Robust Estimation of Autoregressive Processes using a Mixture-Based Filter-Bank

Václav Šmídl [a], Anthony Quinn [a], Miroslav Kárný [b],

Tatiana V. Guy [b]

[a] *Department of Electronic & Electrical Engineering, Trinity College Dublin, Ireland.*

[b] *ÚTIA, Academy of Sciences of the Czech Republic, Prague, Czech Republic.*

Abstract

A mixture-based framework for robust estimation of ARX-type processes is presented. The ARX process is presumed to suffer from an unknown noise and/or distortion. The approach taken here is to model the overall degraded process via a mixture. Each component of this mixture uses the *same* ARX model but explores a different noise/distortion process. Estimation of this mixture unifies the preprocessing and process modelling tasks. The Quasi-Bayes (QB) procedure for mixture identification is extended to yield a fast recursive update of the estimator statistics. This allows non-stationary noise/distortion effects to be tracked. An application in on-line outlier-robust estimation of an AR process is given.

Key words: Bayesian estimation, probabilistic mixtures, recursive estimation, filter-bank, adaptive systems, outlier detection.

---

1 Introduction

Preprocessing of observations is a necessary step prior to estimation of model parameters. Various filters for removal of outliers, high frequency noise, and other artifacts may be sequentially applied to measured data [5,7]. Tuning of the filters is a demanding operation, often taking hours of experimentation by an experienced designer. It may be intractable if the processed data files are extensive. Various semi-automatic techniques for filter tuning have been developed for particular filter classes [4,10]. Despite this, preprocessing filters are often selected in a heuristic way and manually tuned for optimal performance by setting various thresholds and tuning knobs. This determines the success of the preprocessing, and, therefore, of subsequent process modelling, prediction and control design [14].

The method presented in this paper unifies preprocessing and estimation. Preprocessing is seen as an inevitable part of estimation of the model parameters. The idea behind the technique is the following: the unknown optimal filter is approximated by a rich bank of *a priori* selected filters. Each filter output yields the regression vector for one component of a mixture. The key property of the model—common regression parameters across all component models—implies an elegant recursive update algorithm. Estimator statistics for these common parameters are formed as a weighted sum of outer products on these filter-dependent regression vectors. A fast, adaptive algorithm is revealed.

In Section 2, Bayesian estimation of an ARX process under Gaussian assumptions is reviewed. In Section 2.2, it is extended to a mixture framework in order to cope with preprocessing uncertainties. A fast recursive estimation algorithm is derived. In Section 3, the framework is shown, in simulation, to yield robust ARX process estimates in the presence of outliers. Properties of the algorithm are discussed briefly in Section 4.

2 Bayesian ARX estimation

It is supposed that the measured data are generated by the extended AutoRegressive model with Exogenous Input (ARX) [13]. This model represents an extension of classical Multi-Input Multi-Output (MIMO) ARX models [11], by allowing *known* non-linear transformations of the output. It embraces incremental regression models, Autoregressive Moving Average (ARMA) models with known MA part, and a rich set of non-linear models.

The sequence of data, $d(t) = [d_1, d_2, \ldots, d_t]$, measured at times $t = 1, \ldots, T$, is formed by data items, $d_t = [y_t', u_t']'$ consisting of the m-dimensional system output, $y_t = [y_{1;t}, \ldots, y_{m;t}]'$, and (possibly empty) externally-controlled vector input, u_t. Here, $'$ denotes transposition. The transformed (filtered) data are modelled as follows (the ARX model):

$$\tilde{y}_t = \theta\tilde{\psi}_t + \Omega^{-\frac{1}{2}}e_t,\ \ f\left(e_t\right) = \mathcal{N}\left(0, I_m\right). \tag{1}$$

Here, the m-dimensional vector, $\tilde{y}_t$, is the filtered output. The ρ-dimensional vector, $\tilde{\psi}_t$, is the filtered regressor. θ is an $m \times \rho$-dimensional matrix of coefficients of the linear, time-invariant model. $e(t) = [e_1, \ldots, e_t]$ is a sequence of m-dimensional, temporally-independent, normally-distributed random variables with zero mean and identity covariance matrix, I_m. Ω is a positive-definite $m \times m$ precision matrix. The filtered quantities, $\tilde{y}_t$ and $\tilde{\psi}_t$, are obtained from the measured data, $d\left(t\right)$, via a *known* filter, $F(d(t))$:

$$\tilde{\Psi}_t = \begin{bmatrix} \tilde{y}_t \\ \tilde{\psi}_t \end{bmatrix} = \begin{bmatrix} F_y\left(d(t)\right) \\ F_\psi\left(u_t, d(t-1)\right) \end{bmatrix} = F\left(d(t)\right). \tag{2}$$

In (2), $F_y(\cdot) = [F_1(\cdot), \ldots, F_m(\cdot)]'$ and $F_\psi(\cdot) = [F_{m+1}(\cdot), \ldots, F_{m+\rho}(\cdot)]'$ are the parts of the transformation, F, generating the filtered output, $\tilde{y}_t$, and the filtered regressor, $\tilde{\psi}_t$, respectively. $\tilde{\Psi}_t$ is called the *extended regressor*.

We assume that the current realization of the external input, u_t, brings no information about the unknown parameters, $\Theta = \{\theta, \Omega\}$, beyond what is available from the data up to that time. This assumption is known as the *Natural Conditions of Control (NCC)* assumption [13], and can be expressed formally as

$$f\left(\Theta | d\left(t-1\right), u_t\right) = f\left(\Theta | d\left(t-1\right)\right). \tag{3}$$

Peterka [13] has shown that the posterior and predictive probability density functions (pdfs) of the extended model (1) have a form identical to the pdfs for the classical ARX model. This theory is reviewed in the next Section.

2.1 ARX estimation with known filter

For the case of a known filter (2), Bayesian estimation and prediction for the model (1) are governed by the following proposition.

Proposition 1 (Bayesian estimation and prediction of ARX model) *Let the data—preprocessed by (2)—be generated by the ARX model (1), and* *let the m-to-m transformation, F_y (2), have non-zero Jacobian:*

$$J_{F_y;t}\left(d(t)\right) = \begin{vmatrix} \frac{\partial F_1}{\partial y_{1;t}}, & \cdots, & \frac{\partial F_m}{\partial y_{1;t}} \\ \vdots & & \vdots \\ \frac{\partial F_1}{\partial y_{m;t}}, & \cdots, & \frac{\partial F_m}{\partial y_{m;t}} \end{vmatrix} \neq 0. \qquad (4)$$

Here, $|\cdot|$ denotes the matrix determinant. Let the prior pdf be of the conjugate Gauss-Wishart (GW) type [2]:

$$f(\Theta) = GW_\Theta(V_0, \nu_0) \equiv \frac{|\Omega|^{0.5(\nu_0)}}{\mathcal{I}(V_0, \nu_0)} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(V_0[-I_m, \theta]'\Omega[-I_m, \theta]\right)\right\}, \qquad (5)$$

where V_0 is an $(m+\rho)\times(m+\rho)$-dimensional positive-definite symmetric matrix, $\nu_0 > 0$, and

$$\mathcal{I}(V, \nu) = \pi^{\frac{1}{4}m(m-1)} \prod_{j=1}^{m} \Gamma\left[\frac{1}{2}(\nu - j + 1)\right] |\Lambda|^{-0.5\nu} |V_\psi|^{-0.5} 2^{0.5\nu} (2\pi)^{0.5\rho}, \qquad (6)$$

$$V = \begin{bmatrix} V_y & V_{\psi y}' \\ V_{\psi y} & V_\psi \end{bmatrix}, \ \Lambda = V_y - V_{\psi y}' V_\psi^{-1} V_{\psi y}.$$

Here, V_y is the $m \times m$-dimensional upper-left sub-matrix of V.

Then, under the NCC assumption, the posterior pdf of Θ, conditioned by $d(t)$ and the known filter, F, is also $GW_\Theta(V_t, \nu_t)$ (5) with extended information matrix, V_t, and number of degrees of freedom ν_t:

$$f(\Theta|d(t), F) = GW_\Theta(V_t, \nu_t). \qquad (7)$$

V_t and ν_t are updated at times $t = 1, 2, 3, \ldots$ according to the following recursions:

$$V_t = V_{t-1} + \tilde{\Psi}_t \tilde{\Psi}_t' = V_0 + \sum_{i=1}^{t} \tilde{\Psi}_i \tilde{\Psi}_i', \tag{8}$$

$$\nu_t = \nu_{t-1} + 1 = \nu_0 + t. \tag{9}$$

Hence the update of V_t *is with respect to an outer product of the extended regressor,* $\tilde{\Psi}_t$ *(2).*

The predictive pdf of y_t*, given* u_t *and* $d(t-1)$*, is of the form:*

$$f\left(y_t|\tilde{\psi}_t, V_{t-1}, \nu_{t-1}, F\right) \propto \left\|J_{F_y;t}\right\| \left(1 + \frac{\hat{e}_t' \Lambda_{t-1}^{-1} \hat{e}_t}{1+\zeta_t}\right)^{-\frac{\nu_{t-1}-\rho+2+m}{2}}, \tag{10}$$

where $\|\cdot\|$ *denotes the absolute value of (4), and*

$$\begin{aligned} \hat{e}_t &= \tilde{y}_t - \hat{\theta}_{t-1}' \tilde{\psi}_t, \\ \hat{\theta}_{t-1} &= V_{\psi;t-1}^{-1} V_{\psi y;t-1}, \\ \zeta_t &= \tilde{\psi}_t' V_{\psi;t-1}^{-1} \tilde{\psi}_t. \end{aligned} \tag{11}$$

In (11), $\hat{\theta}_{t-1} = E[\theta|d(t-1), F]$ *is the mean* a posteriori *estimate of* θ *at time* $t-1$*, obtained from (7). If* $J_{F_y;t}$ *is constant (i.e. if* $F_y(\cdot)$ *(2) is a linear function of* y_t*), then (10) is a* Student-t *distribution [2]. Then, also, the mean predicted value of* y_t *is of the standard AR kind:*

$$\overline{y}_t = E[y_t|u_t, d(t-1), F] = \left\|J_{F_y;t}\right\|^{-1} \hat{\theta}_{t-1} \tilde{\psi}_t. \tag{12}$$

PROOF. See [13].

The extended information matrix, V_t, and the number of degrees of freedom, ν_t, together form the sufficient statistics for estimation of Θ. Their update, (8) and (9), is equivalent to the well-known Recursive Least-Squares (RLS) algorithm [11]. The V_t update is often poorly conditioned, and so its LD decomposition [3,13] must be used in order to counteract the associated numerical instabilities [13].

2.2 Estimation with an unknown filter

If the filter F (2) is unknown, its uncertainty can be included in the modelling procedure as follows:

$$\begin{aligned} f(y_t, F|u_t, d(t-1), \Theta) &= f(y_t|u_t, d(t-1), \Theta, F) f(F|u_t, d(t-1), \Theta) \quad (13) \\ &= f(y_t|\tilde{\psi}_t, \Theta, F) f(F|d(t-1)). \quad (14) \end{aligned}$$

The first term in the final equality follows from the assumed influence of the filter F. The second term follows from two assumptions, namely (i) the NCC assumption (3), and (ii) conditional independence of F and Θ given $d(t-1)$. The distribution on observations $f\left(y_t|\tilde{\psi}_t, \Theta, F\right)$ may be recovered from the model (1) as follows:

$$f\left(y_t|\tilde{\psi}_t, \Theta, F\right) = \left\| J_{F_y;t} \right\| f\left(\tilde{y}_t|\tilde{\psi}_t, \Theta\right). \quad (15)$$

The observed process model is then obtained as the marginal pdf of (14) over all possible filters:

$$f\left(y_t|u_t, d(t-1), \Theta\right) = \int_{F^*} f\left(y_t|\tilde{\psi}_t, \Theta, F\right) f\left(F|d(t-1)\right) dF, \quad (16)$$

where F^* denotes the space of all possible filters. This solution is intractable, since F^* is generally extremely rich, and often uncountable. A reasonable approximation, yielding a tractable solution, is provided by the following proposition.

Proposition 2 (Mixture model for a degraded ARX process) *Let the space of all possible filters, F^*, be approximated by a finite filter-bank, $\boldsymbol{F}$, of $c < \infty$ representative filters:*

$$\boldsymbol{F} = \{F_i, i = 1, \ldots, c\}. \tag{17}$$

Then, the observed process model (16), conditioned by $\boldsymbol{F}$, has the following mixture form asymptotically:

$$f(y_t|u_t, d(t-1), \Theta, \alpha, \boldsymbol{F}) = \sum_{i=1}^{c} f\left(y_t|\tilde{\psi}_{i;t}, \Theta, F_i\right)\alpha_i, \tag{18}$$

where

$\tilde{\psi}_{i;t}$ *denotes the regression vector formed by data filtered by F_i, the ith known filter in the filter-bank (2);*

$f\left(y_t|\tilde{\psi}_{i;t}, \Theta, F_i\right)$ *is the parameterized component which assumes that the data, when filtered by F_i, i.e. $y_{i;t}$, is ARX (1) with F_i-filtered regression vector, $\tilde{\psi}_{i;t}$ (2);*

$\alpha = [\alpha_1, \ldots, \alpha_c]$ *is a vector of* time-invariant *component weights, such that $\alpha_i \in \alpha^* \equiv \{\alpha_i \geq 0,\ \sum_{i=1}^{c}\alpha_i = 1\}$.*

PROOF. Rewriting (16), using (13) and (14), for the assumed finite approximation, $F^* = \boldsymbol{F}$:

$$\begin{aligned} f(y_t|u_t, d(t-1), \Theta, \alpha, \boldsymbol{F}) &= \sum_{i=1}^{c} f(y_t|u_t, d(t-1), \Theta, F_i) \Pr[F_i|d(t-1)] \\ &= \sum_{i=1}^{c} f\left(y_t|\tilde{\psi}_{i;t}, \Theta, F_i\right)\alpha_{i;t}, \end{aligned}$$

where $\alpha_{i;t} = \Pr[F_i|d(t-1)]$ is a t-dependent probability mass function; i.e. $\sum_{i=1}^{c}\alpha_{i;t} = 1$, $\alpha_{i;t} \geq 0, \forall t$. For a fixed F_i, $\alpha_{i;t}$ is a bounded martingale with respect to the σ-algebra generated by the data $d\,(t-1)$, so that $\alpha_{i;t}$ converges almost surely to a constant probability α_i, as $t \to \infty$. See, for example, [12, page 64, item I.A]. □

Hence, the task of parameter estimation with unknown filter, F, can be replaced, approximately, by the task of parameter estimation for the mixture model (18). The model parameters consist of the common ARX parameters, Θ, and t-independent weights, α.

General guidelines for choosing filter candidates is outside the scope of this communication. However, for particular noise models, selection of a suitable filter-bank may follow naturally from analysis of the problem. This is true of the outlier removal problem presented in Section 3. A filter selection approach for the ARMAX model can be found in [8].

The mixture model (18) yields a tractable estimation task. Methods for estimation of the parameters of mixture models are typically based on the Expectation Maximization (EM) algorithm [6,15]. However, a special feature of the model (18)—namely the common parameterization of *each* component model via the *same* set Θ—can be exploited to yield a novel and efficient recursive estimation technique. Success depends on preserving the advantageous recursive properties of ARX estimation outlined in Proposition 1. This is achieved via the Quasi-Bayes approximation [9], as presented in the next proposition.

The following assumptions are made:

(1) The fixed filter-bank, $\boldsymbol{F} = \{F_i, i = 1, \ldots, c\}$, is known *a priori*. The pdf of the observed data, y_t, for a known (active) transformation $F_j, j \in 1, \ldots, c$,

at time t, is $f(y_t|\tilde{\psi}_{j;t}, \Theta, F_j) = \left\| J_{F_{j,y};t} \right\| f\left(\tilde{y}_{j;t}|\tilde{\psi}_{j;t}, \Theta\right)$, as described by model (1). Here, $\tilde{y}_{j;t}$ denotes the output, filtered by F_j.

(2) The posterior pdf of the mixture parameters, Θ, α, at time $t-1$, has the form

$$f\left(\Theta, \alpha | d(t-1), \boldsymbol{F}\right) = GW_{\Theta}(V_{t-1}, \nu_{t-1}) Di_{\alpha}(\kappa_{t-1}), \tag{19}$$

$t = 1, 2, 3, \ldots$ Here, $d(0)$ is empty, $GW_{\Theta}(\cdot)$ is given by (5), and V_0, ν_0, κ_0 are assigned *a priori*. The assumed Dirichlet pdf on α is defined as

$$Di_{\alpha}(\kappa) \equiv \prod_{i=1}^{c} \frac{\alpha_i^{\kappa_i - 1}}{\Gamma(\kappa_i)} \Gamma\left(\sum_{i=1}^{c} \kappa_i\right),$$

where $\kappa = [\kappa_1, \kappa_2, \ldots, \kappa_c]'$, $\kappa_i \geq 0$, and $\Gamma(\cdot)$ is the Gamma function [1].

Proposition 3 (Quasi-Bayes (QB) estimation of mixture model (18)) *Consider a new, unobserved, random variable, $i_t \in \{1, \ldots, c\}$, which points to the component active at time t, and is governed by the time-invariant probabilities, $\mathsf{Pr}[i_t = i | d(t-1), \alpha] = \alpha_i$ (Proposition 2). (14) may be re-written in terms of i_t (using the same assumptions as those governing (14)), as follows:*

$$f(y_t, i_t | u_t, d(t-1), \Theta, \alpha, \boldsymbol{F}) = f(y_t | \tilde{\psi}_{i_t;t}, \Theta, F_{i_t}) \alpha_{i_t}. \tag{20}$$

Under the NCC assumption (3), and assumptions (1)–(2) above, the following statements are true:

(1) The marginal pdf $f(y_t | u_t, d(t-1), \Theta, \alpha, \boldsymbol{F})$ of the introduced pdf (20) is the mixture model (18).

(2) The joint posterior pdf of Θ and α, augmented by the pointer i_t, can be written in the form:

$$f(\Theta, \alpha, i_t|d(t), \boldsymbol{F}) \propto GW_\Theta \left(V_{t-1} + \sum_{i=1}^{c} \delta_{i,i_t} \tilde{\Psi}_{i;t} \tilde{\Psi}'_{i;t},\ \nu_{t-1} + 1 \right) \times \tag{21}$$

$$Di_\alpha \left(\kappa_{t-1} + \sum_{i=1}^{c} \delta_{i,i_t} [\underbrace{0, \ldots, 0}_{i-1}, 1, 0, \ldots]' \right),$$

where $\delta_{i,i_t} = \left\{ \begin{smallmatrix} 1 \text{ if } i = i_t \\ 0 \text{ if } i \neq i_t \end{smallmatrix} \right.$ *, is the Kronecker function, and* $\tilde{\Psi}_{i;t} = \left[\tilde{y}'_{i;t}, \tilde{\psi}'_{i;t} \right]'$*,* $i = 1, \ldots, c$*, is the filtered extended regressor (2) generated by the ith filter,* F_i*, of the filter-bank,* $\boldsymbol{F}$*.*

(3) The unknown value of the Kronecker function, δ_{i,i_t}*, has posterior expectation*

$$w_{i;t} \equiv E\left[\delta_{i,i_t}|d(t), \boldsymbol{F}\right] \propto \mathcal{I}\left(V_{t-1} + \tilde{\Psi}_{i;t}\tilde{\Psi}'_{i;t},\ \nu_{t-1} + 1\right)(\kappa_{i;t-1} + 1). \tag{22}$$

Since $\sum_{i=1}^{c} w_{i;t} = 1$*, the constant of proportionality in (22) is available.*

(4) Adopting the certainty equivalent approximation,

$$\delta_{i,i_t} \approx E[\delta_{i,i_t}|d(t)] = w_{i;t}, \quad i = 1, \ldots, c, \tag{23}$$

then (22) yields an approximate marginal for (21), which preserves the GW form:

$$f(\Theta, \alpha|d(t), \boldsymbol{F}) = GW_\Theta(V_t, \nu_t) Di_\alpha(\kappa_t). \tag{24}$$

The statistics in (24) are updated according to the following recursions:

$$V_t = V_{t-1} + \sum_{i=1}^{c} w_{i;t} \tilde{\Psi}_{i;t} \tilde{\Psi}'_{i;t}, \tag{25}$$

$$\nu_t = \nu_{t-1} + 1, \tag{26}$$

$$\kappa_{i;t} = \kappa_{i;t-1} + w_{i;t}. \tag{27}$$

This constitutes Quasi-Bayes (QB) estimation of the mixture model (18). (23) is the QB approximation*, which preserves the GW-Di distributional form (19) during the update (24).*

PROOF. The first statement is directly implied by marginalization of the model (20) over i_t. The second statement (i.e. exact updating via random variables, δ_{i,i_t}) uses the extended ARX version of the model (1), the assumed *GW-Di* form of the parameter distribution (19), and Bayes' rule, applied under the NCC assumption (3). The third statement requires marginalization of $GW_\Theta(\cdot)$ in (21) over Θ and $Di_\alpha(\cdot)$ over α. The result (22) invokes the predictive Student pdf form (10) of the former and the elementary property of the Dirichlet pdf, namely $E[\alpha_i|\kappa] \propto \kappa_i$, for the latter. The final recursion (25) is implied directly by substituting the adopted QB approximation (23) into (21). □

Note, from (22), that $w_{i;t} = \mathsf{Pr}[i_t = i|d(t), \boldsymbol{F}]$, which is the active filter probability distribution at time t. The time-variant nature of this measure will be important in the implied data preprocessing algorithm (Section 3).

2.3 Output prediction

The predictive model can be obtained by marginalization, substituting (18) and (19), and using the NCC assumption (3):

$$
\begin{aligned}
f(y_t|u_t, d(t-1), \boldsymbol{F}) &= \textstyle\int_{\Theta^*}\int_{\alpha^*} f(y_t, \Theta, \alpha|u_t, d(t-1), \boldsymbol{F})\, d\Theta d\alpha \\
&= \int_{\Theta^*}\int_{\alpha^*} f(y_t|\Theta, \alpha, u_t, d(t-1), \boldsymbol{F})\, f(\Theta, \alpha|u_t, d(t-1), \boldsymbol{F})\, d\Theta d\alpha \\
&= \int_{\Theta^*}\int_{\alpha^*} \sum_{i=1}^{c} f\left(y_t|\tilde{\psi}_{i;t}, \Theta, F_i\right) \alpha_i GW_\Theta(V_{t-1}, \nu_{t-1}) Di_\alpha(\kappa_{t-1})\, d\Theta d\alpha \\
&= \sum_{i=1}^{c} \int_{\alpha^*} \alpha_i Di_\alpha(\kappa_{t-1})\, d\alpha \int_{\Theta^*} f\left(y_t|\tilde{\psi}_{i;t}, \Theta, F_i\right) GW_\Theta(V_{t-1}, \nu_{t-1}) d\Theta \qquad (28) \\
&\propto \sum_{i=1}^{c} \hat{\alpha}_{i;t-1} f\left(y_t|\tilde{\psi}_{i;t}, V_{t-1}, \nu_{t-1}, F_i\right), \qquad (29)
\end{aligned}
$$

where integration over the space, Θ^*, of Θ in (28) yields predictor (10) for the ith filter, F_i, and integration over the space, α^*, of α yields the mean value of the Dirichlet distribution, namely

$$\hat{\alpha}_{i;t-1} = \kappa_{i;t-1} / \left(\sum_{i=1}^{c} \kappa_{i;t-1} \right). \tag{30}$$

The overall predictive pdf is therefore a mixture of filter-dependent predictors (10), weighted by the expected filter weights, $\hat{\alpha}_{i;t-1}$.

In many applications, only the moments of (29) are required. Linearity of expectation implies that the overall *non-central moments* are obtained as the convex combination of the non-central moments of the individual predictors (12), weighted by $\hat{\alpha}_{i;t-1}$.

3 Example: Outlier-Robust AR Estimation

In this simulation, scalar data were realized from a second-order AR process ($m = 1$, $\rho = 2$, $\Theta = [\theta_1, \theta_2, \omega]'$), degraded by a random, high-variance outlier on every 30th sample. In the absence of a filter, the extended information matrix, V_t (8)—forming, with ν_t (9), sufficient statistics for estimation of Θ—is updated by the outer product of the extended regressor $\Psi_t = [y_t, y_{t-1}, y_{t-2}]'$. Therefore, an outlier degrades a parameter estimate if it appears in at least one entry of this regressor. An ideal preprocessing filter, F, should have no influence on the standard ARX update when there is no outlier, but in other cases it should replace the affected entry (assumed isolated for simplicity) with an appropriate reconstructed value, $\overline{y}_\cdot$. This motivates the design of a filter-bank, $\boldsymbol{F}$, with the following members:

$F_1:$	$\tilde{y}_{1;t} = y_t$	$\tilde{\psi}_{1;t} = [y_{t-1}, y_{t-2}]'$	standard AR model
$F_2:$	$\tilde{y}_{2;t} = y_t/h$	$\tilde{\psi}_{2;t} = [y_{t-1}, y_{t-2}]'/h$	AR model with higher noise variance, $h^2\omega^{-1}$
$F_3:$	$\tilde{y}_{3;t} = y_t$	$\tilde{\psi}_{3;t} = \left[\overline{y}_{t-1}, y_{t-2}\right]'$	replacing 1-step-delayed output by expected value
$F_4:$	$\tilde{y}_{4;t} = y_t$	$\tilde{\psi}_{4;t} = \left[y_{t-1}, \overline{y}_{t-2}\right]'$	replacing 2-step-delayed output by expected value

A key advantage of this choice is that $\boldsymbol{F}$ is parameterized with respect to just one single scalar, $h > 1$, which appears in F_2. Selection of this parameter may be interpreted as setting a tuning knob for $\boldsymbol{F}$, but simulations indicate that estimates are quite robust with respect to the choice of h. However, if the presence of a tuning knob is undesirable, F_2 can be replaced by several filters, each parameterized using a candidate value of h.

The filters F_3 and F_4 replace a hypothesized outlier in the regressor by a reconstructed value, $\overline{y}_{\cdot}$. An optimal process reconstruction for each filter in the filter-bank is the mean value of the predictor (12), $\overline{y}_{i;t}$, associated with this filter. The filtered value, conditioned by knowledge of the active component, i_t, at time t is therefore

$$\overline{y}_{i_t;t} = \sum_{i=1}^{c} \delta_{i,i_t} \overline{y}_{i;t}. \tag{31}$$

In order to remove the dependence on the unknown pointer random variable, i_t, the QB approximation (23) is applied once again, replacing δ_{i,i_t} by its

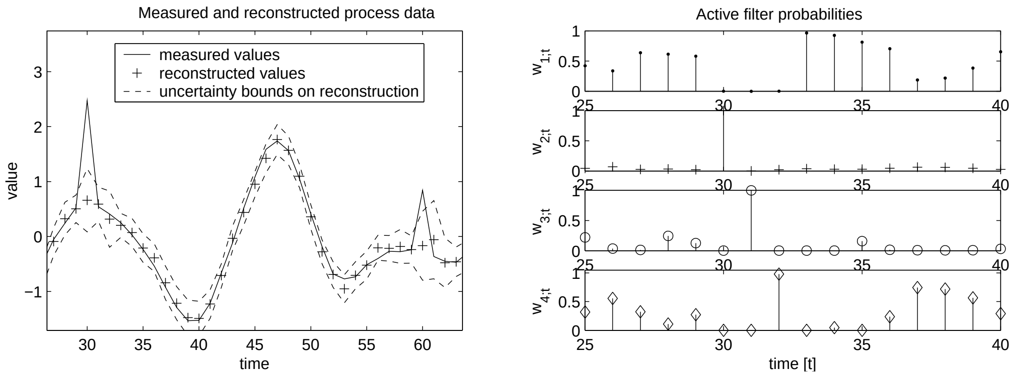

Figure 1. **Left:** outlier-degraded 2nd-order AR process data. Note the high-variance outlier at $t = 30$. Reconstruction and uncertainty bounds are also given. **Right:** active filter probabilities, $w_{i;t}$, around $t = 30$.

expected value, $w_{i;t}$ (22), for each $i = 1, \ldots, c$. Therefore, substituting (12) and (23) into (31), the reconstructed value of the process output for the given filter-bank is

$$\overline{y}_t = \hat{\theta}_{t-1} \left(w_{1;t}\tilde{\psi}_{1;t} + h w_{2;t}\tilde{\psi}_{2;t} + w_{3;t}\tilde{\psi}_{3;t} + w_{4;t}\tilde{\psi}_{4;t} \right), \tag{32}$$

using parameter estimates (11). The measured and reconstructed outputs of the process are compared in Figure 1 (left).

Note that the required Jacobian (12) for each of the filters, F_i, $i = 1, \cdots, 4$, in the filter-bank above is, respectively, $J_{F_{i,y};t} = 1$, $i = 1, 3, 4$, and $J_{F_{2,y};t} = h^{-1}$. In fact, the AR-type predictor (12) is valid in this instance only because these Jacobians are constant, thereby preserving the Student-t form of the predictive pdf (10) in each case (Section 2.1). Nevertheless, in more general cases of non-linear filters, F_i, the proposed QB reconstruction would still be formed via the same $w_{i;t}$-weighted linear combination of appropriate predictors as that in (32).

Figure 2. Degraded AR model estimation, with and without filter-bank.

The main objective of this work was to achieve robust estimation of the model parameters, Θ. A comparison of poles of the estimated process — with and without the filter-bank preprocessing — is given in Figure 2, indicating the success of the approach. Note, finally, that this robust estimation is achieved on-line, i.e. in one sweep through the data.

4 Discussion

The introduced algorithm, (24)-(27), for estimation of the probabilistic mixture model (18) is based on the Quasi-Bayes (QB) approximation (23). The algorithm is a Bayesian alternative to the standard EM algorithm [6]. Specifically, in the introduced QB algorithm, estimates of active component probabilities, w_t (22), are evaluated using the full distribution (5) of the component parameters, Θ. In contrast, point estimates, $\hat{\Theta}$, of the component parameters are used in the EM algorithm. This ability to account for the uncertainty in Θ is an important advantage of the QB approach over the EM approach to mixture estimation.

Accuracy and convergence issues for the QB approximation (23) were studied in [15]. It was shown, for example, that estimation of mixture weights is asymptotically efficient if the supports of the components, $f(y_t|F_i)$, do not overlap. This condition is not satisfied for the filter-bank mixture (18). However, it is approximately true when the chosen filters in the filter-bank $\boldsymbol{F}$ (17) are sufficiently distinct. Our experiments support this statement.

Computational complexity of the algorithm for estimation of the degraded ARX model, (25)-(27), can be compared to the computational complexity for estimation of the standard ARX model (Proposition 1). Each step of the QB algorithm requires $2c$ updates of the statistics V_t (i.e. c in (22) and c in (25)) rather than the single update required in the standard algorithm (8). Furthermore, c normalizations of the kind in (6) must be evaluated in the QB algorithm, one for each $w_{i;t}$, $i = 1, \ldots, c$ (22). If the algorithm is implemented using LD decompositions [3, 13], then the V_t-update is the most expensive operation. Hence, the QB algorithm for estimation of the mixture model (18) of a degraded ARX process, using c filters, is approximately $2c$-times more computationally expensive than standard estimation of an ARX model of the same order.

5 Conclusion

A mixture framework for robust recursive estimation of ARX processes has been presented in the paper. It has been shown that an ARX process degraded by an unknown noise/distortion can be approximated by a mixture of processes with identical ARX parameterization but different known candidate preprocessing filters. The unknown weights of the filters in the filter-bank are estimated (30) using model predictions, (22), (27), and so the task of preprocessing is unified with the task of estimation, within a consistent model. Expertise may be required in the choice of filter-bank members. However, this effort replaces the considerable experience and skill typically demanded for the tuning and supervision of *ad hoc* preprocessing filters.

An adaptive algorithm for parameter estimation has been derived. This confers two principal benefits: (i) non-stationary noise/distortion effects can be tracked via step-wise re-estimation of the filter weights; and (ii) the preprocessing and recursive estimation tasks are accomplished together in *one* sweep of the data. This is of potentially great importance in the adaptive control of processes whose observations are noisy and/or distorted.

The outlier-robust estimation results presented in this paper are promising, but a broader class of noise and distortion phenomena is amenable to this framework.

Acknowledgements

This research was supported by the following grants: EU-IST-12058, AV ČR S1075351, GAČR 102/03/0049, and GAČR 102/03/P010.

References

[1] M. Abramowitz and I.A. Stegun. *Handbook of mathematical functions.* Dover Publications, Inc., New York, 1972.

[2] J.M. Bernardo and A.F.M. Smith. *Bayesian theory.* John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore, 1997. 2nd edition.

[3] G.J. Bierman. *Factorization Methods for Discrete Sequential Estimation.* Academic Press, New York, 1977.

[4] S. Cao and R.R. Rhinehart. A self-tuning filter. *Journal of Process Control*, 7:139–148, 1997.

[5] L. H. Chiang, R. J. Pell, and M. B. Seasholtz. Exploring process data with the use of robust outlier detection algorithms. *Journal of Process Control*, 13:437–449, 2003.

[6] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society, Series B*, 39:1–38, 1977.

[7] F. Doymaz, J. Chen, J. A. Romagnoli, and A. Palazoglu. A robust strategy for real-time process monitoring. *Journal of Process Control*, 11:343–358, 2001.

[8] L. He and M. Kárný. Estimation and prediction with ARMMAX model: a mixture of ARMAX models with common ARX part. *International Journal of Adaptive Control and Signal Processing*, 17(4):265–283, 2003.

[9] M. Kárný, J Böhm, T. V. Guy, and P. Nedoma. Mixture-based adaptive probabilistic control. *International Journal of Adaptive Control and Signal Processing*, 17(2):119–132, 2003.

[10] V. Katkovnik, K. Egiazarian, and J. Astola. Application of the ICI principle to window size adaptive median filtering. *Signal Processing*, 83:251–257, 2003.

[11] L.Ljung. *System Identification-Theory for the User.* D. van Nostrand Company Inc., Prentice-hall. Englewood Cliffs, N.J, 1987.

[12] M Loeve. *Probability Theory II, 4th edition.* Springer Verlag, 1977.

[13] V. Peterka. Bayesian approach to system identification. In P. Eykhoff, editor, *Trends and Progress in System identification*, pages 239–304. Pergamon Press, Oxford, 1981.

[14] A. Quinn, P. Ettler, L. Jirsa, I. Nagy, and P. Nedoma. Probabilistic advisory systems for data-intensive applications. *International Journal of Adaptive Control and Signal Processing*, 17(2):133–148, 2003.

[15] 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.