previously). Thus, the advanced expert knowledge injected here has allowed to obtain a robust solution to the inverse problem; however, it is to be reminded that such a precise expert judgement on a probability distribution (and not only its mean) is often difficult to achieve in practice. Comparing results obtained with EM and ECME algorithms, mean parameter estimates m_{Z_m} and m_{Z_v} are the same but NECs are smaller with ECME: This comes from the slower convergence of the EM algorithm, which results in a dependency between the starting point and the algorithm output. For variance parameters results are similar (estimates and NECs).

Remarks:

The variance $\mathrm{Var}(\hat{m}_j)$, with $j = Z_m$ and $j = Z_v$, have been computed by replacing $\mathrm{Var}(X^{(j)})$ with its estimate in the formulae given in section 3.3. It leads to $\mathrm{Var}(\hat{m}_{Z_m}) = 0.0306$ and $\mathrm{Var}(\hat{m}_{Z_v}) = 0.0021$. Using the true values of $\mathrm{Var}(X^{(j)})$, we obtain $\mathrm{Var}(\hat{m}_{Z_m}) = 0.0334$ and $\mathrm{Var}(\hat{m}_{Z_v}) = 0.0019$. The proposed approximation works well for this identifiable model. Moreover these results give analogous indications on the estimate quality that the Normal Error (NEC).

If we know that the variables Z_m and Z_v are independent, a possibility is to fix covariance terms equal to 0 in the variance matrix C. With the ECME algorithm, estimates of m_{Z_m} and m_{Z_v} are not changed ($m_{Z_m} = 55.2087$ and $m_{Z_v} = 50.0785$) but estimates of σ_{Z_m} and σ_{Z_v} are better ($\sigma_{Z_m} = 0.3467$ and $\sigma_{Z_v} = 0.4196$). Thus it could be profitable, (as soon there are making sense ...), to use such constraints because it leads to more parsimonieus models.

Parameters	True values	Max Likelihood Estimates	Mean Estimates	Standard deviation	Norm. Error
m_{Z_m}	55	54.1364	55.0313	0.2276	0.4835
m_{Z_v}	50	50.1088	50.1047	0.0123	0.0229
σ_{Z_m}	$\sqrt{1/6} \simeq 0.41$	0.5688	0.4707	0.1128	0.2397
σ_{Z_v}	$\sqrt{1/6}$	0.5368	0.5353	0.0011	0.0015
σ_{K_s}	7.5	7.4553	7.9726	2.3354	0.2929

Table 2: Parameters estimates for the flooding model via EM, from 25 random initial positions, with fixed $m_{K_s} = 30$.

4.2 Illustration on a thermohydraulics case study

The dataset studied in this paragraph comes from experiments carried out by EDF/R&D between 1990 and 1991 to study the redistributions of diphasic flows. The aim is to evaluate the capacity of a thermohydraulics physical model, named THYC, to assess void fractions in a beam of tubes.

The following elements are available:

Parameters	True values	Max Likelihood Estimates	Mean Estimates	Standard deviation	Norm. Error
m_{Z_m}	55	55.1718	55.1202	0.0318	0.0671
m_{Z_v}	50	50.1087	50.1082	0.0017	0.0032
σ_{Z_m}	$\sqrt{1/6} \simeq 0.41$	0.5689	0.4741	0.1338	0.2400
σ_{Z_v}	$\sqrt{1/6}$	0.5370	0.5355	0.0011	0.0024
σ_{K_s}	7.5	7.4591	7.9002	3.1452	0.3981

Table 3: Parameters estimates for the flooding model via ECME, from 25 random initial positions, with fixed $m_{K_s} = 30$.

Parameters	True values	Max Likelihood Estimates	Mean Estimates	Standard deviation	Norm Error
m_{Z_m}	55	55.2110	55.2419	0.1432	0.3098
m_{Z_v}	50	50.0786	50.0827	0.0196	0.0432
σ_{Z_m}	$\sqrt{1/6} \simeq 0.41$	0.5995	0.4622	0.1063	0.2230
σ_{Z_v}	$\sqrt{1/6}$	0.4610	0.4530	0.0079	0.0175

Table 4: Parameters estimates for the flooding model via EM, from 25 random initial positions, with fixed $m_{K_s} = 30$ and $\sigma_{K_s} = 7.5$.

Parameters	True values	Max Likelihood Estimates	Mean Estimates	Standard deviation	Norm Error
m_{Z_m}	55	55.2117	55.2128	6.7859e-04	0.0015
m_{Z_v}	50	50.0788	50.0787	3.5020e-05	7.7290e-05
σ_{Z_m}	$\sqrt{1/6} \simeq 0.41$	0.5978	0.4619	0.1063	0.2302
σ_{Z_v}	$\sqrt{1/6}$	0.4608	0.4531	0.0079	0.0174

Table 5: Parameters estimates for the flooding model via ECME, from 25 random initial positions, with fixed $m_{K_s} = 30$ and $\sigma_{K_s} = 7.5$.

- Experimental results $(Y_i)_{1 \le i \le n}$ (void fractions measurements), obtained for different thermohydraulic conditions (e.g. flows) and at different coordinates, these experimental conditions being denoted by (d_i); the experiments are considered as independent,
- the standard deviation of the measurement errors $(U_i)_{1 \le i \le n}$,
- the void fractions $(H(x_0, d_i))_{1 \le i \le n}$ computed by THYC for the different experimental conditions (d_i), the uncertain input variables of THYC being fixed at "best estimate" values $x_0 = 0$,

- the Jacobian matrices $(J_H(x_0, d_i))_{1 \le i \le n}$, which gather the partial derivatives of THYC's outputs with respect to the uncertain input variables.

Hence, the inverse problem studied is characterized by the equation:

$$Y_i = H(X_i, d_i) + U_i, \quad 1 \le i \le n = 643$$

where:
(Yi): scalar value (void fraction measurement),
(X_i) : non observed data vector of dimension $\mathbb{R}^{11}$, assumed i.i.d with a Gaussian distribution $\mathcal{N}(m, C)$,
(d_i) : deterministic variables, with dimension q_2,
H : function from $\mathbb{R}^q$ to $\mathbb{R}$ $(q = 11 + q_2)$,
(U_i) : measurement errors assumed i.i.d with a Gaussian distribution $\mathcal{N}(0, R)$, with $R = 9.10^{-4}$.

The linearised model is $Y_i = H(x_0, d_i) + J_H(x_0, d_i)(X_i - x_0) + U_i, \quad 1 \le i \le n$. And the aim is to estimate the mean parameter m and the variance matrix C.

In a first time, to ensure the model identifiability, the Jacobian matrix has been studied as recommended in section 2. The determinant of HTH is actually quite close to zero, though not exactly null; this means that identifiability is here an issue, even if the model is not exactly non-identifiable in theory. Unlike the flooding model in which ECME detects non-identifiability via a non-inversible matrix, ECME then provides results and misleading appearance of convergence, but the results are actually littered by numerical inversion problems. "Close to colinear" variables have then been identified from graphical analyses, as for example Figure 2 for variables six and ten. Thence seven variables have been kept (number 1, 2, 3, 5, 8, 9, 10). For the other variables (numbers 4, 6, 7, 10), the approach used to overcome non-identifiability is different from that described in the flooding example. Indeed, it has been judged too difficult to fix arbitrarily relevant values of their standard deviations; these parameters have therefore been fixed at their best estimated values, thus neglecting their stochastic nature.

In Tables 6 and 7 the estimates are given for the mean vector m, using the ECME and EM algorithms, with $nb_{iter} = 5000$ and the same initial values. Here the NEC is equal to $\frac{\sqrt{C_{i,i}}}{m_i}$.
Several differences appear between the results of ECME and EM algorithms. But the loglikekihood values are respectively: $\mathcal{L}(Y, \hat{\theta}_{ECME}) = 1836.03$ and $\mathcal{L}(Y, \hat{\theta}_{EM}) = 1834.27$. This simply means that EM algorithm has not yet converged despite the large number of iterations (see Figure 1 for a graphical confirmation). ECME fulfills in this example its original goal: It outperforms EM in terms of convergence speed.

Parameters	Estimates	Variation coefficient
m_1	0.0670	0.4452
m_2	0.8736	1.0402
m_3	-0.8432	0.5334
m_5	-0.1186	3.2669
m_8	0.5123	0.6086
m_9	0.3313	3.1854
m_{10}	-0.0005	9.8094

Table 6: Estimates and Normalised Error Coefficients for the mean vector with the ECME algorithm, for the thermohydraulics model.

Parameters	Estimates	Norm. Error
m_1	0.0662	0.4485
m_2	0.6910	1.2650
m_3	-0.7239	0.6007
m_5	-0.0893	4.3887
m_8	0.5744	0.5806
m_9	0.2569	4.2284
m_{10}	-0.0017	2.8562

Table 7: Estimates and Normalised Error Coefficient for the mean vector with the EM algorithm, for the thermohydraulics model.

An important question is to decide if all variables have to be kept or is it possible to select the more significant ones among the seven variables ? In the following it is analysed if it is possible to select six variables among the seven variables.

In that purpose, the AIC (Akaike, 1974) and BIC (Schwarz 1978) model selection criteria are used:

$$AIC = -2\mathcal{L}_n(Y, \hat{\theta}) + 2q$$

$$BIC = -2\mathcal{L}_n(Y, \hat{\theta}) + q\log(n)$$

where n is the dimension of the sampling Y, $\mathcal{L}_n$ the loglikelihood function, $\hat{\theta}$ the maximum likelihood estimate of θ and q the number of free parameters to be estimated.

In Table 8 are given the values of the AIC and BIC criteria according to parameter estimates obtained with the ECME algorithm and $nb_{iter} = 5000$. Selected variables are successively removed and in the last line the seven parameters are kept.

For both model selection criteria, the selected model is the one where the seven parameters are kept.

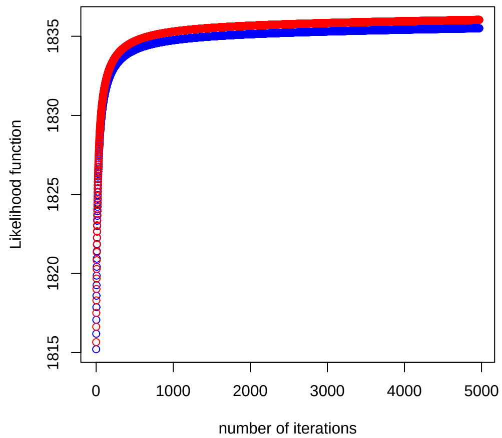

Figure 1: It represents the log-likelihood function according to the iteration number with ECME (red curve) and EM (blue curve).

Removed variable	AIC	BIC
m_1	-3291.36	-3264.57
m_2	-3572.98	- 3546.18
m_3	-3571.96	- 3545.16
m_5	-3631.04	-3604.24
m_8	-3649.77	- 3622.98
m_9	-3645.04	-3618.21
m_{10}	-3589.95	-3563.15
none	-3658.07	-3626.81

Table 8: Values of AIC and BIC according to parameter estimates obtained with ECME, for the thermohydraulics model.

5 Discussion

Linearisation of complex multivariate systems could be regarded as a reference method for uncertainty modelling. In many situations, this approach leads without too much human effort and quite rapidly to a reasonable estimation of uncertainty factors. Parameters of the linearised model (2) can be estimated efficiently without too much computational burden with EM-like algorithms. In the two examples presented in Section 4, the ECME algorithm

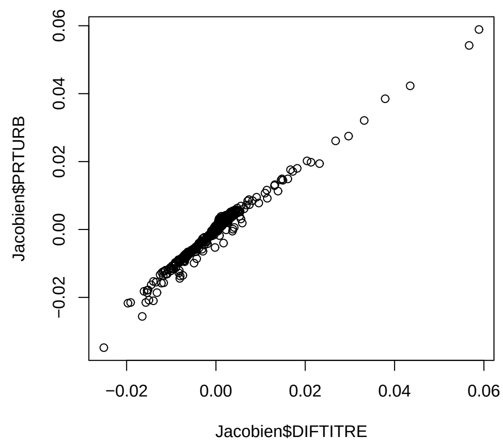

Figure 2: It represents the values of the gradient of the H function for the variables six in function of the variable ten.

appears to outperform the EM algorithm in terms of convergence speed (in particular for the thermohydraulics example). The choice of this algorithm seems therefore sensible to solve the inverse problem at hand and identify the intrinsic variability of the uncertain variables studied. However, the choice of the stopping criterion of the algorithm should be considered cautiously: slow convergence may sometimes affect ECME and give a misleading impression of convergence, even if it is less critical than for EM. And in any case, the use of this algorithm should always be preceded in practice by an analysis of the identifiabiliy issue: As shown in this paper, it is rather simple to detect and resolve unidentifiable models.

Obviously, the relevancy of these algorithms is intrinsically tight to that of the main assumptions (linear approximation and Gaussian distributions). If at least one of these assumptions appears irrelevant, then the estimation problem will require more complex procedures. First, linearisation could be considered with alternative dispersion distributions as a Weibull or an extreme value distribution. On the other hand, in order to treat the inital statistical model (1) with a controlled number of runs of the physical model, several approaches are currently studied: coupling a Stochastic version of the EM algorithm with Monte Carlo Markov Chains tools, or using an Importance Sampling procedure to approximate properly the likelihood. But even in that setting, linear approximation methods could be quite helpful to overcome some possible difficulties and tune efficiently the methods (choice of a starting point for complex stochastic algorithms, selection of variables to get an identifiable model, etc.).

Acknowledgements

The authors would like to thank Agnès de Crecy (CEA) and Franck Maurel (EDF/R&D) for their support, comments and suggestions.

Appendix: EM algorithm for an exponential family

It is assumed that the vector Z can be written $Z = (X, Y)$ where X is the non-observed data vector and Y is observed. It is assumed that the density of Z can be written under the form

$$p(z|\lambda) = \frac{b(z)}{a(\lambda)} \exp(\lambda^T t(z)).$$

where $\lambda = \lambda(\theta)$ is the vector of parameters called canonical parameter, $t(Z)$ being the corresponding sufficient statistic and $a(\lambda), b(z)$ real functions.

The E Step of the EM algorithm consists of computing $Q(\theta, \theta^{(k)}) = E[L(\theta, Z)|Y, \theta^{(k)}]$ where L is the completed loglikelihood.

Using the canonical parameter λ, we can write

$$\tilde{Q}(\lambda, \lambda^{(k)}) = E[\tilde{L}(\lambda, Z)|Y, \lambda^{(k)}] = \int \tilde{L}(\lambda, z) p(x|Y, \lambda = \lambda^{(k)}) dx$$

with $\tilde{L}(\lambda, z) = \ln b(z) - \ln a(\lambda) + \lambda^T t(z)$ and $p(x|Y, \lambda = \lambda^{(k)})$ the density of non-observed data X conditionally to the observed data Y and the current value of parameter $\lambda^{(k)}$.

Then $\tilde{Q}(\lambda, \lambda^{(k)}) = \lambda^T E[t(Z)|Y, \lambda^{(k)}] - \ln a(\lambda) + Cste$.

And the E Step consists of computing the conditional expectation of $t(Z)$ conditionally to observed data Y and the current value of the parameter $\lambda = \lambda^{(k)}$.

The M Step consists of computing $\lambda^{(k+1)} = \arg\max_{\lambda \in \Lambda} \tilde{Q}(\lambda, \lambda^{(k)})$. Solving $\frac{\partial \tilde{Q}}{\partial \lambda}(\lambda, \lambda^{(k)}) = 0$, $\lambda^{(k+1)}$ satisfied

$$E[t(Z)|Y, \lambda^{(k)}] = \frac{\partial \ln a}{\partial \lambda}(\lambda^{(k+1)}). \tag{10}$$

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716-723.

Beck, J.V. and Arnold, K.J. (1977). *Parameter Estimation in Engineering and Science.* Wiley.

De Crecy, A. (1996). Determination of the uncertainties of the constitutive relationships in the Cathare 2 Code. *Proceedings of the 1996 4th ASME/JSME International Conference on Nuclear Engineering.*

Dempster, E. J., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via EM algorithm. *Annals of the Royal Statistical Society, Series B*, **39**, 1-38.

Liu, C. and Rubin, D.B. (1994). The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. *Biometrika*, **81**, 633-648.

Mahé, P. and de Rocquigny, E. (2005). Incertitudes non observables en calcul scientifique industriel - Etude d'algorithmes simples pour intégrer dispersion intrinsèque et ébauche d'expert. *Actes des* $37^{ème}$ *Journées de Statistique de la SFdS.*

Meng, X.L. and Rubin, D.B. (1993). Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika*, **80**, 267-278.

Saporta, G. (1990). *Probabilités, analyse des données et Statistique.* Edition Technip.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, **6**, 461-464.

Talagrand, O. (1997) Assimilation of observations, an introduction. *J. of Meteorological Soc. Of Japan*, **75**, 191-201.

Tarantola, A. (1987). *Inverse Problem Theory and methods for data fitting and model parameter estimation.* Elsevier - Amsterdam.

Unité de recherche INRIA Futurs
Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399