Quasi-Bayes Approach to Multi-Model Fault Detection and Isolation

Miroslav Kárný, Ivan Nagy, Jana Novovičová

Institute of Information Theory and Automation, AVČR, P.O. Box 18, 182 08 Prague 8, Czech Republic, school@utia.cas.cz

December 7, 1999

Contents

1 Introduction

Early recognition/isolation of a faulty behavior of a dynamic system is the main task of a fault detection and isolation (FDI). Its importance has stimulated a lot of research activities that have provided a wide set of general as well as specific results. For reviews see [1, 2]. The development is definitely not finished yet. Complexity of the inspected systems as well as complexity of the formally optimal solutions call for a further progress. This research report contributes to a solution of the complexity aspect of FDI. It proposes a quasi-Bayesian estimation algorithm [3] for resolving the specific instance of the "curse of dimensionality" connected with use of the multi-model approach to FDI [4, 5].

The research report is organized as follows. After introducing the necessary mathematical preliminaries the FDI problem is formulated in probabilistic terms. Section 4 presents the Bayes solution of FDI problem. Section 5 concerns the approximation of the formal solution of FDI problem using the quasi-Bayes procedure ([6] and the references therin). The result is then elaborated for practically significant and feasible exponential family of models [7]. A simple example illustrates properties of the proposed algorithm. It deals with two faulty states only but handles both discrete and continuous data, i.e. it handles so called mixed data model. Finally, conclusions are drawn.

The main contributions of the research report are:

- adaptation of the novel quasi-Bayes algorithm for an efficient treatment of FDI problems
- novel solution of on-line, possibly adaptive estimation of a dynamic mixture model that is able to cope both with continuous and discrete value data.

2 System description and problem formulation

In this section we introduce first some concepts and definitions concerning the fault behavior of a dynamic system, the loss function and the type of probabilistic models used throughout the research report. The notation generally follows the standard conventions of mathematics and statistics. By x^* we denote the *set* with common elements x and $\overline{x}$ means *cardinality* of the set x^*. We use the term probability density function in both the continuous and discrete cases, abbreviating it in the usual way to pdf. The difference between a random variable and its realization is omitted. We denote the density function of x as $p(x)$ and y as $p(y)$, even though the functions differ.

The considered behavior of the system can be described as follows. At each time $t \in t^* \equiv \{1, 2, \ldots, \overline{t}\}$, the system dynamics may exhibit one fault from a finite set of faults, say $f^* \equiv \{0, 1, \ldots, \overline{f} - 1\}$. As usual, the value $f = 0$ corresponds with a normal (non-faulty) behavior of the system. At each time instance $t \in t^*$, some data d_t are observed on the system reflecting the current fault is $f_t \in f^*$. The sequence of all observed values d_τ, respectively of all faults f_τ, $\tau = 1, \ldots, t$ is denoted by $d(t) \equiv \{d_1, \ldots, d_t\}$, respectively $f(t) \equiv \{f_1, \ldots, f_t\}$.

The fault detection task consists of deciding whether $f_t = 0$ or $f_t \neq 0$. It is a simpler version of the fault isolation that should estimate the true f_t at time t after having observed data $d(t-1)$. Thus, a decision $\hat{f}_t \in f^*$ should be taken that classifies the behavior of the system as exhibiting the fault $f_t = \hat{f}_t$. This decision can be based on observed data only, i.e. $\hat{f}_t = \hat{f}_t(d(t-1))$. A sequence of rules generating fault estimates $\hat{f}(t) \equiv \{\hat{f}_t\}_{t\in t^*}$, is called the FDI strategy. We would like to take an optimal FDI strategy. Intuitively it is clear that a model relating the unknown f_t to the available data $d(t-1)$ is needed. Its form results from a formalization of the notion "optimal".

The FDI detection strategy generating the estimates $\hat{f}^0(t) \equiv \{\hat{f}^0_t\}_{t\in t^*}$, is said to be *optimal FDI strategy* if the expectation $\mathcal{E}$ of the loss function

$$L(f(\bar{t}), \hat{f}(\bar{t})) \equiv \sum_{t\in t^*} l(f_t, \hat{f}_t),$$

determined by a table of particular loss functions $l(f_t, \hat{f}_t) \geq 0,\ t \in t^*$, reaches its minimal value for $\hat{f}(\bar{t}) = \hat{f}^0(\bar{t})$.

It is known that the optimal FDI strategy can be constructed value-wise using dynamic programming technique. It leads to a backward recursive optimization

$$V(d(t-1)) = \min_{\hat{f}_t \in \hat{f}^*} \mathcal{E}[l(f_t, \hat{f}_t) + V(d(t))|d(t-1), \hat{f}_t],\ t \in t^*,\quad V(d(\bar{t})) = 0 \tag{1}$$

where the function $V(d(t))$ is known as the optimal cost-to-go.

As we have seen, the construction of the optimal FDI strategy relies on our ability to evaluate the conditional expectations in the recursion (1). We need to have the probability density function $p(d_t, f_t|d(t-1), \hat{f}_t)$ for all $t \in t^*$. This pdf is the marginal pdf of the joint pdf

$$p(d_t, F_t|d(t-1), \hat{F}_t). \tag{2}$$

Here we introduced the *fault-configuration* F_t of the length $(n+1)$, $n \geq 0$ as a finite sequence of faults $F_t = \{f_t, f_{t-1}, \ldots, f_{t-n}\}$. Similarly, we used the fault-configuration estimate $\hat{F}_t = \{\hat{f}_t, \hat{f}_{t-1}, \ldots, \hat{f}_{t-n}\}$ of the same length $(n+1)$.

Remark 1 *In practice, it is sufficient to consider the small n, $n \leq 2$ of fault-configurations F_t and their estimates $\hat{F}_t$.*

A design of the algorithm for an efficient (approximate) on-line construction of the pdf (2) is addressed in this research report.

For constructing the needed pdf (2), relationships of observed data to faults and fault estimates have to be modeled. The desired pdf is then obtained through parameter estimation. It is easy to be lost in the variety of pdfs we deal with. We introduce them formally in order to minimize this danger. At the same time, we introduce some assumptions and definitions related to the described system and solved task.

$\mathcal{A}1$. The pdf predicting data d_t after observing data $d(t-1)$ and given faults $f(t)$ and their estimates $\hat{f}(t)$ satisfies

$$p(d_t|d(t-1), f(t), \hat{f}(t)) = p(d_t|d(t-1), F_t, \hat{F}_t).$$

The pdf $p(d_t|d(t-1), f(t), \hat{f}(t))$ satisfying the condition specified in $\mathcal{A}1$ is called *observation model*. The observation model is therefore fully determined by observed data $d(t-1)$ and the newest fault-configuration F_t as well as its estimate $\hat{F}_t$.

$\mathcal{A}2$. The pdf predicting data d_t after observing data $d(t-1)$ and given the fault-configuration F_t, its estimate $\hat{F}_t$ and unknown parameters $\theta \in (\theta)^*$ satisfies

$$p(d_t|d(t-1), F_t, \hat{F}_t, \theta) = p(d_t|\Psi_{t-1}, F_t, \hat{F}_t, \theta)$$

where we denote by Ψ a finite fix dimensional vector of past data items (current input is not included).

The pdf $p(d_t|d(t-1), F_t, \hat{F}_t, \theta)$ satisfying the condition $\mathcal{A}2$ is called *parameterized observation model*.

$\mathcal{A}3$. The pdf predicting the fault-configuration F_t after observing data $d(t-1)$ and fault estimates $\hat{f}(t)$ satisfies

$$p(F_t|d(t-1), \hat{f}(t)) = p(F_t|d(t-1), \hat{F}_t).$$

The pdf $p(F_t|d(t-1), \hat{f}(t))$ satisfying the condition specified in $\mathcal{A}3$ is called *fault-configuration model*. The fault estimates $\hat{f}(t)$ are supposed to be generated by FDI strategies among which the optimal FDI strategy is searched for.

$\mathcal{A}4$. The pdf predicting the fault-configuration F_t after observing data $d(t-1)$, fault estimates $\hat{f}(t)$ and given by the unknown parameters $\alpha \in \alpha^*$ satisfies

$$p(F_t|d(t-1), \hat{f}(t), \alpha) = p(F_t|\hat{F}_t, \alpha).$$

The pdf $p(F_t|d(t-1), \hat{f}(t), \alpha)$ satisfying the condition specified in $\mathcal{A}4$ is called *parameterized fault-configuration model*. The assumed model is of time-invariant Markov type. Throughout the research report, the parameterized fault-configuration model of the following type

$$p(F_t|d(t-1), \hat{f}(t), \alpha) = \alpha_{F|\hat{F}} \tag{3}$$

is used, where the time-invariant array $[\alpha_{F|\hat{F}}]$ of transition probabilities α belong to the probabilistic simplex

$$\alpha^* = \left\{ \alpha : \alpha_{F|\hat{F}} \geq 0, \sum_{F \in F^*} \alpha_{F|\hat{F}} = 1, \; \forall \hat{F} \in \hat{F}^* \right\}.$$

The joint pdf

$$p(\alpha, \theta, d_t, F_t|d(t-1), \hat{F}_t) \tag{4}$$

of all unobservable quantities,

$$\alpha \in \alpha^*, \; \theta = \{\theta\}_{F \in F^*}, \; F_t \in F^*$$

and the unobserved d_t, given by observed data $d(t-1)$ and by the fault-configuration estimate $\hat{F}_t$ is called *joint model.*

The pdf $p(d_t|d(t-1), \hat{f}(t), \theta, \alpha) = p(d_t|d(t-1), \hat{F}_t, \theta, \alpha)$ predicting data d_t after observing data $d(t-1)$ and fault-configuration estimates $\hat{F}_t$, given parameters θ, α is called *parameterized data model.* It follows from $\mathcal{A}2$, $\mathcal{A}4$ and from the computation formula for the marginal density that the parameterized data model has the following mixture form

$$p(d_t|d(t-1), \hat{F}_t, \theta, \alpha) = \sum_{F_t \in F^*} p(d_t|d(t-1), F_t, \hat{F}_t, \theta)\alpha_{F_t|\hat{F}_t}. \tag{5}$$

3 Multi-model structure description

Structure vector

We deal with multi-model components which correspond to particular faulty configuration $F \in F^* = \{1, 2, ..., \overline{F}\}$. The pdfs of the components are either known or can be expressed in a parameterized exponential form with linear (in parameters) exponent

$$\exp\{b'^F c^F(\psi)\}$$

where

b^F is $\overline{b^F}$-dimensional column vector of parameters connected with the faulty configuration F; (b'^F is an abbreviation for $(b^F)'$),

c^F is a column vector of statistics with dimension $\overline{c^F} = \overline{b^F}$ depending on lately measured data involved in the regressor ψ.

We suppose, that the dimensions of vector parameters b^F can be different for each $F \in F^*$. For description of such a structure we introduce *structure vector* ν_b

$$\nu_b = [\overline{b^1}, \overline{b^2}, ..., \overline{b^{\overline{F}}}]' \tag{6}$$

which is $\overline{F}$-dimensional vector consisting of positive integers – dimensions of particular parameter vectors.

Now, the model components are dependent on both continuous and discrete variables. The dependence on the discrete ones can be expressed again by a decomposition of the components to $\bar{c}$ sub-models, each of them valid for a particular configuration of values of the discrete data. So, the structure of each component can be similarly expressed by a structure vector ν^F which is now determined by discrete data, so we connect it with the statistics c rather than with the parameters b. For each component $F \in F^*$ we define

$$\nu^F = [\overline{(c_1^F)}, \overline{(c_2^F)}, ..., \overline{(c_{\bar{c}}^F)}]' \tag{7}$$

where $\overline{(c_i^F)}$, $i = 1, 2, ..., \overline{c^F}$ are dimensions of statistics corresponding to the i-th configuration of discrete data within the F-th model component.

The whole multi-model structure gets finer and can now be described by *structure multi-vector* $\mu_c = [\nu'^1|\nu'^2|...|\nu'^{\overline{F}}]'$

$$\mu_c = \left[\overline{(c_1^1)}, \overline{(c_2^1)}, \cdots, \overline{(c_{\overline{c}}^1)} \mid \overline{(c_1^2)}, \overline{(c_2^2)}, \cdots, \overline{(c_{\overline{c}}^2)} \mid \cdots\cdots \mid \overline{(c_1^{\overline{F}})}, \overline{(c_2^{\overline{F}})}, \cdots, \overline{(c_{\overline{c}}^{\overline{F}})}\right]' \tag{8}$$

Remark 2 *Indeed, in the introduced notation there must hold*

$$\overline{b^F} = \sum_{i=1}^{\overline{c}} \overline{c_i^F} = \nu^F \, \underline{1} \ \text{ for all } \ F \in F^*$$

where $\underline{1}$ *is a column vector consisting of ones with appropriate dimension.*

δ-vector
For selecting $j-$th item of an m-vector $p = [p_1, p_2, ..., p_m]'$ we introduce *δ-vector* $\underline{\delta}(j)$

$$\underline{\delta}(j) = [\delta(1,j), \delta(2,j), ..., \delta(m,j)]' \ ; \quad j = 1, 2, ..., \overline{j} \tag{9}$$

where
$\delta(i,j)$ is Kronecker delta, i.e. $\delta(i,j) = 1$ if $i = j$ and zero otherwise.
The $j-$th item of p can be expressed as $p_j = p' \, \underline{\delta}(j)$.

Vector of ones
For summing items of an m-vector $p = [p_1, p_2, ..., p_m]'$ we introduce *vector of ones* $\underline{1}_k$ which is a k-dimensional column vector of ones. Then $\sum_{i=1}^m p_i = p' \, \underline{1}_m$

A generalization of the previously introduced symbol is a *block-vector of ones* $\underline{1}_\kappa$ where $\kappa = [k_1, k_2, ..., k_m]'$ is an m-vector of nonnegative integers defined as

$$\underline{1}_\kappa = [\, \underline{1}'_{k_1}, \, \underline{1}'_{k_2}, ..., \, \underline{1}'_{k_m}]' \tag{10}$$

which is a vector consisting of blocks of ones with dimensions $k_1, k_2, ..., k_m$.

I-matrix
For selecting the model components according to the current fault and values of the discrete data described by structure multi-vector, we introduce *I-matrix*

$$I(\mu, r) = I([\nu'^1|\nu'^2| \cdot |\nu'^{\overline{F}}], r) = \text{diag}\{\, \underline{1}'_{\nu^1} \odot r, \, \underline{1}'_{\nu^2} \odot r, ..., \, \underline{1}'_{\nu^{\overline{F}}} \odot r\} \tag{11}$$

where

$\mu = [\nu'^1|\nu'^2| \cdot |\nu'^{\overline{F}}]'$ is a structure multi-vector, ν^i being subvectors of integers with equal dimension $\overline{c}$

$\underline{1}'_{\nu^F}$, $F = 1, 2, \cdots, \overline{F}$ is a block-vector of ones corresponding to the fault F with block dimensions $\nu_1^F, \nu_2^F, ..., \nu_{\overline{r}}^F$.

r is a real vector with the dimension $\overline{c} = \overline{\nu^1} = \overline{\nu^2} = ... = \overline{\nu^{\overline{r}}}$

$\odot$ is so called Khatri-Rao product defined for block vectors p, q by means of Kronecker product as follows

$$p \odot q = [p_1 \otimes q_1, p_2 \otimes q_2, ..., p_{\overline{p}} \otimes q_{\overline{p}}] \tag{12}$$

If p, q are column vectors with scalar items, this product can be expresses in an ordinary matrix notation $p \odot q = p' \operatorname{diag}\{q\} = (\operatorname{diag}\{p\}q)'$

particular items in diagonal of (11) are Khatri-Rao products of unit block-vectors $\underline{1}'_{\nu^i}$ with $\overline{c}$ blocks by real $\overline{c}$-dimensional vector r. The dimensions of blocks are $\nu^i_j,\ j = 1, 2, ...\overline{c},\ i = 1, 2, ..., \overline{F}$.

In the case concerning only the upper-level structure (the one induced by faulty configurations) the *I-matrix* is

$$I(\nu, r) = \operatorname{diag}\{\nu \odot r\} = \operatorname{diag}\{\ \underline{1}_{\nu_1} r_1,\ \underline{1}_{\nu_2} r_2, ...,\ \underline{1}_{\nu_{\overline{r}}} r_{\overline{r}}, \} \tag{13}$$

where particular items represent multiplications of vectors $\underline{1}_{\nu_F} = \underline{1}_{\overline{b^F}}$ (see (6)) by scalars $r_F,\ F = 1, 2, ..., \overline{F}$.

Remark 3 *The name I-matrix is derived from the fact that the vector r is mainly substituted by the δ-vector $\underline{\delta}$ of Kronecker delta functions. It chooses a specified component of the multi-model. By substituting $\underline{1}$ for r, we choose all components.*

With the introduced notation introduced, the exponent of the F-th model component pdf, corresponding to i-th configuration of observed discrete data can be written

$$b'^F_i c^F_i = b'\ I(\nu_b,\ \underline{\delta}(F)) I(\mu_c,\ \underline{\delta}(i))\ c$$

where

ν_b and μ_c are structure-vectors of parameters b and statistics c.

b and c are full parameter and statistics structures, ordered according to their structure-vectors (b'^F means $(b^F)'$).

$\underline{\delta}(F)$ and $\underline{\delta}(i)$ are δ-vectors choosing proper parameter and data for constructing the model of the active component for actually observed discrete data.

Remark 4 *By the product of I-matrices $I(\nu_b,\ \underline{1}) I(\mu_c,\ \underline{\delta}(i))$ we choose all sub-models corresponding to discrete data i and all components $F \in F^*$.*

4 Bayesian estimation

In this section we present the formal Bayes estimation of the target pdf (2). The resulting formulas are converted into a practicable algorithm in Sections 5, 6.

Constructing $p(d_t, F_t|d(t-1), \hat{F}_t)$
The investigated pdf (2) can be expressed by using the joint model (4) as

$$p(d_t, F_t|d(t-1), \hat{F}_t) = \int\int p(\alpha, \theta, d_t, F_t|d(t-1), \hat{F}_t)\, d\alpha\, d\theta. \tag{14}$$

Here and in the sequel $\int f(x)\, dx$ is the integral over all values of x for which the function $f(x)$ is nonzero and well defined. Let us assume that the joint pdf for the unknown parameters θ and α can be factorized

$$p(\theta, \alpha|d(t-1), \hat{F}_t) = p(\theta|d(t-1)\, p(\alpha|d(t-1)),$$

i.e. the parameters θ and α are conditionally independent. Later we show that this property is preserved during estimation. Under assumptions $\mathcal{A}2$ and $\mathcal{A}4$ and conditional independence of parameters α and θ, the integrated pdf in (14) can be decomposed as follows

$$\begin{aligned} &p(\theta, \alpha, d_t, F_t|d(t-1), \hat{F}_t) = \\ &\quad p(d_t|d(t-1), F_t, \hat{F}_t, \theta)\, p(F_t|\hat{F}_t, \alpha)\, p(\theta|d(t-1))\, p(\alpha|d(t-1)). \end{aligned} \tag{15}$$

After reordering factors in (15), substituting it in (14) and using Fubini theorem we obtain

$$p(d_t, F_t|d(t-1), \hat{F}_t) = \tag{16}$$

$$\int p(d_t|d(t-1), F_t, \hat{F}_t, \theta) p(\theta|d(t-1), \hat{F}_t)\, d\theta \int p(F_t|\hat{F}_t, \alpha) p(\alpha|d(t-1), \hat{F}_t)\, d\alpha. \tag{17}$$

The integral

$$\int p(d_t|d(t-1), F_t, \hat{F}_t, \theta)\, p(\theta|d(t-1), F_t, \hat{F}_t) d\theta = p(d_t|d(t-1), F_t, \hat{F}_t) \tag{18}$$

represents the observation model for the given fault-configuration F_t, and

$$\int p(F_t|\hat{F}_t, \alpha)\, p(\alpha|d(t-1), \hat{F}_t) d\alpha = p(F_t|d(t-1), \hat{F}_t) \tag{19}$$

is the fault-configuration model.

Parameter estimation
It follows from (17) and (19) that the product of these pdfs gives the needed pdf (2). Altogether, prior pdf of unknown parameters has to be updated according to the Bayes formula

$$p(\theta, \alpha|d(t)) = \frac{p(d_t|d(t-1), \hat{F}_t, \theta, \alpha)\, p(\theta, \alpha|d(t-1))}{p(d_t|d(t-1), \hat{F}_t)} \tag{20}$$

and then the integrals (18) and (19) can be computed.

Data prediction
The data predictor $p(d_t|d(t-1),\hat{F}_t)$ exploiting only observed data can be obtained by summation the pdf (14) over all fault configurations from the set F^*

$$p(d_t|d(t-1),\hat{F}_t)=\sum_{F_t\in F^*}p(d_t,F_t|d(t-1),\hat{F}_t).$$

The mixture form of the parameterized observation model (5) in the Bayes formula (20) makes its real time exact use impossible. In other words, the computational difficulties, we face and solve approximately in the subsequent sections, stem from the fact that the pdf does not possess natural conjugate. Consequently, its functional form does not reproduce within a feasible set while estimating unknown parameters θ,α. The pdf (18) and/or (19) are available if parameters θ and/or α are known, i.e. pdfs $p(\theta|d(t-1),\hat{F}_t)$ and $p(\alpha|d(t-1),\hat{F}_t)$ concentrate of a known point.

Fault-configuration estimation
The estimate of the current fault-configuration F_t based on data $d(t)$, i.e. with the current data d_t included, can also be constructed using the Bayes formula. It holds

$$p(F_t|d(t),\hat{F}_t)=p(F_t|d_t,d(t-1),\hat{F}_t)=\frac{p(d_t|d(t-1),F_t,\hat{F}_t)p(F_t|d(t-1),\hat{F}_t)}{p(d_t|d(t-1),\hat{F}_t)}. \tag{21}$$

The denominator in (21) is independent of F_t and thus we can (21) rewrite as follows

$$p(F_t|d(t),\hat{F}_t)\propto p(d_t|d(t-1),F_t,\hat{F}_t)p(F_t|d(t-1),\hat{F}_t)=p(d_t,F_t|d(t-1),\hat{F}_t) \tag{22}$$

where $\propto$ denotes proportionality. Therefore $p(F_t|d(t),\hat{F}_t)$ is proportional to the pdf (2).

This completes the formulas relating the parameterized models. Our aim is to *compute the pdf (2) at least approximately in the way applicable in on-line mode.* An algorithm for solving this problem is described in the next section.

5 Quasi-Bayes estimation: known observation model

We start with the simpler case when the observation model contains no unknown parameter θ. Various formulae below are simplified by using the δ-vector $\underline{\delta}$ in definition (9) and vector of ones $\underline{1}$ of appropriate dimension.

According to (18), (17), the needed pdf (2) is

$$p(d_t,F_t|d(t-1),\hat{F}_t)=p(d_t|d(t-1),F_t,\hat{F}_t)\int p(F_t|\hat{F}_t,\alpha)p(\alpha|d(t-1),\hat{F}_t)\,d\alpha. \tag{23}$$

The pdf $p(\alpha|d(t-1),\hat{F}_t)$ of the unknown α have to be updated in order to get the desired (approximately) Bayesian solution of the considered problem.

α-prior pdf
We assume that the prior pdf for α has the form of Dirichlet density [8]

$$p(\alpha|d(t-1), \hat{F}_t) \propto \prod_{F \in F^*} \prod_{\hat{F} \in \hat{F}^*} \alpha_{F|\hat{F}_t}^{n_{F|\hat{F}_t}(t-1)-1}, \tag{24}$$

determined by the $\overline{F}^2 \equiv (\overline{f}+1)^{2n}$-way table of positive statistics $n_{F|\hat{F}}(t-1)$. It can be written in an exponential form

$$p(\alpha|d(t-1), \hat{F}_t) \propto \exp\{\mathrm{tr}\ \lambda'(\alpha)[n(t-1) - \underline{1}\,\underline{1}']\}$$

where

$\lambda(\alpha)$ is $\overline{F}^2$-dimensional known matrix function of α with entries

$$\lambda_{F|\hat{F}} = \ln(\alpha_{F|\hat{F}}),\ F \in F^*,\ \hat{F} \in F^* \tag{25}$$

$n(t)$ is $\overline{F}^2$-dimensional matrix of statistics with entries $n_{F|\hat{F}}(t)$ "counting" the number of occurrences of the configurations $F_\tau|\hat{F}_\tau$ for $\tau \leq t$.

Parameterized fault-configuration model
To give the parameterized fault-configuration model the same multiplicative form as the α-prior pdf, we formally express

$$p(F_t|\hat{F}_t, \alpha) = \prod_{F \in F^*} \prod_{\hat{F} \in \hat{F}^*} \alpha_{F|\hat{F}}^{\delta(F,F_t)\delta(\hat{F},\hat{F}_t)}. \tag{26}$$

Using (9), we can rewrite (26) in the form

$$p(F_t|\hat{F}_t, \alpha) = \exp\{\,\underline{\delta}'(F_t)\lambda(\alpha)\,\underline{\delta}(\hat{F}_t)\} = \exp\{\mathrm{tr}\ \lambda'(\alpha)\,\underline{\delta}(F_t)\,\underline{\delta}'(\hat{F}_t)\}. \tag{27}$$

Computing the fault-configuration prediction
After evaluating the integral in (23) by using relations (24) and (26), we obtain the pdf (2) in the following form

$$p(d_t, F_t|d(t-1), \hat{F}_t) = p(d_t|d(t-1), F_t, \hat{F}_t)\frac{\underline{\delta}'(F_t)n(t-1)\,\underline{\delta}(\hat{F}_t)}{\underline{1}'n(t-1)\,\underline{\delta}(\hat{F}_t)}. \tag{28}$$

Updating the α-prior pdf
By the Bayes chain formula [9] we have

$$p(\alpha, F_t|d(t), \hat{F}_t) \propto p(d_t|d(t-1), F_t, \hat{F}_t)\exp\{\mathrm{tr}\ \lambda'(\alpha)[n(t-1) - \underline{1}\,\underline{1}' + \underline{\delta}(F_t)\,\underline{\delta}'(\hat{F}_t)]\} \tag{29}$$

where the δ-vector $\underline{\delta}(F_t)$ on the right-hand side of (29) is the only object that prevents updating because the true value of F_t is not observed. Its formally correct removal by marginalization leads to the constructed posterior pdf of a mixture form. Its propagation through time increases the number of terms too quickly and some approximation is needed.

Approximation
The quasi-Bayes estimation [3] simply replaces $\delta(F, F_t)$ by its expectation

$$\delta(F, F_t) \approx \mathcal{E}[\delta(F, F_t)|d(t), \hat{F}_t] = p(F_t = F|d(t), \hat{F}_t) \equiv w_F(t),$$

i.e. by the pdf $p(F_t|d(t), \hat{F}_t)$ given in (19) and proportional to pdf (28) (with the given $d(t)$ and $\hat{F}_t$).

Probability weights
Using this approximation, the posterior pdf of α preserves the form of the Dirichlet density with the matrix $n(\cdot)$ evolving according to the simple formula

$$n(t) = n(t-1) + w(t)\, \underline{\delta}'(\hat{F}_t) \tag{30}$$

where F-th entry of the $\bar{F}$-vector $w(t)$ contains probability of the fault-configuration F (conditioned on the data $d(t)$ and the fault-configuration estimate $\hat{F}_t$). Its entries are given by the formula (implied by (23) and (28))

$$w_F(t) = p(F_t = F|d(t), \hat{F}_t) \propto p(d_t|d(t-1), F_t = F, \hat{F}_t)\, \underline{\delta}'(F_t) n(t-1)\, \underline{\delta}(\hat{F}_t). \tag{31}$$

Note that

$$p(F_t = F|d(t-1), \hat{F}_t) = \frac{\underline{\delta}'(F_t) n(t-1)\, \underline{\delta}(\hat{F}_t)}{\underline{1}' n(t-1)\, \underline{\delta}(\hat{F}_t)} \tag{32}$$

is a prediction of the fault configuration occurrence. It can be used for constructing FDI strategy when the dynamic nature of (1) can be neglected and we are not willing or able to specify loss function in detail. The vector $w(t)$ consists of probabilities of fault configurations F_t when even d_t is included in condition. They are generally better indicators of the occurrence of fault configurations. If the selection of the fault-configuration estimate can be postponed then these weights can be used for the construction of $\hat{f}_t(d(t))$.

Practical usefulness of the final formulas (23), (30) and (31) depends heavily on our ability to provide the observation model $p(d_t|d(t-1), F_t = F, \hat{F}_t)$ indexed by the arbitrary fault-configuration $F_t = F$. This problem is addressed in the subsequent section.

6 Quasi-Bayes estimation: exponential family

The observation models $p(d_t|d(t-1), F_t, \hat{F}_t)$ is rarely available from a direct modeling. Thus, we have to find it through the parameter and fault estimation. We have restricted ourselves to the input-output models and thus to the parameter estimation. A slight generalization to state space models is possible when using Daume's class of feasible problems [10]. Discussion of this case is, however, out of scope of this research report.

Observation model
Now, let the observation model (18) be assumed unknown and have to be computed by integration of the product of parameterized observation model and prior pdf of θ (θ-prior pdf). Similarly as in the previous section, we express the parameterized observation model as a product

$$p(d_t|d(t-1), F_t, \hat{F}_t, \theta) \quad = \quad p(u_t|d(t-1), \hat{F}_t) p(y_t|u_t, \Psi_{t-1}, F_t, \hat{F}_t, \theta). \tag{33}$$

Here k is $\theta-$independent factor; $d_t = (y_t, u_t)$ where y_t denotes the output of the system, i.e. the predicted part of the data item d_t; u_t denotes the input of the system and it is generated without knowledge of θ and F_t. The couple $[u_t, \Psi_{t-1}]$ is often denoted by ψ_t.

The natural conditions of control [9] are supposed to be fulfilled. They express the fact, that both the parameter estimates and the control signal are constructed on the same information taken from past measured data. Consequently, the information about input u_t extracted from parameter estimates θ at time t adds nothing to information carried by data $d(t-1)$ and the parameters can be omitted in the condition of the first pdf on the right-hand-side of (33).

The observation models of components we will consider here belong to the *exponential family*, i.e. their model pdf can be, for each $F \in F^*$, expressed in the following exponential form

$$p(y_t|\psi_t, F_t, \hat{F}_t, \theta) = \exp\{b'^F(\theta) c^F(\Psi_t)\} \tag{34}$$

where

$b^F(\theta)$ is a vector function of parameters, concerning the F-th fault configuration;

$c^F(\Psi_t)$ is a vector data statistics, corresponding to F-th fault configuration concerning data $\Psi_t = [y_t, \psi_t']'$.

For dealing with a multi-model it is advantageous to use I-matrices introduced in (13). We suppose that our multi-model has $\overline{F}$ components, partitioning the vector parameter into $\overline{F}$ blocks $b = [b^1, b^2, ..., b^{\overline{F}}]'$. Each block is allowed to have different dimension and to correspond to differently constructed statistics.

The parameterized observation model from the exponential family and conjugate θ-prior pdf, i.e. the general term on the right-hand side of (33) can be expressed as follows

$$p(y_t|\psi_t, F_t, \hat{F}_t, \theta) = \exp\{b'(\theta) I(\nu_b, \underline{\delta}(F)) c(\Psi_t)\} \tag{35}$$

where

$b(\theta)$ is $\overline{b}$ dimensional known vector function of θ composed of b^F for $F \in F^*$;

$I(\cdot, \cdot)$ is I-matrix defined by (13);

ν_b is the structure vector introduced in (6);

$\underline{\delta}(F)$ is the δ-vector defined by (9);

$c(\Psi_t)$ is a known $\overline{c}$-dimensional vector function of the data vector Ψ_t composed of c^F for $F \in F^*$;

Ψ_t is data vector that can be recursively evaluated using the observed data $d(t)$ and the fault-configuration estimate $\hat{F}_t$.

(θ, α)-prior pdf
The pdf of the unknown parameters (θ, α) is assumed to be of the form

$$p(\theta, \alpha | d(t-1), \hat{F}_t) \propto \exp\{b'(\theta) C(t-1)\} \exp\{\mathrm{tr}\ \lambda'(\alpha)(n(t-1) - \underline{1}\,\underline{1}')\} \tag{36}$$

where $C(t-1)$ is the $(\overline{b}, \overline{F})$ dimensional matrix of statistics, containing the prior information and that brought by the past observed data. Then similarly as in the previous section we can perform the prediction and the data update of statistics.

Computing the fault-configuration prediction

$$p(d_t, F_t | d(t-1), \hat{F}_t) \propto$$
$$\propto \int \exp\{b'(\theta)[C(t-1) + I(\nu_b, \underline{\delta}(F)) c(\Psi_t)]\} d\theta \ \frac{\underline{\delta}'(F_t) n(t-1)\, \underline{\delta}(\hat{F}_t)}{\underline{1}' n(t-1)\, \underline{\delta}'(\hat{F}_t)}. \tag{37}$$

Updating the α-prior and θ-prior pdfs
The Bayes rule and natural conditions of control [9] imply

$$\begin{aligned} p(\alpha, \theta, F_t | d(t), \hat{F}_t) \quad &\propto \quad \exp\{b'(\theta)[C(t-1) + I(\nu_b, \underline{\delta}(F)) c(\Psi_t)]\} \\ &\qquad \exp\{\mathrm{tr}\ \lambda'(\alpha)[n(t-1) - \underline{1}\,\underline{1}' + \underline{\delta}(F_t)\, \underline{\delta}'(\hat{F}_t)]\} \end{aligned} \tag{38}$$

Approximation
Again, the correct removal of the unknown matrix $\underline{\delta}(F_t)$ would lead to computational burst. For this reason, we use the approximation

$$\delta(F, F_t) \approx p(F_t = F | d(t), \hat{F}_t)$$

for the known $d(t)$ and $\hat{F}_t$. It can be computed as the marginal pdf of (38).

Probability weights

$$w_F(t) = p(F_t = F | d(t), \hat{F}_t) \propto$$
$$\propto \mathcal{I}(I(\nu_b, \underline{\delta}(F))[C(t-1) + c(\Psi_t)])\, \underline{\delta}'(F)(n(t-1) + \underline{1}\,\underline{1}')\, \underline{\delta}(\hat{F}_t) \tag{39}$$

where

$$\mathcal{I}(I(\nu_b, \underline{\delta}(F))C) \equiv \int \exp[b'(\theta) I(\nu_b, \underline{\delta}(F))C] \, d\theta^F.$$

Substituting approximation of the weight (39) into the I-matrix $I(\nu_b, \underline{\delta}(F))$ in (38), we reproduce approximately the form of the parameter distribution. The key influence of the approximation adopted is that the increments of the statistics $c(\Psi_t)$ are added with the weight (39). Having the (approximate) pdf $p(\theta|d(t), \hat{F}(t))$ we can simply compute the needed observation model (17).

Let us summarize the results found.

Quasi-Bayes estimation: multi-model exponential distribution family

The quasi-Bayes algorithm consists of two stages: it is initialized during the preparatory stage and then continues with the on-line stage.

Preparatory stage

1. Choose the parameterized models, i.e. specify the form of the functions $b(\theta)$, $c(\Psi)$ determining the members of the exponential family (34) and the form of the vector Ψ.
2. Specify the I-matrix $I(\nu_b, \underline{\delta}(F_t))$ determining structure of the mixture model.
3. Specify initial values of the statistics $C(0)$.
4. Specify "counters" $n_{F|\hat{F}}(0)$ for all $F \in F^*$ and $\hat{F} \in \hat{F}^*$ with positive items
5. Set time $t = 1$
6. Select the first input u_1

On-line stage

1. Compute the characteristics of the pdf,

$$p(y_t, F_t|\psi_t, \hat{F}_t) \propto \mathcal{I}\left(I(\nu_b, \underline{\delta}(F))(C(t-1) + c(\Psi_t))\right) \;\; \underline{\delta}'(F_t) n(t-1) \, \underline{\delta}(\hat{F}_t)$$

   if need be
2. Select the prediction $\hat{F}_t$ of the fault configuration
3. Observe the output y_t
4. Evaluate the observed value of the data vector Ψ_t
5. Make "trial" updating of the following statistics

$$\begin{aligned} \tilde{C}(t) &= C(t-1) + I(\nu_b, \underline{1}) c(\Psi_t) \\ \tilde{n}(t) &= n(t-1) + \underline{1}\, \underline{\delta}'(\hat{F}) \end{aligned}$$

6. Compute the weights $w_F(t)$ for all $F \in F^*$ by the formula

$$w_F(t) = \frac{\mathcal{I}(I(\nu_b, \underline{\delta}(F_t))\tilde{C}(t)) \;\; \underline{\delta}'(F_t)\tilde{n}(t)\, \underline{\delta}(\hat{F}_t)}{\sum_{F \in F^*} \mathcal{I}(I(\nu_b, \underline{\delta}(F))\tilde{C}(t)) \;\; \underline{\delta}'(F)\tilde{n}(t)\, \underline{\delta}(\hat{F}_t)}$$

7. Select the filtered estimate $\hat{F}(d(t))$ of F_t (if need be).
8. Compute for the chosen fault prediction $\hat{F}_t$ the statistics

$$\begin{aligned} C(t) &= C(t-1) + I(\nu_b, w(t))c(\Psi_t) \\ n(t) &= n(t-1) + w(t)\,\underline{\delta}'(\hat{F}_t) \end{aligned}$$

9. Compute the characteristics of the pdf

$$p(\theta, \alpha | d(t), \hat{f}(t)) \propto \exp\{b'(\theta) I(\nu_b,\, \underline{\delta}(F)) C(t)\} \exp\{\mathrm{tr}\ \lambda'(\alpha)(n(t) - \underline{1}\underline{1}')\}$$

   if need be
10. Increase the time counter
11. Choose the next system input u_t
12. Return to the beginning of the on-line stage.

It is worth noticing that the presented algorithm for the solution of the stated problem can be easily extended to the case of slowly varying parameters by adopting stabilized forgetting [11]. Then the recursions for statistics are modified for all $F \in F^*$, chosen $\hat{F}_t$ and forgetting factors $\varphi \in [0, 1]$ as follows

$$\begin{aligned} C(t) &= C(t-1) + w(t) I(\nu_b,\, \underline{\delta}(F)) c(\Psi_t) + (1-\varphi) C(0) \\ n(t) &= n(t-1) + w(t)\,\underline{\delta}'(\hat{F}_t) + (1-\varphi) n(0) \end{aligned}$$

7 Example – mixture model for mixed data

In this section we investigate the particular case of the parameterized observation model which is static with F_t of the length 0, i.e. $F_t = f_t$.

We look in some detail at an instructive example. Let us simulate real-like case of a control of rolling mill. The controlled output y_t is the thickness of the produced metal sheet. At the time t the controlled system can exhibit one of the two possible faults

$$f_t = 0 \text{ (non-fault)}, \quad f_t = 1 \text{ (fault)}$$

(i.e. $f^* = \{0; 1\}$, $\bar{f} = 2$). An operator observes the system and estimates its behavior. He can choose his estimate from the set f^* of all possible faults:

$$\hat{f}_t = 0 \quad \text{or} \quad \hat{f}_t = 1.$$

As an action in dependence on his estimate $\hat{f}_t$ the operator switches between two types of control:

$$u_t = 0 \text{ (automatic)} \quad \text{or} \quad u_t = 1 \text{ (manual)},$$

for $u \in u^* = \{0; 1\}, \;\; \bar{u} = 2.$

The continuous output y_t and discrete input u_t form the mixed data pair $d_t = [y_t, u_t]$. The parameterized observation model corresponding to the fault f_t is static, $n = 0$, thus

$$p(d_t|d(t-1)F_t, \hat{F}_t, \theta) = p(d_t|f_t, \theta) = p(y_t|u_t, f_t, \theta)p(u_t|\hat{f}_t)$$

where the parameterized model of the output y_t is supposed to be normal

$$p(y_t|u_t, \theta) \sim \mathcal{N}(\mu_{u_t|f_t}; \sigma^2_{u_t|f_t}). \tag{40}$$

The model of the operator actions (fulfilling the natural conditions of control) is a two-way table of the fix elements $\beta_{u_t|\hat{f}_t}$ representing at time t the probability corresponding with individual values of pairs $u_t, \hat{f}_t$

$$p(u_t|\hat{f}_t) = \beta_{u_t|\hat{f}_t}.$$

In the presented example we have

	fault estimate: non-fault $\hat{f}_t = 0$	fault estimate: fault $\hat{f}_t = 1$
control: automatic $u_t = 0$	$\beta_{0\|0} = 0.6$	$\beta_{0\|1} = 0.2$
control: manual $u_t = 1$	$\beta_{1\|0} = 0.4$	$\beta_{1\|1} = 0.8$

The parameterized fault-configuration model defined in (3) takes in this example the form

$$p(f_t|\hat{f}_t, \alpha) = \alpha_{f_t|\hat{f}_t}$$

and it is represented again by the two-way table of coefficients $\alpha_{f_t|\hat{f}_t}$

	fault estimate: non-fault $\hat{f}_t = 0$	fault estimate: fault $\hat{f}_t = 1$
non-fault $f_t = 0$	$\alpha_{0\|0}$	$\alpha_{0\|1}$
fault $f_t = 1$	$\alpha_{1\|0}$	$\alpha_{1\|1}$

These probabilities are unknown and have to be estimated.

The fault-configuration model (27) is

$$p(f_t|\hat{f}_t, \alpha) = \exp\{\mathrm{tr}\;\; \underline{\delta}'(f_t)\lambda(\alpha)\,\underline{\delta}(\hat{f}_t)\}$$

with

$$\lambda = \begin{bmatrix} \ln(\alpha_{1|1}) & \ln(\alpha_{1|2}) \\ \ln(1-\alpha_{1|1}) & \ln(1-\alpha_{1|2}) \end{bmatrix}$$

and

$$\underline{\delta}(f_t) = \begin{bmatrix} \delta(1, f_t) \\ \delta(2, f_t) \end{bmatrix}; \qquad \underline{\delta}(\hat{f}_t) = \begin{bmatrix} \delta(1, \hat{f}_t) \\ \delta(2, \hat{f}_t) \end{bmatrix}$$

The prior pdf for the parameter α is assumed to be Dirichlet in the exponential form

$$p(\alpha|d(t-1),\hat{f}_t) \propto \prod_{f\in f^*}\prod_{\hat{f}\in\hat{f}^*} \alpha_{f|\hat{f}_t}^{n_{f|\hat{f}_t}(t-1)-1} = \exp\{\mathrm{tr}\ \lambda'(\alpha)(n(t-1)-\underline{1}\underline{1}')\}$$

where the statistics $n(t-1)$ has the matrix form

$$n(t-1) = \begin{bmatrix} n_{1,1} & n_{1,2} \\ n_{2,1} & n_{2,2} \end{bmatrix}.$$

The parameterized observation model according to (40) has the form of normal density and it can be expressed in the exponential form (35)

$$p(y_t|u_t,f_t,\theta) \propto \frac{1}{\sigma_{u_t|f_t}}\exp\{-\frac{1}{2\sigma^2_{u_t|f_t}}(y_t-\mu_{u_t|f_t})^2\} \tag{41}$$

with the parameters $\theta = [\mu_{u_t|f_t}, \sigma^2_{u_t|f_t}]'$. It depends not only on the fault $f \in f^*$ but also on the discrete control variable $u \in u^*$. This dependence induces further decomposition of components to sub-models, each of them valid for a particular value of the discrete data u_t. So, the structure of each component can be similarly expressed by a structure vector ν^F defined in (7) which is now determined by values of u_t. For each component $F \in F^*$ we define

$$\nu^F = [\overline{(c_1^F)}, \overline{(c_2^F)}, ..., \overline{(c_{\overline{c}}^F)}]'$$

where $\overline{(c_i^F)},\ i = 1,2,...,\overline{c^F}$ are dimensions of statistics corresponding to the i-th configuration of discrete data within the F-th model component. The structure multi-vector μ_c defined in (8) is composed of these structure vectors ν^F.

The exponential form of the observation model pdf (34) is

$$p(y_t|u_t,f_t,\theta) \propto \exp\{b'(\theta)I(\nu_b,\,\underline{\delta}(F))I(\mu_c,\,\underline{\delta}(u))c(y_t)]\}.$$

The parameter structure is

$$b = [b'^1, b'^2]' = [[b_1^1, b_2^1], [b_1^2, b_2^2]]'$$

with

$$b_u^f = \left[-\frac{1}{2\sigma^2_{u|f}},\ \frac{\mu_{u|f}}{\sigma^2_{u|f}},\ -\frac{\mu^2_{u|f}}{2\sigma^2_{u|f}} - \ln(\sigma_{u|f})\right]',\ u \in u^*,\ f \in f^*. \tag{42}$$

Similarly the structure of the statistic is

$$c = [c_1'^1, c_2'^1, c_1'^2, c_2'^2]'$$

with

$$c_u^f(y(t)) = \left[y_t^2,\, y_t,\, 1\right]',\ u \in u^*,\ f \in f^*. \tag{43}$$

The exponential form of the observation model can be finally written

$$p(y_t|u_t, f_t, \theta) \propto \exp\{b'(\theta)[I(\nu_b, \underline{\delta}(f))I(\mu_c, \underline{\delta}(u))]c(y_t)\}$$

where (denoting $f_i = \delta(i, f),\ u_i = \delta(i, u),\ i = 1, 2$)

$$I(\nu_b, \underline{\delta}(f)) = \mathrm{diag}\{f_1, f_1, f_1, f_1, f_1, f_1, f_2, f_2, f_2, f_2, f_2, f_2\} \tag{44}$$

$$I(\mu_c, \underline{\delta}(u)) = \mathrm{diag}\{u_1, u_1, u_1, u_2, u_2, u_2, u_1, u_1, u_1, u_2, u_2, u_2\} \tag{45}$$

are the I-matrices for given structure of the mixed data-model.

Remark 5 *As, in this example, the statistics is really dependent only on the observed output and neither on the control u nor the fault f we can reduce the statistics to the three dimensional vector (43). The product of the I-matrices is then a column composed of four diagonal matrices (using the introduced notation)*

$$\Delta_{f,u} = diag\{f_i u_j, f_i u_j, f_i u_j\},\ i = 1, 2;\ j = 1, 2.$$

The prior pdf for the parameter θ according to (36) is

$$p(\theta|d(t-1), \hat{f}_t) \propto \mathrm{e}^{b'(\theta)C(t-1)}$$

where
b is defined in (41) and

$$C = [C'_{1|1}, C'_{2|1}, C'_{1|2}, C'_{2|2}]', \quad \text{with} \quad C_{u|f} = [V_{u|f}, W_{u|f}, \kappa_{u|f}]'.$$

is the three dimensional vector of statistics.

Having introduced particular forms of models and corresponding prior pdfs we can apply the quasi-Bayes algorithm described in Section 5.

ALGORITHM

Preparatory stage

1. Choose the data-model, i.e. specify the parameter vector $b = [b_1^{'1}, b_2^{'1}, b_1^{'2}, b_2^{'2}]$' with items

$$b_u^f(\theta) = \left[-\frac{1}{2\sigma^2_{u|f}}, \frac{\mu_{u|f}}{\sigma^2_{u|f}}, -\frac{\mu^2_{u|f}}{2\sigma^2_{u|f}} - \ln(\sigma_{u|f})\right]', \quad u = 1, 2; \quad f = 1, 2;$$

the data vector and the statistics are common for all $f,\ u$

$$\Psi_t = y_t, \quad c(\Psi_t) = \left[y_t^2, y_t, 1\right]'.$$

The Δ-matrix $\Delta(\underline{\delta}(f_t), \underline{\delta}(u_t)) = I(\nu_b, \underline{\delta}(f_t))I(\mu_c, \underline{\delta}(u_t))$ determining structure of the data-model mixture is (for the same statistic common to all sub-models)

$$\Delta(\underline{\delta}(f_t), \underline{\delta}(u_t)) = \begin{bmatrix} \Delta_{1,1} \\ \Delta_{1,2} \\ \Delta_{2,1} \\ \Delta_{2,2} \end{bmatrix}, \qquad \Delta_{i,j} = \begin{bmatrix} \delta(i,f_t)\delta(j,u_t) & 0 & 0 \\ 0 & \delta(i,f_t)\delta(j,u_t) & 0 \\ 0 & 0 & \delta(i,f_t)\delta(j,u_t) \end{bmatrix}$$

and the fault-configuration-model with parameter vector is

$$\lambda_{f,\hat{f}} = \ln(\alpha_{f|\hat{f}}), \quad f = 1,2; \quad \hat{f} = 1,2.$$

2. Specify initial values of statistics

$$C(0) = [V_{1|1}(0), W_{1|1}(0), \kappa_{1|1}(0), V_{2|1}(0), W_{2|1}(0), \kappa_{2|1}(0),$$
$$V_{1|2}(0), W_{1|2}(0), \kappa_{1|2}(0), V_{2|2}(0), W_{2|2}(0), \kappa_{2|2}(0)]'$$

and

$$n(0) = \begin{bmatrix} n_{1,1}(0) & n_{1,2}(0) \\ n_{2,1}(0) & n_{2,2}(0) \end{bmatrix}, \; n_{i,j}(0) > 0, i,j = 1,2$$

3. Set $t = 1$

4. Select the input u_1

On-line stage

1. MAKE PREDICTION

2. Select prediction of the fault $\hat{F}_t = \hat{f}_t$

3. Observe y_t

4. Evaluate Ψ_t for the observed data ($\Psi_t = y_t$, trivial)

5. Make trial updating of the statistics $\tilde{C}(t)$ using the following recursions

$$\tilde{C}(t) = C(t-1) + \Delta(\underline{1}, \underline{\delta}(u_t)) \left[y_t^2, \, y_t, \, 1\right]'$$

which means: for given u_t and f_t update

$$\tilde{V}_{u_t|f_t}(t) = V_{u_t|f_t}(t-1) + y_t^2$$
$$\tilde{W}_{u_t|f_t}(t) = W_{u_t|f_t}(t-1) + y_t$$
$$\tilde{\kappa}_{u_t|f_t}(t) = \kappa_{u_t|f_t}(t-1) + 1$$

and similarly for $\tilde{n}(t)$

$$\tilde{n}(t) = n(t-1) + \underline{1}\,\underline{\delta}'(\hat{f}_t)$$

which means: for all $f_t \in \{0,1\}$ and given $\hat{f}_t$ update

$$\tilde{n}_{f,\hat{f}}(t) = n_{f,\hat{f}}(t-1) + \underline{1}\,\underline{1}'$$

6. Compute the weights $w_f(t)$ (probabilities of $f_t = f$ for $f \in f^*$ and actual u_t and $\hat{f}_t$) by the formula

$$w_f(t) = \frac{\mathcal{I}_{u_t,f}\tilde{n}_{f|\hat{f}}(t)}{\sum_{i \in f^*} \mathcal{I}_{u_t,i}\tilde{n}_{i|\hat{f}}(t)}$$

where

$$\mathcal{I}_{u,f} = \frac{1}{\tilde{\kappa}_{u,f}} \Gamma\left(\frac{\tilde{\kappa}_{u,f} - 3}{2}\right)\left(\frac{2}{\tilde{R}_{u,f}}\right)^{\frac{\tilde{\kappa}_{u,f}-3}{2}}, \quad \tilde{R}_{u,f} = \tilde{V}_{u|f} - \frac{\tilde{W}^2_{u|f}}{\kappa_{u,f}}$$

7. Select filtration-based estimate of the fault f_t; it is reasonable to take $\arg\max_f w_f$

8. Compute the statistics

$$C(t) = C(t-1) + w_f(t)\Delta(w(t), \underline{\delta}(u_t))\left[y_t^2, y_t, 1\right]'$$

$$\tilde{n}(t) = n(t-1) + w'(t)\,\underline{\delta}'(\hat{f}_t)$$

9. Compute the parameter estimates

$$\hat{\mu}(t) = \frac{W(t)}{\kappa(t)}, \qquad \hat{\sigma}^2(t) = \frac{R(t)}{\kappa(t)}, \qquad \hat{\alpha}_{f|\hat{f}}(t) = \frac{n_{f|\hat{f}}(t)}{\sum_{i \in f^*} n_{i|\hat{f}}(t)}$$

10. Choose the next input u_t

11. Compute prediction of the fault, i.e. recompute the vector $w_f,\ f \in f^*$ for newly updated statistic and predict the fault which corresponds with the largest item of w (it has the largest probability)

12. Return to the beginning of the on-line stage

Detailed evaluations are presented in Appendices.

8 Results of experiments

In picture d) of the following figures the true faulty state values are depicted (circles) and values of their estimates (crosses) in terms of probability weights, corresponding to estimated vector of probabilities w.

Figure 1. shows the ideal case of relatively detached faults when the particular states and data belonging them can be easily recognized. Moreover, the program starts with the full knowledge of the true parameters. A result is, that it can classify measured data and estimate the current state with no errors.

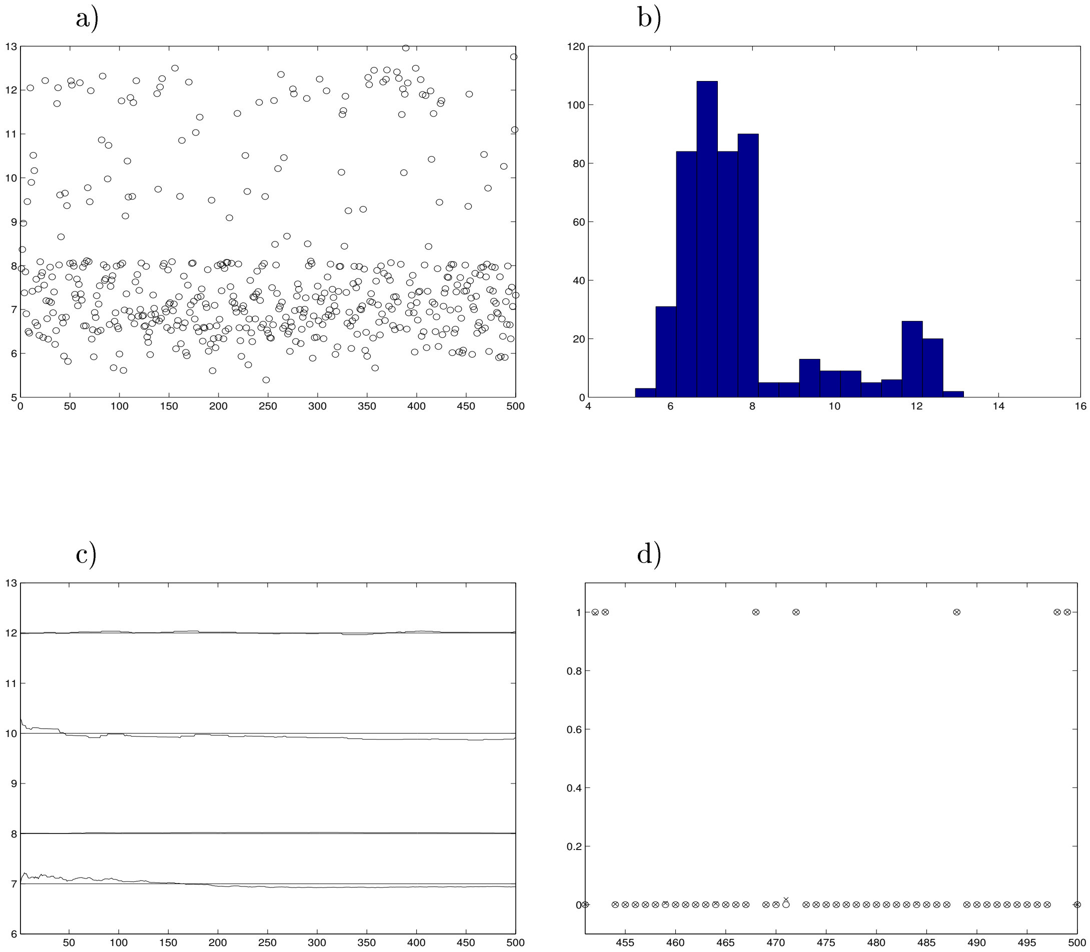

Fig. 1:
a) simulated data in time, b) histogram,
c) estimated parameters,
d) estimated state probabilities

Figure 2 represents almost the same case as the previous one but the dispersions of data belonging to particular components are slightly larger and the parameter estimation algorithm starts with initial parameter values which are not precisely true. The final parameter estimates are good as well as the current faulty state one. (In 500 sample simulation there were 13 mistakes, in the last 100 samples only 1.)

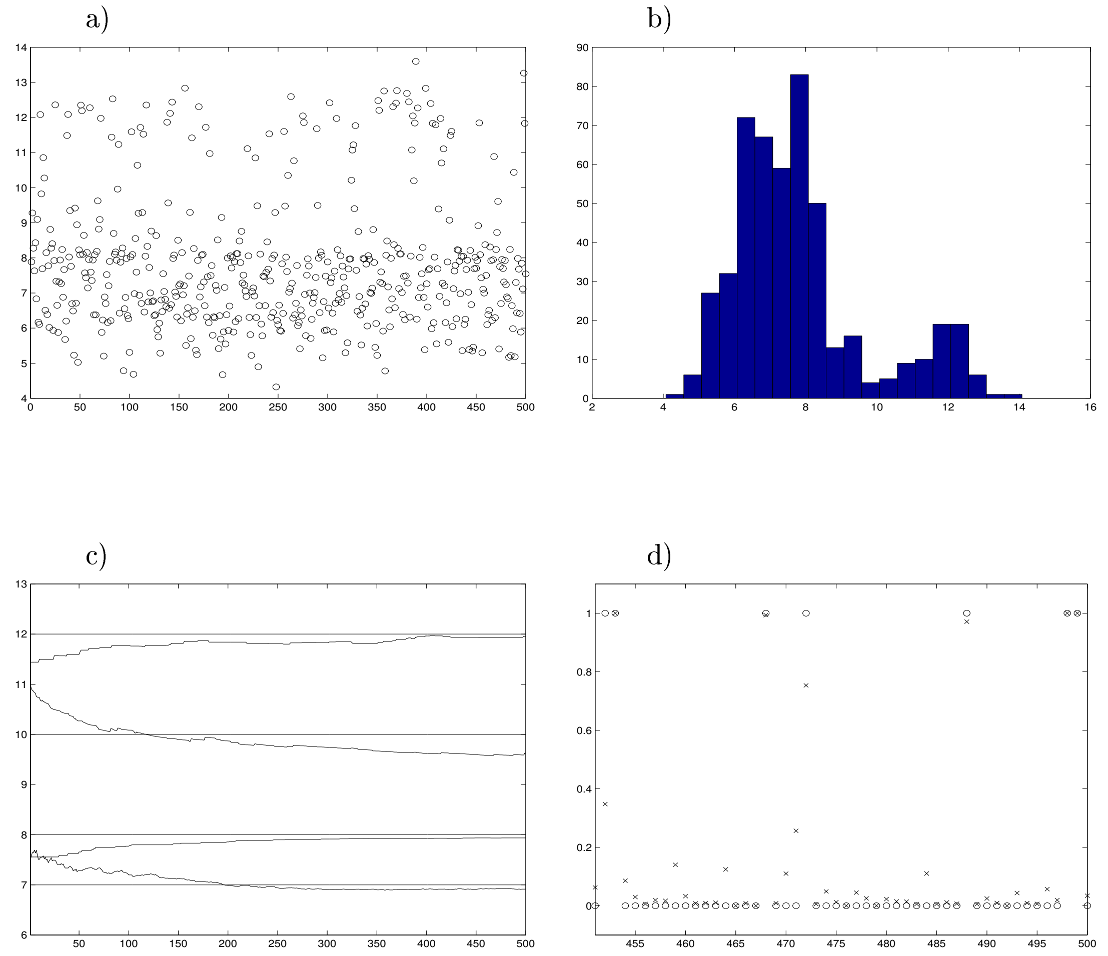

Fig. 2:
a) simulated data in time, b) histogram,
c) estimated parameters,
d) estimated state probabilities

Figure 3 exhibits still bigger data noise which makes particular faulty states more unrecognizable. This can be judged in the pictures a) and b), showing data in time and through the histogram. The estimates of parameters are worse but still sufficient for very good classification. The ratio of badly determined faulty state is 0.07 (at the end of simulation only 0.06). Notice, that due to overlapping of component pdfs the decisions about the true faulty states begin to be far from unique. Mostly are predicted both states but the true has bigger probability.

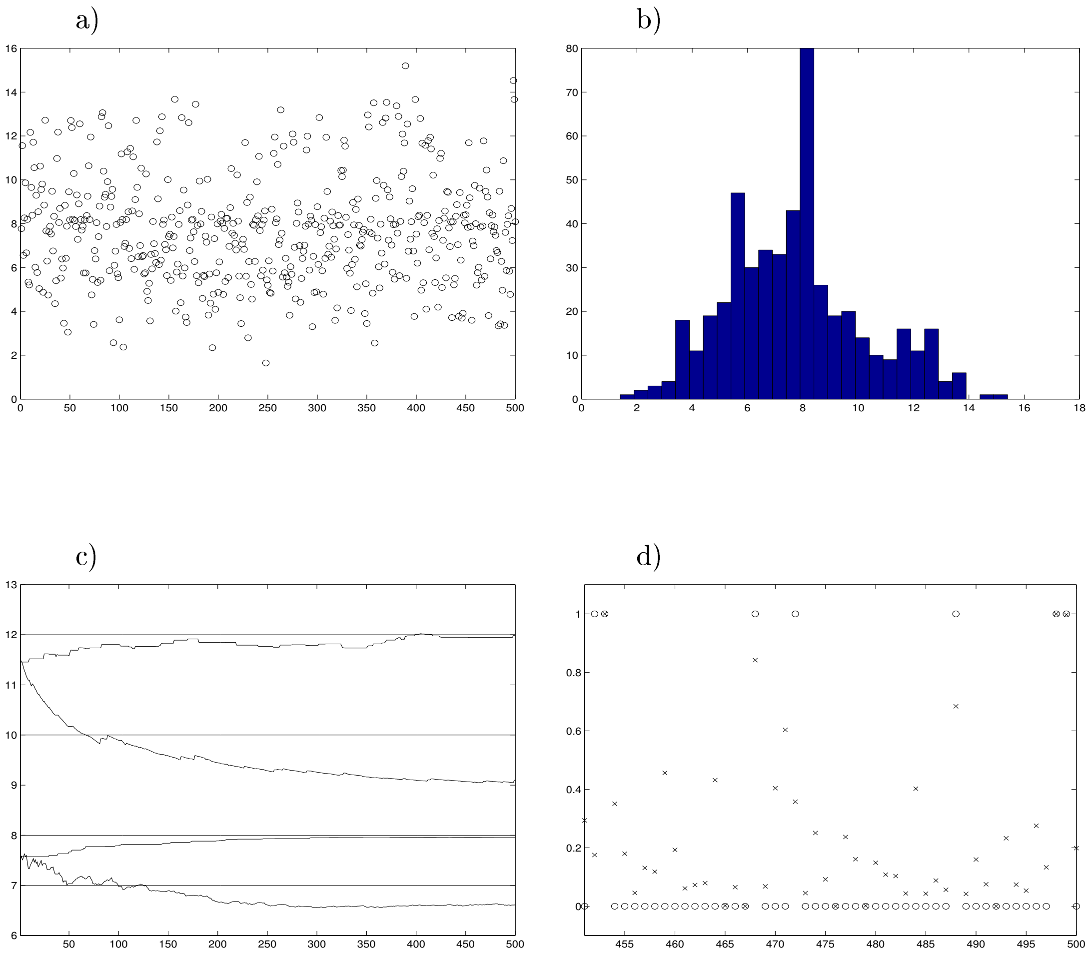

Fig. 3:
a) simulated data in time, b) histogram,
c) estimated parameters,
d) estimated state probabilities

The data in the figure 4 are afflicted with so large noise that the components cannot be practically distinguished. Due to the good (even if not precise) start of identification the estimates of important parameters are sufficiently good for faulty state predictions (the ratio of success is 0.17 through the whole simulation).

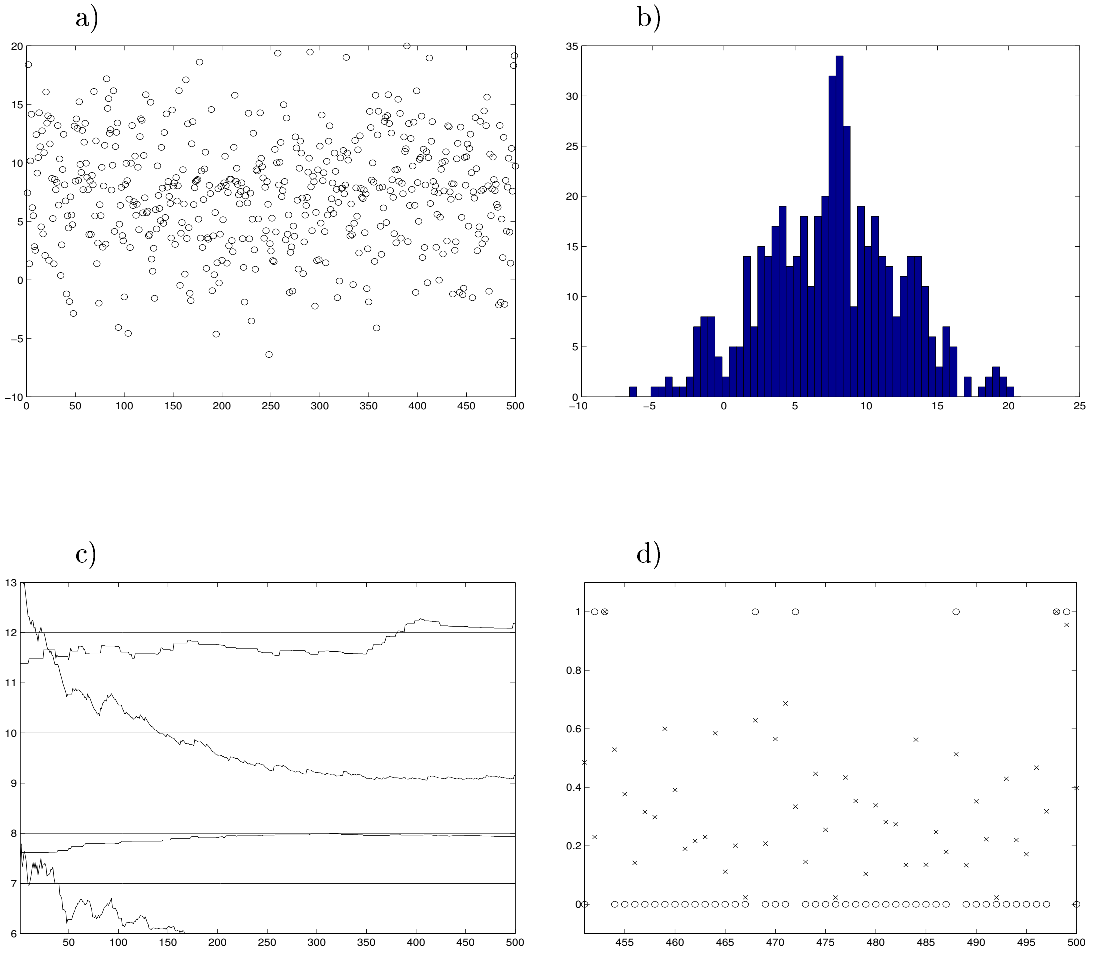

Fig. 4:
a) simulated data in time, b) histogram,
c) estimated parameters,
d) estimated state probabilities

Figure 5 starts another serial of demonstrations, now, for rather wrong start of estimation. At the beginning the noise is small. In this case the faulty state prediction is very good (ratio 0.018 during the simulation, 0 at the end).

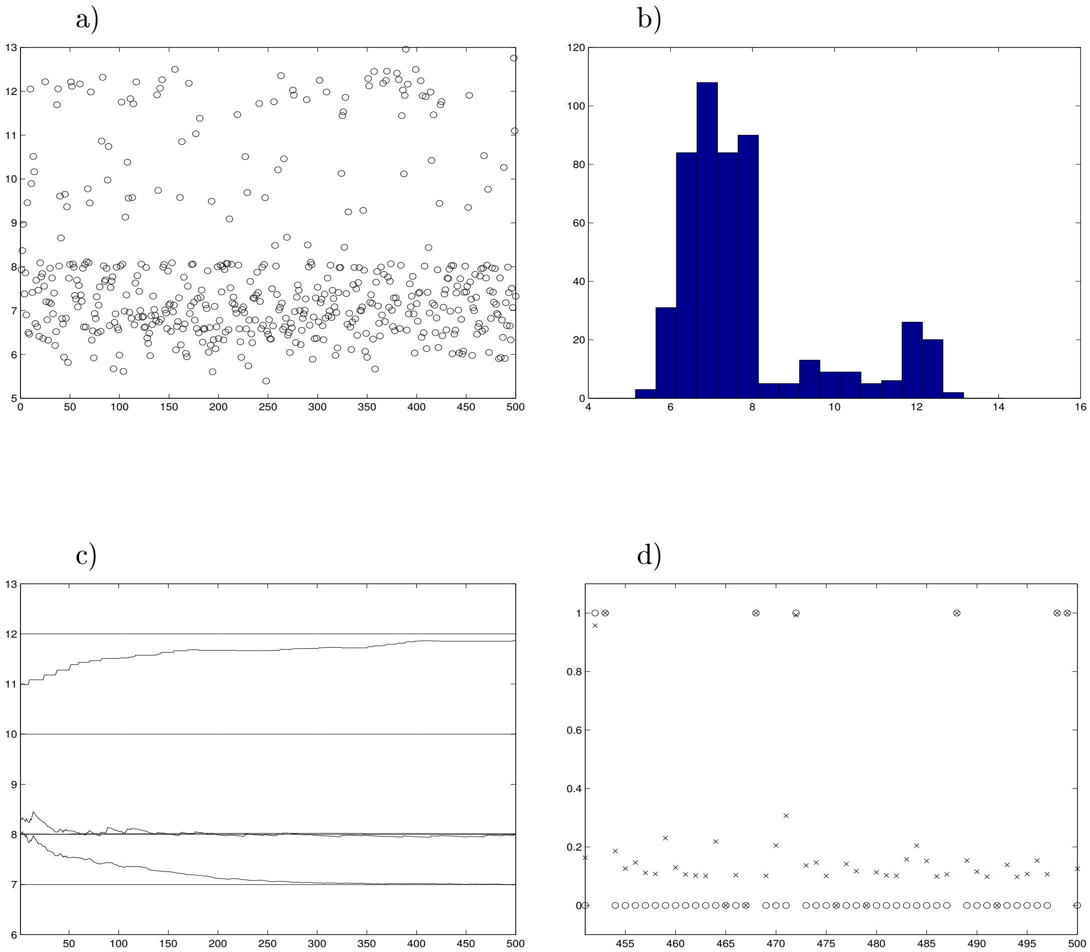

Fig. 5:
a) simulated data in time, b) histogram,
c) estimated parameters,
d) estimated state probabilities

Figure 6 continues the serial with bigger noise. Even here the faulty states predictions are good (ratio 0.02)

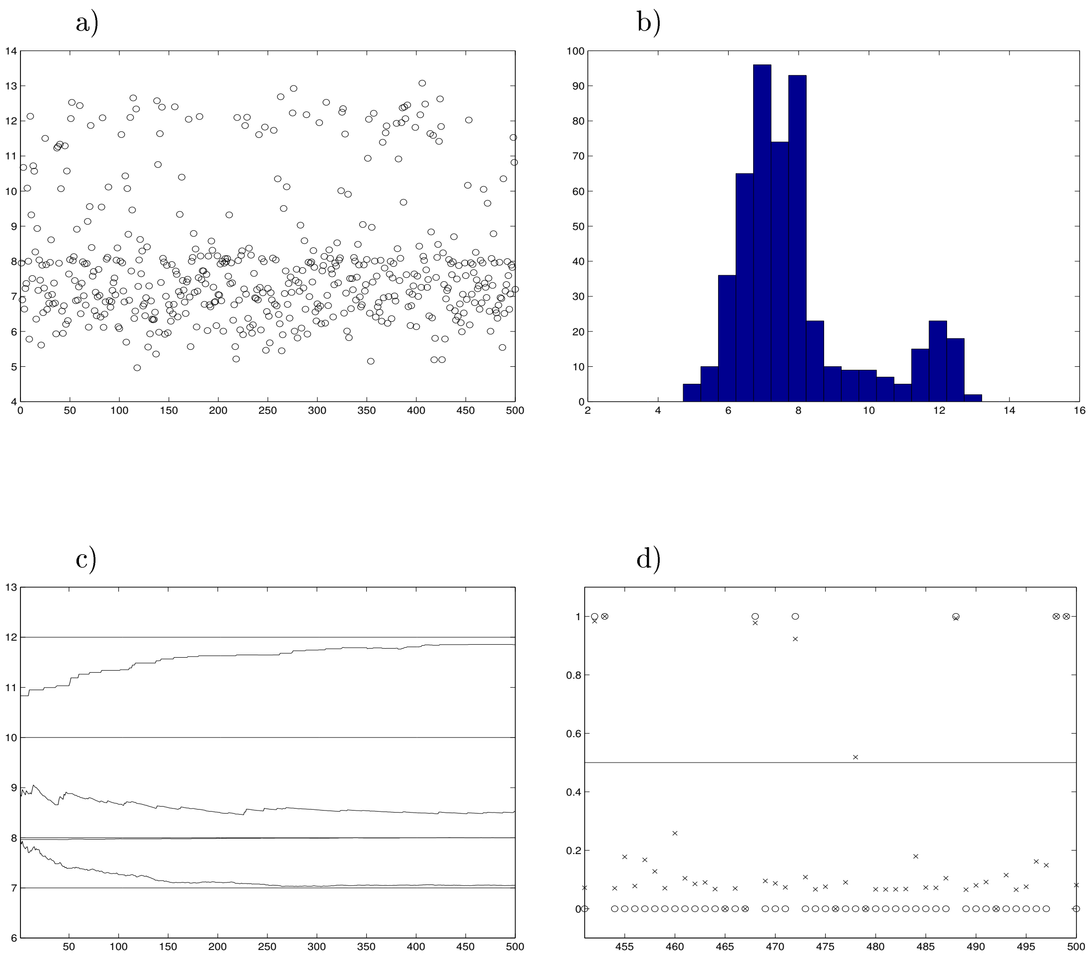

Fig. 6:
a) simulated data in time, b) histogram,
c) estimated parameters,
d) estimated state probabilities

Finally in figure 7 the noise is relatively big and its magnitude seems to be bordering for still good predictions (nevertheless, the ratio for such a case is still good 0.06, 0.04 at the end).

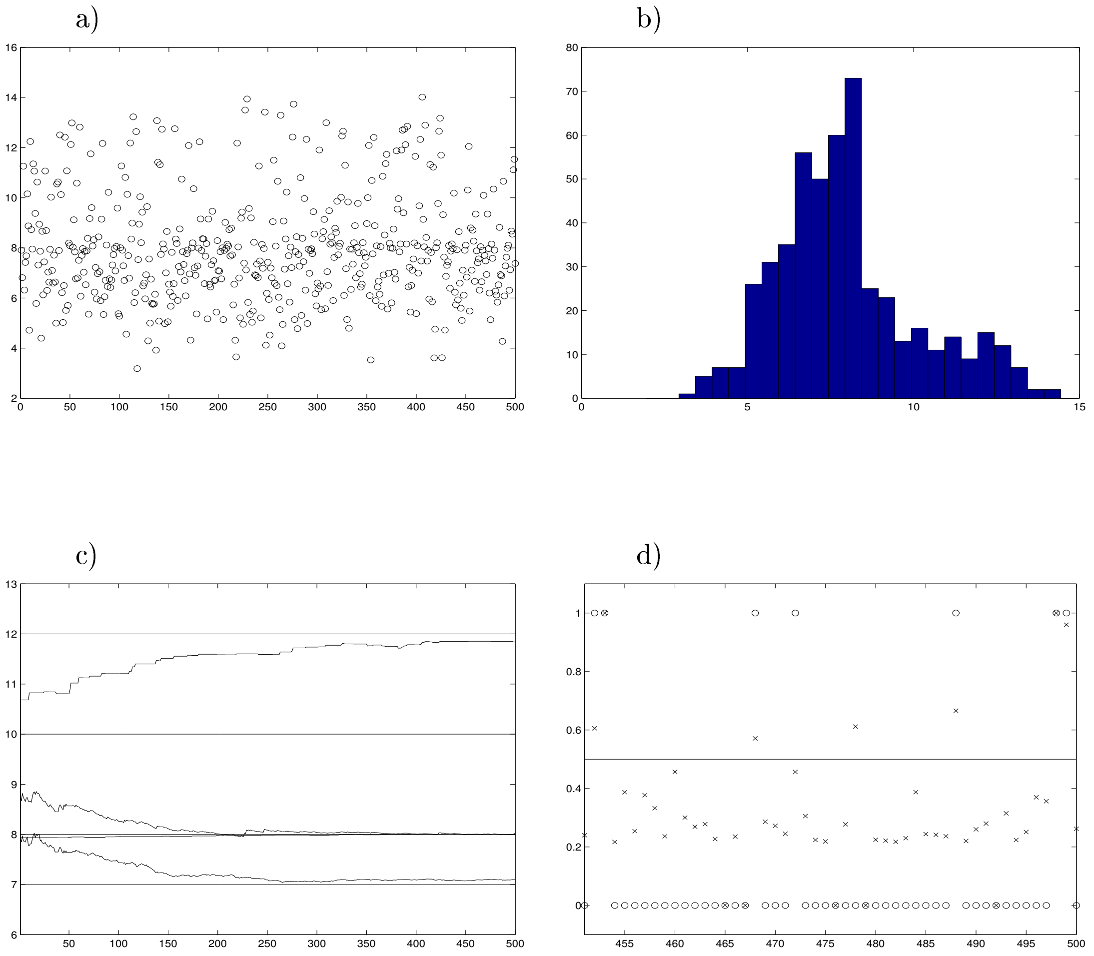

Fig. 7:
a) simulated data in time, b) histogram,
c) estimated parameters,
d) estimated state probabilities

The following simulations try to investigate the real-like case, when the the states of real fault can suffer from very rare number of occurrence. Figure 8 shows the case. In the picture b) it can be seen that the great majority of samples lay in one narrow group while the other few are spread around. The particular groups (up to the one) are not easily distinguishable. Picture c) shows, that some parameters are difficult to determine, due to lack of data. The classification is good (ratio 0.02).

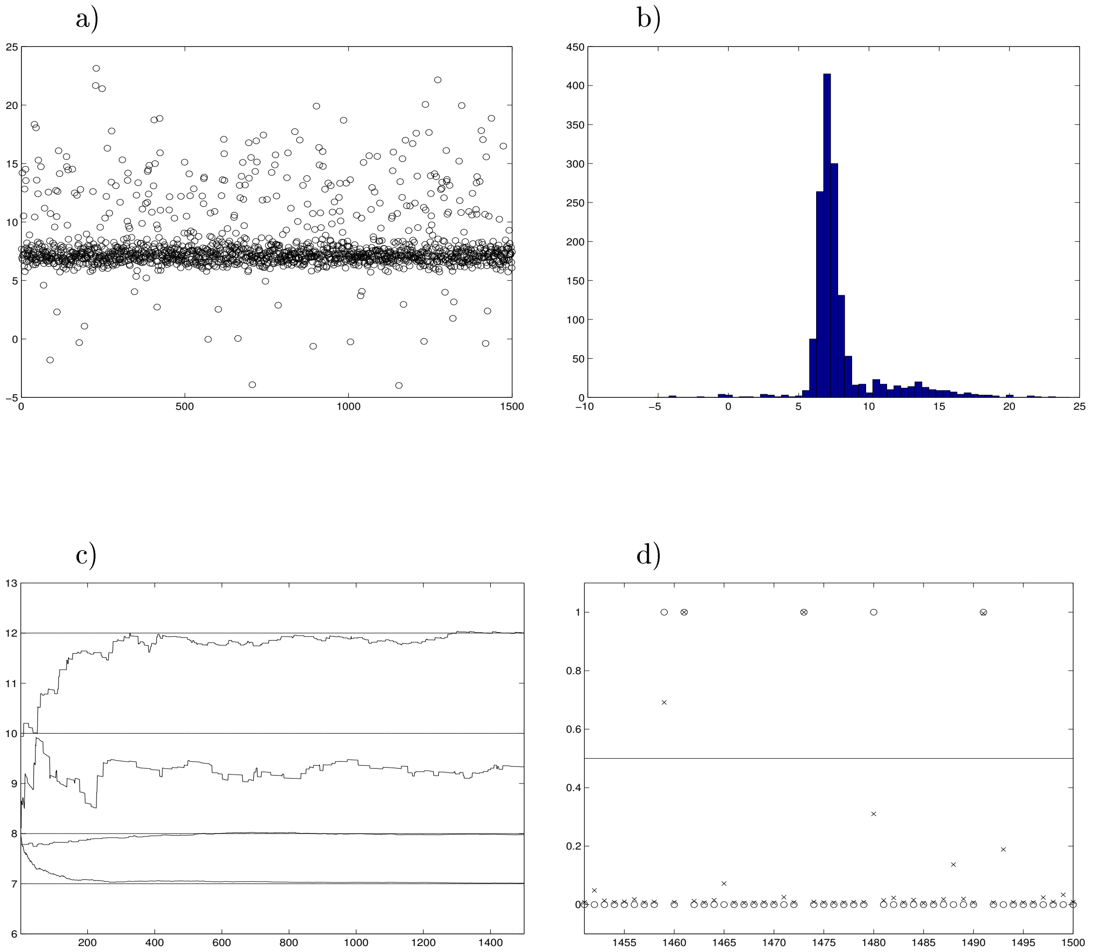

Fig. 8:
a) simulated data in time, b) histogram,
c) estimated parameters,
d) estimated state probabilities

Figure 9 presents more uniform distribution of data among the minority states but extreme dispersions of data. The ratio of classification was 0.08, at the end 0.04.

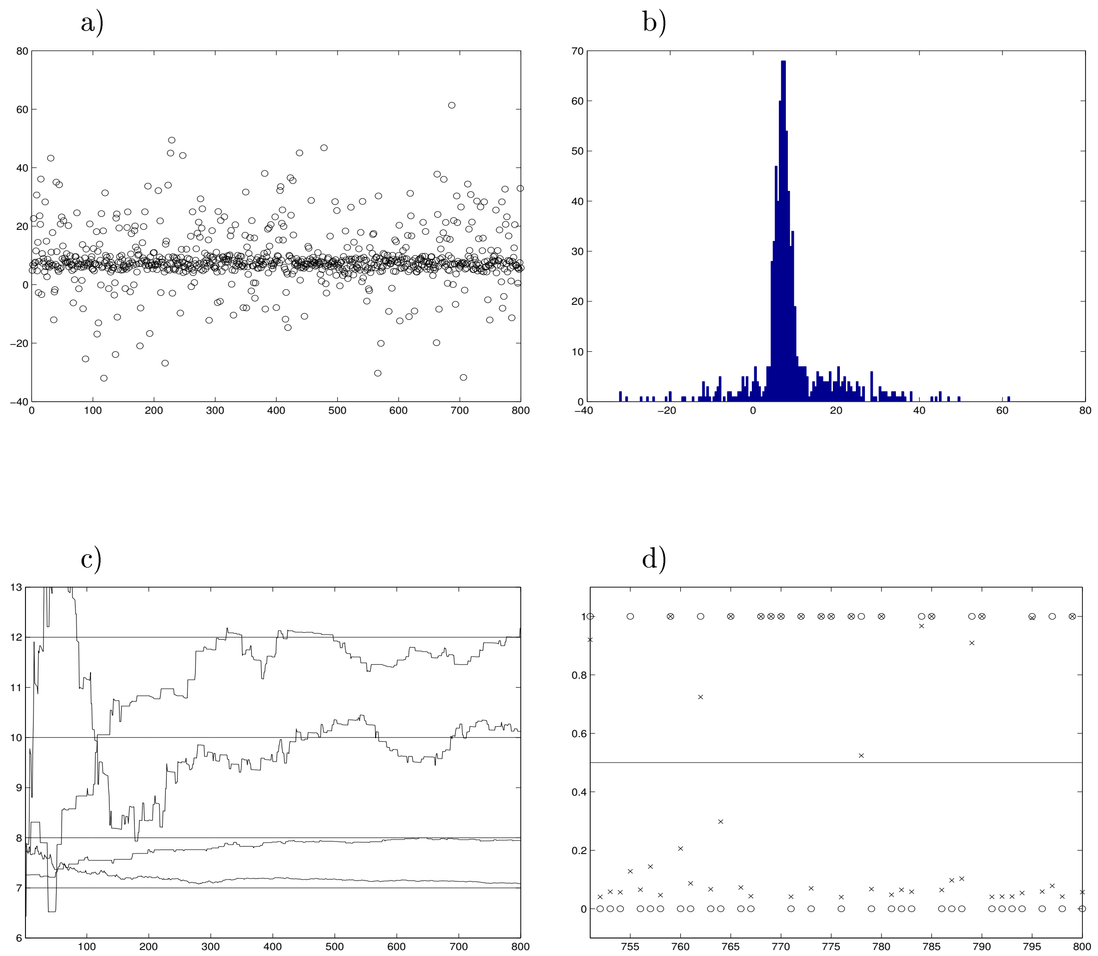

Fig. 9:
a) simulated data in time, b) histogram,
c) estimated parameters,
d) estimated state probabilities

Figure 10 tests still more extreme differences in number of occurrences between states with and without fault. The result is approximately the same as the previous one.

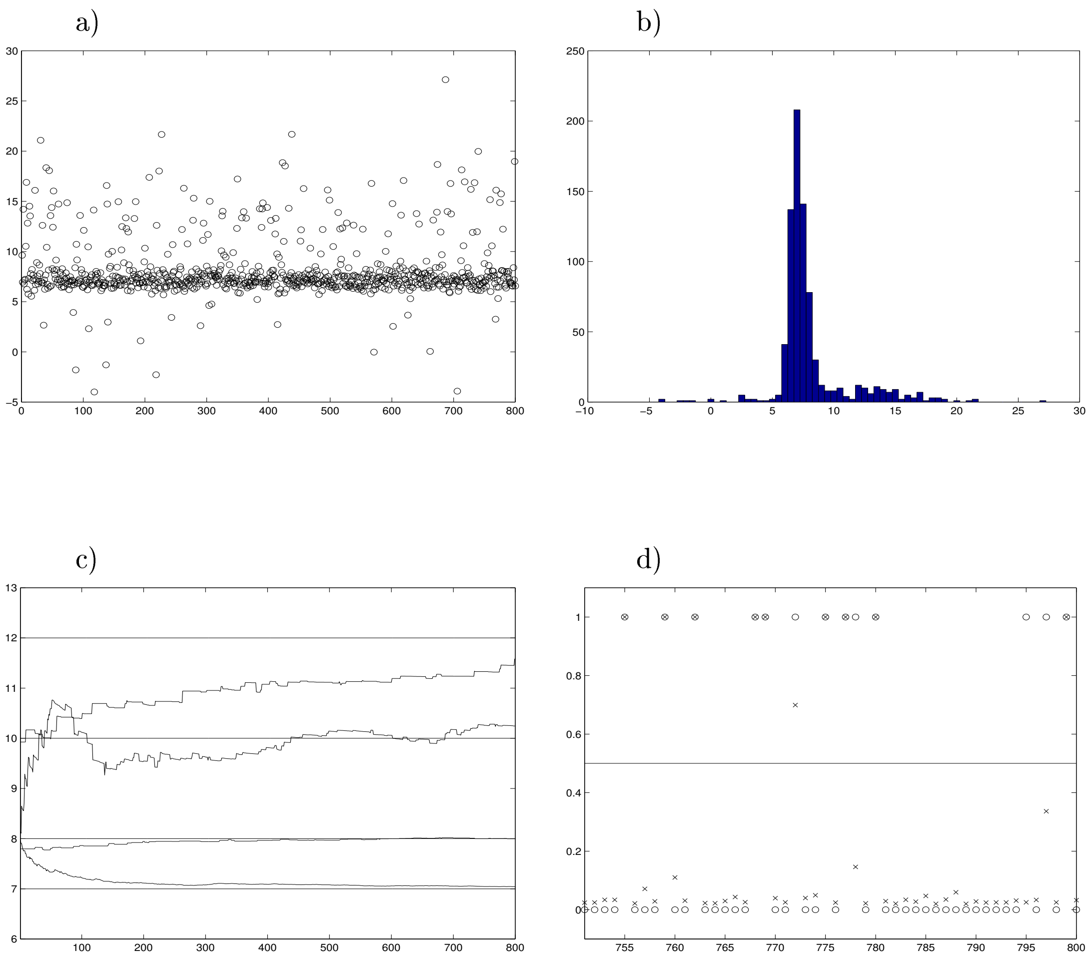

Fig. 10:
a) simulated data in time, b) histogram,
c) estimated parameters,
d) estimated state probabilities

Figure 11 tries to identify the parameters with minimum data used in some parts of the multi-model. It can be seen that the estimates of afflicted parameters are rather biased. The classification is still good, but if the deviations of parameter estimates cross some border, all could be destroyed.

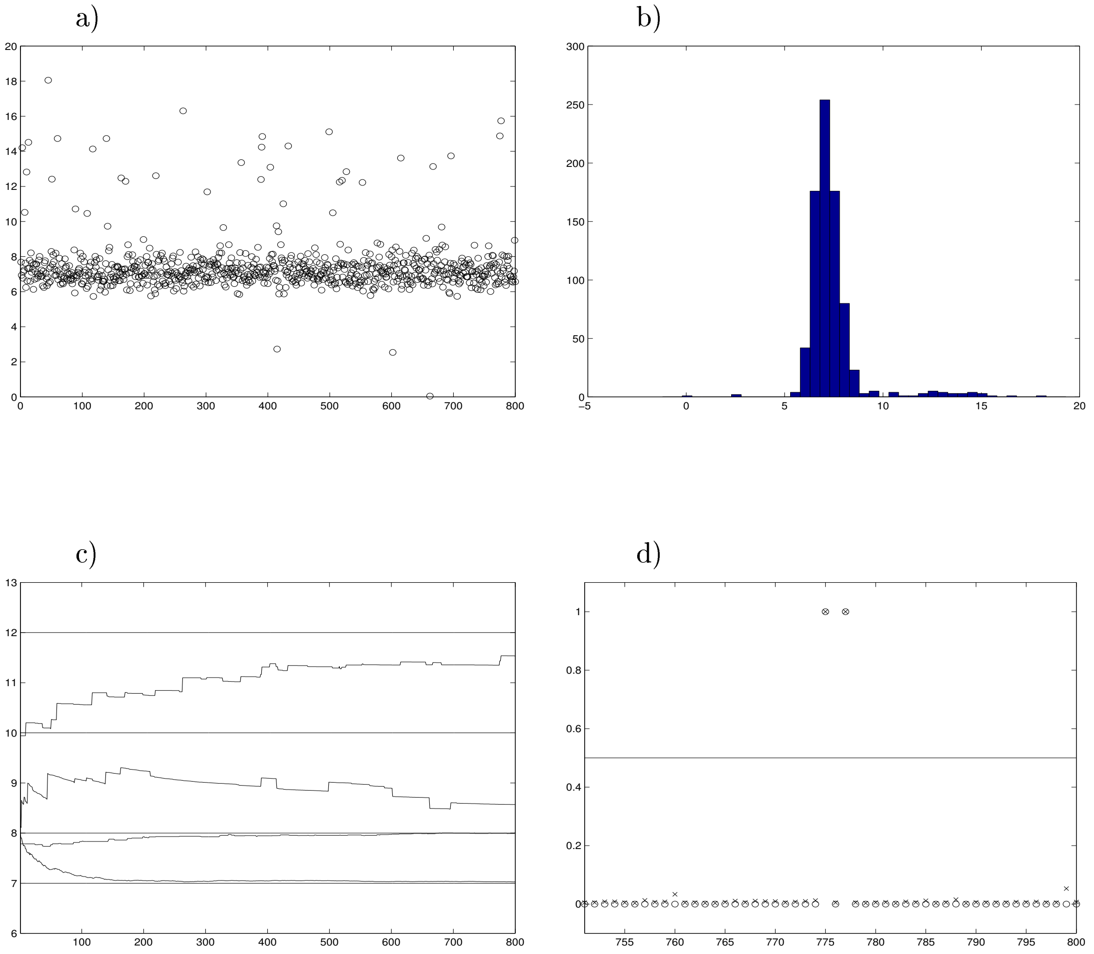

Fig. 11:
a) simulated data in time, b) histogram,
c) estimated parameters,
d) estimated state probabilities

Figure 12 closes the demonstrations with the case of extreme noise influencing the simulation with unbalanced frequencies of realization. The parameter estimates are on the border of applicability, the classification ratio is still 0.02 but utility of the algorithm at this case is disputable.

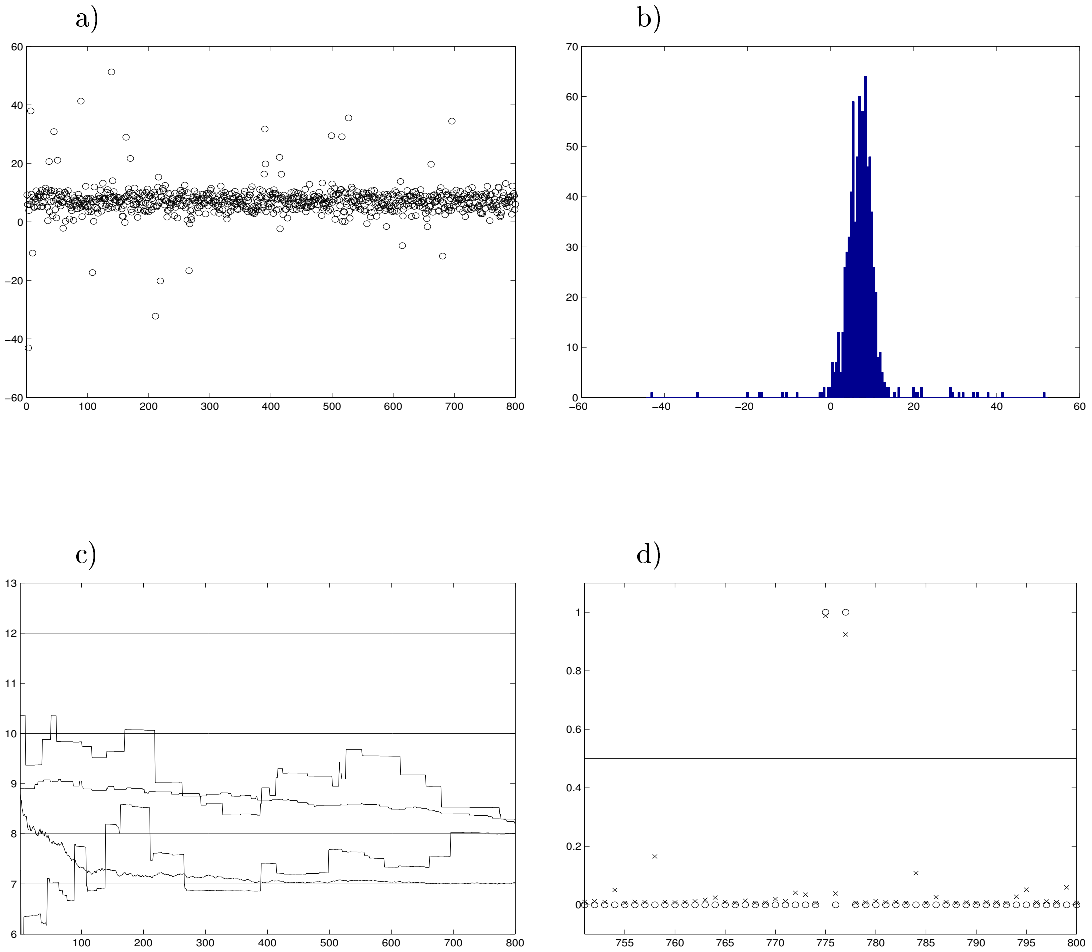

Fig. 12:
a) simulated data in time, b) histogram,
c) estimated parameters,
d) estimated state probabilities

9 Conclusion

In this research report we present the contribution to a solution of the complexity aspect of FDI. We proposed a quasi-Bayesian estimation algorithm for resolving the specific instance of the "curse of dimensionality" connected with use of the multi-model approach to FDI.

The main contributions are (i) adaptation of the novel quasi-Bayes algorithm for an efficient treatment of FDI problems; (ii) novel solution of on-line, possibly adaptive estimation of a dynamic mixture model that is able to cope both with continuous and discrete value data.

Acknowledgments

This work has been partially supported by grants from the Grant Agency of the Czech Republic (No102/99/1564), Grant Agency of the Academy of Sciences of the Czech Republic (No A2147701), from the Complex Research Project of the Academy of Sciences of the Czech Republic (No K1075601) and from the Grant of the Czech Ministry of Education (MSMT No VS96063 and No ME245).

References

[1] A.S. Wilsky, "A survey of design methods for failure detection in dynamic systems", *Automatica*, vol. 12, pp. 601–611, 1976.

[2] R. Isermann, "Process fault detection based on modeling and estimation – a survey", *Automatica*, vol. 20, pp. 387–404, 1984.

[3] M. Kárný, J. Kadlec, and E. L. Sutanto, "Quasi-Bayes estimation applied to normal mixture", in *Preprints of the 3rd European IEEE Workshop on Computer-Intensive Methods in Control and Data Processing*, J. Rojíček, M. Valečková, M. Kárný, and K. Warwick, Eds., Praha, September 1998, pp. 77–82, ÚTIA AV ČR.

[4] L. Berec, "A multi-model method to fault detection and diagnosis: Bayesian solution. An introductory treatise", *International Journal of Adaptive Control and Signal Processing*, vol. 12, no. 1, pp. 81–92, 1998.

[5] L. Berec, *Model Structure Identification: Global and Local Views. Bayesian Solution*, Ph.D. Thesis, Czech Technical University, Faculty of Nuclear Sciences and Physical Engineering, Prague, Czech Republic, 1998.

[6] D.M. Titterington, A.F.M. Smith, and U.E. Makov, *Statistical Analysis of Finite Mixtures*, John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore, 1985, ISBN 0 471 90763 4.

[7] O. Barndorff-Nielsen, *Information and exponential families in statistical theory*, Wiley, New York, 1978.

[8] T.S. Fergusson, "A Bayesian analysis of some nonparametric problems", *The Annals of Statistics*, vol. 1, pp. 209–230, 1973.

[9] V. Peterka, "Bayesian system identification", in *Trends and Progress in System Identification*, P. Eykhoff, Ed., pp. 239–304. Pergamon Press, Oxford, 1981.

[10] E.F. Daum, "New exact nonlinear filters", in *Bayesian Analysis of Time Series and Dynamic Models*, J.C. Spall, Ed. Marcel Dekker, New York, 1988.

[11] R. Kulhavý and M. B. Zarrop, "On general concept of forgetting", *International Journal of Control*, vol. 58, no. 4, pp. 905–924, 1993.

Appendices

Before presenting the derivation of fault-configuration prediction (Appendix A), construction of the observation model (Appendix B) and the parameter estimation (Appendix C) for the example presented in Section 6, we introduced some notations. For brevity, we use the following abbreviations:

$\delta_{f,i} = \delta(f_t; i), \quad \delta_i = \delta(f_t = i; i)$ (Kronecker δ-function)

$n_f = n_{f_t|\hat{f}_t}, \quad n_i = n_{f_t=i|\hat{f}_t};$

$\alpha_f = \alpha_{f_t|\hat{f}_t}, \quad \alpha_i = \alpha_{f_t=i|\hat{f}_t};$

$C_i(t) = [V_i(t), W_i(t), \kappa_i(t)]' = C_{u_t|f_t=i};$

$m_i(t) = m_{u_t|f_t=i};$

$R_i(t) = R_{u_t|f_t=i}$; of $m_i(t)$;

$b_f = [\mu_f, \sigma_f^2]' = b_{u_t|f_t}, \quad b_i = [\mu_i, \sigma_i^2]' = b_{u_t|f_t=i},$

$w_i = w_{f_t=i|\hat{f}_t};$

$c_t = c(y_t).$

Appendix A: Fault-configuration prediction

By using the above introduced notation, we can rewrite the integral (19) in the form

$$p(f_t = i|d(t-1), \hat{f}_t) \propto \int \prod_{f=1}^{2} \alpha_f^{\delta_{f,i}} \alpha_f^{n_f(t-1)-1} \mathrm{d}\alpha = \int \prod_{f=1}^{2} \alpha_f^{n_f(t-1)+\delta_{f,i}-1} \mathrm{d}\alpha = \mathcal{I}_{f=i}$$

where $\alpha' = [\alpha_1, \alpha_2]$. The integration is performed over all α_i, $\alpha_i \geq 0$ for all $i = 1, 2$ and for all $\hat{f}_t$, $\alpha_1 + \alpha_2 = 1$.

Now, for the first fault ($i = 1$) and the fault estimate ($\hat{f}_t$), the integral can be written as follows

$$\mathcal{I}_{i=1} = \int_0^1 \alpha_1 \alpha_1^{n_1(t-1)-1} (1-\alpha_1)^{n_2(t-1)-1} \mathrm{d}\alpha_1 \propto \frac{n_1(t-1)}{n_1(t-1) + n_2(t-1)}.$$

Similarly for $i = 2$ we obtain

$$\mathcal{I}_{i=2} = \int_0^1 (1-\alpha_1) \alpha_1^{n_1(t-1)-1} (1-\alpha_1)^{n_2(t-1)-1} \mathrm{d}\alpha_1 \propto \frac{n_2(t-1)}{n_1(t-1) + n_2(t-1)}.$$

The resulting fault-configuration predictor (after normalization) has the form

$$p(f_t = i|d(t-1), \hat{f}_t) = \frac{n_i(t-1)}{n_1(t-1) + n_2(t-1)} \tag{46}$$

for $i = 1, 2$.

Appendix B: Observation model

The observation model (18) for the fault $f_t = i$ and the observed input u_t is

$$p(y_t|u_t, d(t-1), f_t = i) \propto \int_0^\infty \int_{-\infty}^\infty \prod_{f \in f^*} \mathrm{e}^{b'_f \left[C(t-1)+\delta_{f,i} c_t\right]} \mathrm{d}\mu \mathrm{d}\sigma^2 = \mathrm{e}^{b'_i \tilde{C}(t)} = \mathcal{I}.$$

Both the integrations $\int(\cdot)\,\mathrm{d}\mu$ and $\int(\cdot)\,\mathrm{d}\sigma^2$ are formally multi-dimensional as μ and σ^2 are two dimensional vectors $\mu = [\mu_1, \mu_2]$ and $\sigma^2 = [\sigma_1^2, \sigma_2^2]$. In practice, only integrals corresponding to the active fault i have to be performed because for $f \neq i$ it holds $\delta_{f,i} = 0$. The model pdf $\mathrm{e}^{b'_f \delta_{f,i} c_t} = \mathrm{e}^{b'_f 0 c_t} = 1$ and only prior pdf remains in the integrand. Such an integral equals to 1 according to the definition of the probability density function.

For the integration over μ_i we complete the exponent of the integrand to a square in the variable μ_i. We obtain

$$\begin{aligned} I &= \int_0^\infty \int_{-\infty}^\infty \frac{1}{\sigma_i^{\tilde{\kappa}_i(t)}} \mathrm{e}^{\frac{1}{2\sigma_i^2}(\tilde{\kappa}_i(t)\mu_i^2 - 2\tilde{W}(t)\mu + \tilde{V}(t))} \mathrm{d}\mu_i \mathrm{d}\sigma_i^2 = \\ &= \int_0^\infty \int_{-\infty}^\infty \frac{1}{\sigma_i^{\tilde{\kappa}_i(t)}} \mathrm{e}^{\frac{1}{2\sigma_i^2}\left[\tilde{\kappa}_i(t)(\mu_i - \tilde{m}_i(t))^2 + \tilde{R}_i(t)\right]} \mathrm{d}\mu_i \mathrm{d}\sigma_i^2 \end{aligned}$$

where we denoted

$$\tilde{m}_i(t) = \frac{\tilde{W}(t)}{\tilde{\kappa}_i(t)} \quad \text{and} \quad \tilde{R}_i(t) = \tilde{V}_i(t) - \tilde{m}_i(t)\tilde{W}_i(t).$$

Remind, that the tilde denotes the first update of the statistics for which $\delta_i \in \{0; 1\}$.

Now, the last integrand can be viewed as a product of two pdfs. The former is a normal distribution of μ_i, the latter does not depend on μ_i. So the integral over μ_i gives

$$I = \frac{1}{\sqrt{\tilde{\kappa}_i(t)}} \int_0^\infty \frac{1}{\sigma_i^{\tilde{\kappa}_i(t)}} \mathrm{e}^{-\frac{\tilde{R}_i(t)}{2\sigma_i^2}} \mathrm{d}\sigma_i^2.$$

Using the Gamma function we finally obtain

$$I = \frac{1}{\tilde{\kappa}_i(t)} \Gamma\left(\frac{\tilde{\kappa}_i(t) - 3}{2}\right) \left(\frac{2}{\tilde{R}_i(t)}\right)^{\frac{\tilde{\kappa}_i(t)-3}{2}},$$

the computed data predictor up to the normalization.

The probabilities (17) of faults for $i = 1,\ 2$ are

$$w_i \propto \frac{1}{\tilde{\kappa}_i(t)} \Gamma\left(\frac{\tilde{\kappa}_i(t) - 3}{2}\right) \left(\frac{2}{\tilde{R}_i(t)}\right)^{\frac{\tilde{\kappa}_i(t)-3}{2}} n_i(t-1) \tag{47}$$

and after normalization they are used for the approximate estimation of parameters.

Appendix C: Parameter estimation

Recall that all statistics are after the second update for which it is $\delta_i = w_i$.

The product of the models (data and fault-configuration) and prior pdfs for α and θ gives the joint pdf (the integrand of (14))

$$p(\alpha, \theta, f_t = i|d(t), \hat{f}_t) \propto \frac{1}{(\sigma_i^2)^{\frac{\kappa_i(t)}{2}}} \mathrm{e}^{-\frac{1}{2\sigma_i^2}\left[\kappa_i(t)(\mu_i - m_i(t))^2 + R_i(t)\right]} \alpha_1^{n_1(t)} \alpha_2^{n_2(t)} \tag{48}$$

By the optimal values $\hat{\theta}_i = [\hat{\mu}_i, \hat{\sigma}_i^2]$ and $\hat{\alpha} = [\hat{\alpha}_1, \hat{\alpha}_2]$ we mean any choice of θ_i and α at which the joint pdf attains its maximum.

To obtain equations for determining an estimates $\hat{\theta}_i$ and $\hat{\alpha}$ we consider the partial derivatives of the joint pdf (48) with respect to θ_i and parameters α_i (which are constrained to be nonnegative and to sum to one) and set their derivatives to zero. For $\hat{\theta}_i = [\hat{\mu}_i, \hat{\sigma}_i^2]$ one has the equations determined the unconstrained parameters $\theta_i, i = 1, 2$

$$\hat{\mu}_i - m_i(t) = 0, \quad \frac{R_i(t)}{2\hat{\sigma}_i^4} - \frac{\kappa_i(t)}{2\hat{\sigma}_i^2} = 0.$$

Therefore

$$\hat{\mu}_i = m_i(t) = \frac{W_i(t)}{\kappa_i(t)}, \quad \hat{\sigma}_i^2 = \frac{R_i(t)}{\kappa_i(t)} = \frac{W_i(t) - V_i(t)m_i(t)}{\kappa_i(t)}.$$

The equation determined $\hat{\alpha} = [\hat{\alpha}_1, \hat{\alpha}_2]$ is

$$n_1(t)(1 - \alpha_1) = n_2(t)\alpha_1$$

for all $\alpha = (\alpha_1, \alpha_2)$ such that and $\alpha_i \geq 0$, $i = 1, 2$, $\alpha_1 + \alpha_2 = 1$. Then the above equation holds for α satisfying the given constrains if and only if

$$\hat{\alpha}_1 = \frac{n_1(t)}{n_1(t) + n_2(t)} \quad \text{and} \quad \hat{\alpha}_2 = 1 - \hat{\alpha}_1.$$

Listing of MATLAB program

```
% Quasi-Bayes solution to FDI problem for static multi-model
%
%       Description of the most important variables of the program
%..t            indicates actual value of variable
%..o            denotes old values of variables ..
%..h            denotes hat (estimates of ..)

%model:
%y              system output
%ut             actual system input
%e              system noise
%ft             actual fault
%fht            estimate of the actual fault
%f              index denoting faults

%statistics:
%V,W,ka         data model statistics
%n              fault model statistics
%R              rest after the least-square minimization

%simulation:
%ns             length of simulation
%t              current time (t=1,2..ns)
%mu,sig         data model parameters (mean and standard deviation)
%al1,al2        probabilities of faulty states
%be1,be2        probabilities for input generator
%ga             probability of estimation of non-faulty state

%results:
%muh            regression parameter estimates
%s2h            noise covariance estimate
%al             estimates of faults probabilities
%wft            probabilities of all possible values of actual faulty state
%

clc, clear all, clear functions randn('seed',22);   rand('seed',17);
z14=zeros(1,4); z22=zeros(2,2); Go=zeros(12,1); wft=zeros(1,2); alh=z22;
Lw=wft; Lwo=wft; muh=z22; s2h=z22; nw=z22; nw1=z22; nf=z22; nu=z22;

% Simulation parameters
ns =800;              % length of simulation
```

```
lp =50;             % length for plotting
ga =.8;             % probability that estimate is 'good' (fht)
al1=.95; al2=.2;  % probabilities of classes (ft)
be1=.9;  be2=.3;    %              and control (ut)
mu =[ 7  10   8  12];               % matrices stored to
sig=[.1 1.1 .08  .6];               % vectors row-vise
phi=1;              % exponential forgetting

% Prior simulation
ni=20;              % length of prior simulation (for each component)
si=1;               % magnitude of prior noise standard deviation
%mi=mu;             % initial mean values
mi=[7.5 10.5 7.5 11.5];
ai=10;              % weight for prior alpha
for ut=1:2,
 for f=1:2,
  for t=1:ni,
   y=mi*Delta(ut,f,2,2,1)+sig*Delta(ut,f,2,2,1)*randn*si;
   Go=Go+Delta(ut,f,2,2,3)*[y^2; y; 1];
  end;
 end;
end; al=[al1 al2; 1-al1 1-al2]; no=[.9; 1.2; .8; 1.1]*ai; G=Go; n=no;

% Initial regression and noise parameters
for ut=1:2
 for f=1:2
  ss=G'*Delta(ut,f,2,2,3); gv=ss(1); gw=ss(2); gk=ss(3);
  muih(ut,f)=gw/gk;                     % regression parameters
  s2ih(ut,f)=(gv-gw*muh(ut,f))/gk;      % noise covariance
 end
end
      yy=zeros(1,ns); ee=zeros(1,ns); ff=zeros(ns,1); % vector definitions
      gg=zeros(1,ns); uu=zeros(1,ns); mm=zeros(8,ns); ww=zeros(ns,1);

% TIME LOOP (begins)
for t=1:ns, %t

% Generation of random variables with fixed probabilities
fht=ceil(rand-ga)+1; if fht==1, ft=ceil(rand-al1)+1; ut=ceil(rand-be1)+1;
else       ft=ceil(rand-al2)+1; ut=ceil(rand-be2)+1; end

% Simulation of system output
% --------------------------
 e=sig*Delta(ut,ft,2,2,1)*randn;    % noise e
```

```
 y=mu*Delta(ut,ft,2,2,1)+e;         % model y

      yy(t)=y; ee(t)=e; ff(t)=ft; gg(t)=fht; uu(t)=ut;  % storing for results
      nf(ft,fht)=nf(ft,fht)+1;  nu(ut,ft)=nu(ut,ft)+1;   % real frequencies

% Estimation of W
% ---------------
 sw=0;
 for f=1:2,
  utf=2*ut+f-2;

%  Computation of I(t-1)
  ss=G'*Delta(ut,f,2,2,3); gv=ss(1); gw=ss(2); gk=ss(3);
  th3=(gk-3)/2; lk=-.5*log(gk);  R=gv-gw^2/gk;
  Lwo(f) = lk+gammaln(th3)+th3*log(2/R);

%  First update of statistics
  G=Go+Delta(ut,f,2,2,3)*[y^2; y; 1];
  n=no+Delta(f,fht,2,2,1);

%  Computation of I(t)
  ss=G'*Delta(ut,f,2,2,3); gv=ss(1); gw=ss(2); gk=ss(3);
  th3=(gk-3)/2;  lk=-.5*log(gk);  R=gv-gw^2/gk;
  Lw(f) = lk+gammaln(th3)+th3*log(2/R);
 end
 Lwd=Lw-Lwo; Lwn=Lwd-max(Lwd);
 Elw=exp(Lwn).*n([fht fht+2])';
 wft=Elw/sum(Elw);

% Predicted probabilities of components W
% ---------------------------------------
 ww(t)=wft(2);

%  Counting of particular fault predictions for displaying results
 if t>0,   nw =nw +[2-ft 2-ft;ft-1 ft-1].*round([wft;wft]); end
 if t>(ns-lp), nw1=nw1+[2-ft 2-ft;ft-1 ft-1].*round([wft;wft]); end

% Estimation of parameters
% ------------------------
 for f=1:2,
  utf=2*ut+f-2;

%  Second update of statistics
  G=Go+wft(f)*Delta(ut,f,2,2,3)*[y^2; y; 1];
```

```
  n=no+wft(f)*Delta(f,fht,2,2,1);

%  Time shift:  old_stat -> new_stat
  Go=G; no=n;
 end

%  Estimates of parameters
% data model
for ut=1:2
 for f=1:2
  ss=G'*Delta(ut,f,2,2,3); gv=ss(1); gw=ss(2); gk=ss(3);
  muh(ut,f)=gw/gk;                        % regression parameters
  s2h(ut,f)=(gv-gw*muh(ut,f))/gk;         % noise covariance
  utf=2*ut+f-2;
  mm(utf,t)=muh(ut,f); mm(4+utf,t)=mu*Delta(ut,f,2,2,1);
 end
end
% fault model
for fht=1:2
 n12=n'*Delta(1,fht,2,2,1)+n'*Delta(2,fht,2,2,1);
 alh(1,fht)=n'*Delta(1,fht,2,2,1)/n12;
 alh(2,fht)=n'*Delta(2,fht,2,2,1)/n12;
 alh(2,fht)=1-alh(1,fht);
 alh(1,fht)=1-alh(2,fht);
end end
% TIME LOOP (ends)

 sigh=sqrt(s2h);                          % standard deviation
 nfn=[nf(1,1)+nf(2,1) nf(1,2)+nf(2,2);
     nf(1,1)+nf(2,1) nf(1,2)+nf(2,2)];
 alr=nf./nfn;
 pnw=[nw(1,2)/sum(nw(1,:))nw(2,1)/sum(nw(2,:))
     (nw(1,2)+nw(2,1))/sum(sum(nw))];
 pnw1=[nw1(1,2)/sum(nw1(1,:)) nw1(2,1)/sum(nw1(2,:))
      (nw1(1,2)+nw1(2,1))/sum(sum(nw1))];
 e2=sum(ee.*ee)/ns;

% Results
% -------
 muh,mu,sigh,sig,alh,alr,al,nw,pnw,nw1,pnw1
 figure(2), x=(min(yy)-1):.5:(max(yy)+1);
 hist(yy,x); print -dps figure2
 figure(1),
 plot(1:ns,yy,'o'), print -dps figure1
```

```
figure(3),
plot(mm'), axis([1 ns 6 13]), print -dps figure3
figure(4), o=0; s=(ns-lp-o+1):(ns-o); wf=[ww ff-1];
plot(s,wf(s,1),'xr',s,wf(s,2),'o',s,ones(size(s))/2),
axis([ns-lp-o+1 ns-o -.1 1.1]), print -dps figure4
```