Robust identification of switched regression models

Elom Ayih Domlan, Biao Huang, José Ragot, Didier Maquin

Robust identification of switched regression models

E. Domlan[1] B. Huang[1] J. Ragot[2] D. Maquin[2]

[1]*Department of Chemical and Materials Engineering, University of Alberta, Edmonton AB T6G 2G6, Canada*
[2]*Centre de Recherche en Automatique de Nancy, UMR 7039, Nancy-Université, CNRS 2, Avenue de la Forêt de Haye, 54516 Vandœuvre-lès-Nancy, France*
E-mail: Biao.Huang@ualberta.ca

Abstract: This study addresses the problem of parameters estimation for switched regression models used to represent systems with multiple operating modes or regimes. For the identification of such models, the collected data are from different operating modes and there is no a priori information holding on the partitioning of the data in regard to the different operating modes. The essential contributions of this study lie first in the estimation procedure of the model parameters that provides an analytical solution, second in the simultaneous resolution of the problem of estimating the model parameters and allocating the data points to the different local models and finally the robustness of the estimation procedure regarding the presence of outliers in the identification dataset.

1 Introduction

The interest of researchers in modelling systems presenting multiple operating regimes with classical linear or multi-linear models can be traced back to the early 1970s in various research areas such as economics, finance or engineering. This early attention to multiple operating regime systems has given rise to many modelling paradigms known as switching regression models or more generally as switched or switching models. Among the pioneering work, those of Hudson [1], Quandt [2] and Goldfeld [3] gave the first principles of such representation. Since then, many subsequent contributions have helped define a clear and precise framework for this formalism that can potentially be applied to any system or process that exhibits, in a natural or forced manner, a switched operating regime behaviour [4, 5]. Switched regression models can be generally classified into two categories:

- Models whose operating regime changes are determined by a known variable, meaning the values of that variable can be obtained directly through the instrumentation system or indirectly through the usage of observers.

- Models whose operating regime changes are determined by an unknown variable. In this later situation, the mechanism that generates the changes of operating regime has to be also modelled using the available measured variables.

It may be noticed that hybrid models cover both situations where the operating regime changes can moreover be described by a Markov process [6, 7] that can potentially bring additional knowledge for the characterisation of these systems.

Before describing the behaviour of a process or a system with a switched regression model, the preliminary question to be addressed is the actual existence of different operating regimes or modes. The most difficult situation is encountered when only input/output measurements exist without further information on the physical existence of many operating regimes. It is then mandatory to discover these operating regimes based on the available measurements. Once their existence is confirmed, they have to be characterised by assessing their number and the parameters of their corresponding local model.

Switched regression models can be viewed as a particular class of hybrid models that regroups different models such as switched auto regressive exogenous (ARX) models [8] or

piece-wise affine (PWA) models [9] and their extension to piece-wise auto regressive exogenous (PWARX) models [10]. From a general point of view, switched regression models are constructed by partitioning the operating space of the system into a finite number of regions and allocating a local autoregressive exogenous model to each obtained region. A switching mechanism is then defined in order to determine the conditions of occurrence of a switch from one local model to another. This can be equivalently viewed as imposing logical constraints on the model [11]. The resulting identification problem for this type of models is generally non-linear [12] because of the logical constraints that generate the switches from one local model to another, translating the physical viewpoint of changes in the process operating regime. The main difficulty results from the coupling between the estimation of the parameters of the local models and the estimation of the validity domains of these models with the arising question being: how is the identification dataset, the data relevant to a particular operating mode selected? It is therefore appropriate to resort to different concepts: data clustering, data classification and parameter estimation.

The conventional approaches typically use a two-stage procedure [13]. In the first stage, the dataset is a priori partitioned in several regions and; in the second stage, one proceeds to the estimation of the parameters of the local models in the resulting partitioned space [14]. The initial partitioning is then updated based on the results of the estimated parameters until the convergence of the successive obtained solutions.

The attention for the estimation problem of switched regressions models has been recently revived by several publications in the literature. In [10], a combination of clustering, classification and linear identification methods is used to provide a solution to the identification problem. Ragot *et al.* [15] introduced a method based on the selection of an adapted weighting function that is used to perform the allocation of the data to the operating modes. A Bayesian inference approach is used in [16] to solve the identification problem. A procedure based on the selection of a bound on the identification error is presented in [17]. The selection of an identification error bound allows us to simultaneously resolve the issue of data allocation and parameters estimation. Vidal *et al.* [8] proposed an algebraic geometric solution to the identification problem and Nakada *et al.* [18] developed an algorithm based on a statistical clustering technique.

Generally, during the identification of a process or a system, the data used for the identification procedure usually come from an instrumentation system that is subject to various perturbations, and even failures. The collected data are then often corrupted by measurement errors. If the measurement errors are of relatively small magnitude (regarding the magnitude of the measured variable) with zero-mean values and can be represented by Gaussian random variables, estimation methods based on the least square estimation technique can handily minimise the influence of these measurement errors. By contrast, the presence of outliers or measurement errors of relatively large magnitude is far more delicate to handle and it is well known that conventional least squared methods show little robustness in regard to this type of error.

Based on the work in [8, 19] presenting a general formulation of the problem of principal component analysis and its application for the identification of switched regression models, a robust identification method for switched regression is proposed for the quite realistic and practical situation where the data used for the estimation of the model parameters are corrupted by possible outliers. The model parameters are first extracted from the so-called hybrid decoupling constraint (HDC) and then the usage of an estimation criterion that simultaneously takes into account both outliers and measurement errors of small magnitude provides a robust estimation method for the model parameters, that is largely insensitive to the presence of outliers.

The problem under investigation is presented in Section 2 through an introductory example. The formalisation of the estimation procedure is given in Section 3. Section 4 deals with the problem of robustness of the estimation in regard to the presence of outliers. An illustrative example is presented in Section 5 to demonstrate the effectiveness of the procedure.

2 Introductory example

Consider the multiple operating regime system, represented by (1), which has two operating modes that are selected according to the switching mechanism described by (2)

$$y_{k+1} = a(\theta_k)u_k \tag{1}$$

$$a(\theta_k) = \begin{cases} a_1 & \text{if } \theta_k - 0.5 \geq 0 \\ a_2 & \text{if } \theta_k - 0.5 < 0 \end{cases} \tag{2}$$

with $a_1 \in \mathbb{R}$ and $a_2 \in \mathbb{R}$. The variable $\theta_{(\cdot)}$ can be regarded as the realisation of a random variable or as a variable parameter under the influence of unknown exogenous variables or even as a parameter that depends on the current operating conditions of the system.

Fig. 1 shows the time evolution of the system input and output for $a_1 = 1.05$ and $a_2 = -0.5$. In this figure, the output data resulting from the resolution of (1) has been corrupted by a multiplicative noise represented as a zero-mean Gaussian distribution with a standard deviation of 0.1. In addition to the measurement noise, three outliers have been added to the data at the time instants 9, 28 and 40. The outliers have been simulated by adding to the corresponding output an offset with a magnitude of 50% of the true value of the measured output.

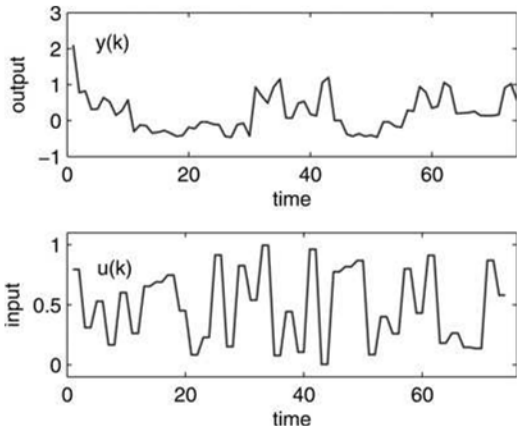

Figure 1 *Output and input of system (1)*

Conversely, knowing the output y_k of the system on a finite time horizon $[1 \ldots N]$, an identification problem can be formulated in the sense of identifying the values a_1 and a_2 taken by the parameter a and determining the time instants at which the operating mode of the system changes, that is, the switching instants. Eventually, if the parameter $\theta_{(\cdot)}$ depends on some internal variables (input, output or state) of the system, it may be interesting to characterise this dependence in the form of a model [20, 21].

The solution to this estimation problem is simple if the switching instants of the parameter a are known and if the ongoing operating mode of the system is known at every time instant. In this case, the estimation of the parameter could be simply done by using the data points belonging to the first operating mode for estimating a_1 and those belonging to the second operating mode for estimating a_2. In what follows, the estimation problem is considered for the situation in which these switching instants are unknown. Thus, the portions of the state trajectory described with the parameter a_1 and those described with parameter a_2 are a priori unknown.

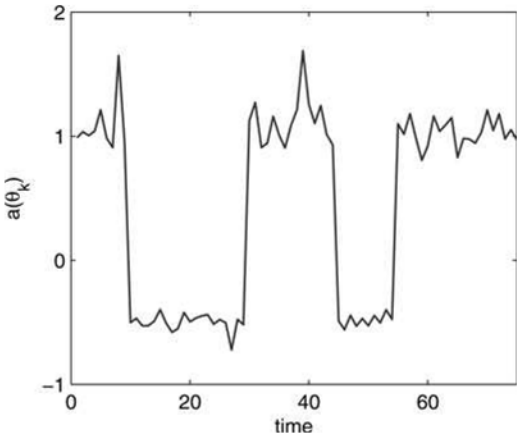

Figure 2 *$a(\theta_k)$ as a function of time*

To estimate the parameter a of the model, it is noticed that, for $u_k \neq 0$, (1) can be written as

$$\frac{y_{k+1}}{u_k} = a(\theta_k) \tag{3}$$

Fig. 2 shows the temporal evolution of the ratio expressed by (3). The changes of operating modes clearly appear and are expressed by the piecewise constant shape of the plot. The distribution (in the form of a histogram bar plot) of the values of $a(\theta_{(\cdot)})$ is presented in Fig. 3 and it clearly reveals the existence of two distinct values for $a(\theta_{(\cdot)})$.

The existence of several operating regimes is also confirmed in Fig. 4 drawn in the Plane $\{u_k; y_{k+1}\}$. The time link with the changes of operating modes no longer appears, the time being a hidden parameter in this representation. By contrast, the presence of two distinct operating modes is clearly highlighted. These two operating regimes must then be characterised.

Thus, in view of these different data representations, one is able to specify the number of operating regimes characterising the system. It remains to identify the models

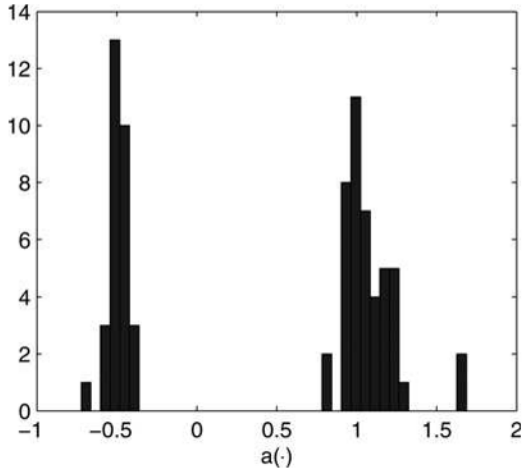

Figure 3 *Histogram of the values of a(·)*

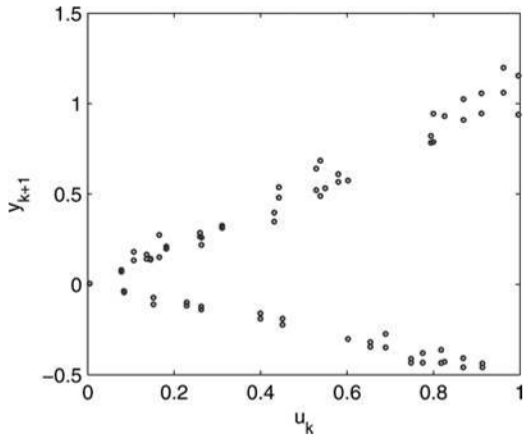

Figure 4 *Trajectory in the plane $\{u_k; y_{k+1}\}$*

related to these two operating modes. In general, this double identification problem is made difficult by the presence of measurement errors and in particular those qualified as accidental of large magnitude.

The search for the values of $a(\theta_{(\cdot)})$ may be done using either previously presented representations. On a more analytical viewpoint, the retaining principle rests on the maximisation of a cost function simultaneously valid for both operating modes. A candidate cost function could be [22]

$$\ell(a(\theta_k)) = \sum_{k=1}^{N-1}\left(\frac{1}{\eta + (y_{k+1} - a(\theta_k)u_k)^2}\right) \tag{4}$$

where η is a small real constant. The cost function in (4) allows an automatic partitioning of the data points. Indeed, given the form of the cost function (4), when $a(\theta_k) \neq a_1$ or $a(\theta_k) \neq a_2$, the value of $\ell(\cdot)$ is small as the term $(y_{k+1} - a(\theta_k)u_k)^2$ is significantly different from zero and larger than η. When $a(\theta_k) = a_1$ or $a(\theta_k) = a_2$, the term $(y_{k+1} - a(\theta_k)u_k)^2$ is equal to zero whenever the data point $(y_{k+1}; u_k)$ belongs to the operating regime described by $a(\theta_k)$. In that case, the value of $\ell(\cdot)$ is relatively large as it is constituted of several terms which values equal $1/\eta$. Therefore the maxima of the cost function (4) correspond to the case where $a(\theta_k) = a_1$ or $a(\theta_k) = a_2$. Note that the effectiveness of the cost function (4) is of course conditioned to how representative is the identification dataset regarding the different operating regimes of the system. This is a prerequisite for any identification method for multiple operating regimes systems. Also, the usage of the cost function (4) is limited by the number of operating modes and the number of identification data points. Indeed, with the increase of the number of operating modes, a larger and 'richer' identification dataset would be required. Fig. 5 shows the form of the cost function in (4) for the data presented in Fig. 1. The function has been plotted with $\eta = 0.001$ and for $a(\theta_{(\cdot)}) \in [-4, 4]$. The function exhibits two local maxima in the vicinity of the values a_1 and a_2.

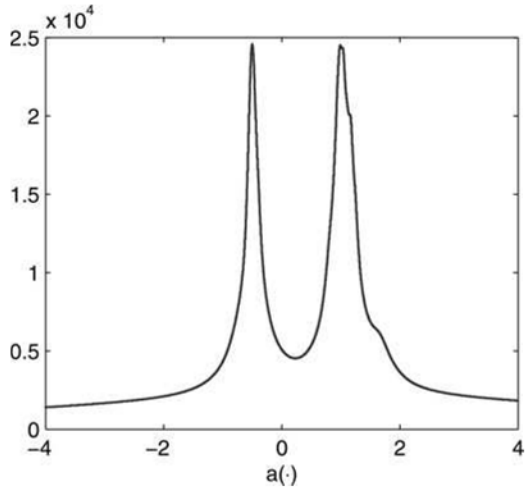

Figure 5 *Cost function of (4) for* $\eta = 0.001$

Although the maximisation of the cost function (4) is intractable from an analytical point of view, iterative optimisation methods, for example the gradient method, can help to extract the values of $a(\theta_{(\cdot)})$ through the optimisation of (4). Based on the hybrid decoupling constraint (HDC) proposed by Vidal *et al.*, one can define another cost function, similar to the one in (4), that presents the advantage to be well posed in terms of analytical tractability. The HDC equation conveys all the information about the different operating modes and is also independent of the switching mechanism of the system. For the system of (1), the HDC is expressed as

$$(y_{k+1} - a_1 u_k)(y_{k+1} - a_2 u_k) = 0 \tag{5}$$

From (5), in order to estimate a_1 and a_2, a new cost function can be defined as

$$\ell(a_1, a_2) = \sum_{k=1}^{N}((y_{k+1} - a_1 u_k)(y_{k+1} - a_2 u_k))^2 \tag{6}$$

The model parameters a_1 and a_2 are obtained by the minimisation of the cost function (6). In fact, by using the substitution $S = a_1 + a_2$ and $P = a_1 a_2$, it is easy to show that S and P are solutions of the (7)

$$\begin{pmatrix} \sum_{k=1}^{N} y_{k+1}^2 u_k^2 & \sum_{k=1}^{N} y_{k+1} u_k^3 \\ \sum_{k=1}^{N} y_{k+1} u_k^3 & \sum_{k=1}^{N} u_k^4 \end{pmatrix} \begin{pmatrix} S \\ P \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{N} y_{k+1}^3 u_k \\ \sum_{k=1}^{N} y_{k+1}^2 u_k^2 \end{pmatrix} \tag{7}$$

The model parameters a_1 and a_2 can be easily assessed from S and P.

This introductory example illustrates a particular case of multiple-operating mode systems and the way to characterise the operating modes. In the general case, the switched regression model is given by (8)

$$\psi_{\mu_k}^T X_k = 0 \tag{8}$$

where $X_k = [-y_k \quad y_{k-1} \cdots y_{k-n_y} \quad u_{k-1} \cdots u_{k-n_u} \quad 1]^T$ is the extended regression vector, n_u and n_y are the model orders. The variable $\mu_k \in \{1, 2, \ldots, s\}$ is the discrete state that determines the local regression model describing the system at each time instant, with s being the number of operating modes. The parameters vector ψ_i, $i \in \{1, 2, \ldots, s\}$, is the regression parameters vector of the ith operating mode. Changes of operating modes are generally the result of a change in the operating conditions (a change in the feed grade of a separation unit for example) or a change in the control strategy (supervisory control scheme for example). The model of (8) has many effective applications in various engineering problems such as air traffic management [23], video segmentation [24] and process control [25]. In the remaining part of this paper, the orders of the model and the number of operating modes s are assumed known. The

problem under investigation is then the determination of the model parameters θ_i, $i \in \{1, 2, \ldots, s\}$ from a collection of input/output data of the system. For the sake of clarity, the method is first discussed on a switched regression model with two operating modes and then extended to the general case.

3 Parameters estimation

3.1 Underlying principle

In (8), let us denote the ith component of the regression vector $X_k \in \mathbb{R}^d$, $d = n_u + n_y + 2$, by x_i. The jth component of the ith parameters vector $\psi_i \in \mathbb{R}^d$, $i \in \{1, 2, \ldots, s\}$ is denoted ψ_{ij}

$$\begin{aligned} X_k &= [x_1 \ldots x_i \ldots x_d]^T \\ \psi_i &= [\psi_{i1} \ldots \psi_{ij} \ldots \psi_{id}]^T \end{aligned} \tag{9}$$

Note the, by definition of the equation model of (8), $\psi_{i1} = 1$. The HDC equation for the switched regression model of (8) is given by

$$\prod_{i=1}^{s} (\psi_i^T X_k) = 0 \tag{10}$$

The HDC equation is in fact the product of the local regression models describing the different operating modes. It can be expressed in a polynomial form [19] constructed from the combinations in $\mathbb{R}^s$ of the regression variables x_i, $i = 1, \ldots, d$. The polynomial expression of the HDC equation is given by

$$\sum_{n_1 \ldots n_d} p_{n_1 \ldots n_d} x_1^{n_1} \ldots x_d^{n_d} = 0 \tag{11}$$

with the constraints

$$\begin{cases} n_1 + n_2 + \cdots + n_d = s \\ 0 \le n_i \le s, \quad i = 1, \ldots, d \end{cases} \tag{12}$$

The coefficients $p_{n_1 \ldots n_d}$ depend on the local model parameters ψ_i, $i = 1, \ldots, s$. The idea proposed in [8] consists of first identifying the parameters $p_{n_1 \ldots n_d}$, and then deducing the local model parameters ψ_i, $i = 1, \ldots, s$ from the coefficients $p_{n_1 \ldots n_d}$. The first step of the identification procedure is equivalent to an identification procedure from available measurements, whereas the second step amounts to a problem of model inversion in which the local models parameters are expressed as functions of the parameters $p_{n_1 \ldots n_d}$. A new approach is presented here for the step of model inversion that allows us to obtain the parameters of the local models from the coefficients $p_{n_1 \ldots n_d}$.

3.2 Case of a two modes switched regression model

Consider the switched regression model of (8) with $n_u = 1$ and $n_y = 1$. The HDC equation for such a model is given by

$$(\psi_1^T X_k)(\psi_2^T X_k) = 0 \tag{13}$$

with $\psi_i = [\psi_{i1}\ \psi_{i2}\ \psi_{i3}]$, $i \in \{1, 2\}$ and $X_k = [x_1\ x_2\ x_3]$. Equation (13) is written in a developed form as

$$\begin{aligned} &(p_{200})x_1^2 + (p_{110})x_1x_2 + (p_{101})x_1x_3 \\ &\quad + (p_{020})x_2^2 + (p_{011})x_2x_3 + (p_{002})x_3^2 = 0 \end{aligned} \tag{14}$$

Using the available measurements X_k, $k = 1, \ldots, N$, the coefficients p_{200}, p_{110}, p_{101}, p_{020}, p_{011} and p_{002} can easily be estimated with a linear estimation technique by minimising a cost function of the type

$$\ell(p_{n_1 \ldots n_d}) = \sum_{k=1}^{N} \left(\sum_{n_1 \ldots n_d} p_{n_1 \ldots n_d} x_1^{n_1} \ldots x_d^{n_d} \right)^2 \tag{15}$$

In fact, one does not even need to estimate the coefficient p_{200} as this coefficient is always equal to 1 because of the fact that ψ_{i1} is always equal to 1 by definition of the structures of $\psi_{\mu_k}^T$ and X_k.

By identifying the two polynomials of (13) and (14), the model inversion step, which involves the estimation of the model parameters θ_i, $i = 1, \ldots, s$ from the coefficients $p_{n_1 \ldots n_d}$, can be analytically addressed. This leads to the following set of equations that explicitly shows the relations between the model parameters ψ_{ij}, $i = 1 \ldots, s$, $j = 1 \ldots, d$ and the coefficients $p_{n_1 \ldots n_d}$

$$p_{200} = \psi_{11}\psi_{21} \tag{16}$$
$$p_{110} = \psi_{11}\psi_{22} + \psi_{12}\psi_{21} \tag{17}$$
$$p_{101} = \psi_{11}\psi_{23} + \psi_{13}\psi_{21} \tag{18}$$
$$p_{020} = \psi_{12}\psi_{22} \tag{19}$$
$$p_{011} = \psi_{12}\psi_{23} + \psi_{13}\psi_{22} \tag{20}$$
$$p_{002} = \psi_{13}\psi_{23} \tag{21}$$

Equations (16)–(21) can be combined in order to extract the values of the pairs of ratios $(\psi_{11}/\psi_{12}; \psi_{21}/\psi_{22})$, $(\psi_{12}/\psi_{13}; \psi_{22}/\psi_{23})$ and $(\psi_{13}/\psi_{11}; \psi_{23}/\psi_{21})$. For instance, the extraction of $(\psi_{11}/\psi_{12}; \psi_{21}/\psi_{22})$ is performed with (16), (17) and (19)

$$\begin{cases} p_{200} = \psi_{11}\psi_{21} \\ p_{110} = \psi_{11}\psi_{22} + \psi_{12}\psi_{21} \\ p_{020} = \psi_{12}\psi_{22} \end{cases} \tag{22}$$

By combining (22), the ratios ψ_{11}/ψ_{12} and ψ_{21}/ψ_{22} are then

obtained as the roots of the polynomial (23) in α_{12}

$$p_{020}\alpha_{12}^2 - p_{110}\alpha_{12} + p_{200} = 0 \tag{23}$$

Similarly, the ratios ψ_{12}/ψ_{13} (respectively, ψ_{13}/ψ_{11}) and ψ_{22}/ψ_{23} (respectively, ψ_{23}/ψ_{21}) can be inferred from (19)–(21) (respectively, (16), (18) and (21)) as the roots of the polynomial (24) (respectively, (25))

$$p_{002}\alpha_{23}^2 - p_{011}\alpha_{23} + p_{020} = 0 \tag{24}$$

$$p_{200}\alpha_{31}^2 - p_{101}\alpha_{31} + p_{002} = 0 \tag{25}$$

The resolution of (23)–(25), respectively, provides two solutions for α_{12}, α_{23} and α_{31}, which leads to eight potential models. However, because of the inherent definition of the ratios α_{ij}, $i = 1, \ldots, s$, $j = 1, \ldots, d$, the quantity

$$\tau = \alpha_{12}\alpha_{23}\alpha_{31} - 1 \tag{26}$$

must be equal to zero. The constraint (26) is then used to determine the ratios related to the true model parameters.

Numerical example: In order to give more insight into this problem of root selection, let us assume that, for a given system with multiple operating modes, the estimation of the coefficients of the polynomial of (11) leads to the following polynomial equation

$$x_1^2 - 5x_1x_2 - 3x_1x_3 + 6x_2^2 + 5x_2x_3 - 4x_3^2 = 0 \tag{27}$$

Equation (27) can be compared with (14) in order to obtain the correspondence with the parameters $p_{n_1 \ldots n_d}$. The resolution of the polynomial (23)–(25) gives the following ratios as the roots of the polynomial equations

$$\alpha_{12} = -\frac{1}{2} \text{ or } \frac{-1}{3}$$

$$\alpha_{23} = -2 \text{ or } \frac{3}{4}$$

$$\alpha_{31} = -4 \text{ or } 1$$

Eight possible models can then be identified. The constraint $\tau = 0$ in (26) is used to determine the acceptable solutions. Table 1 shows all the values of the test variable for all possible roots combinations. From Table 1, one can see that there are only two acceptable solutions, which are indicated by the value $\tau = 0$ (grey cells in Table 1). The first solution is

$$\psi_{11}/\psi_{12} = -1/2,\ \psi_{12}/\psi_{13} = -2,\ \psi_{13}/\psi_{11} = 1$$

which also gives: $\psi_{12} = -2\psi_{11}$ and $\psi_{13} = \psi_{11}$. This allows us to express the first model as

$$\psi_{11}(x_1 - 2x_2 + x_3) = 0$$

The second solution is

$$\psi_{21}/\psi_{22} = -1/3, \quad \psi_{22}/\psi_{23} = 3/4, \quad \psi_{23}/\psi_{21} = -4$$

which leads to: $\psi_{22} = -3\psi_{21}$ and $\psi_{23} = -4\psi_{21}$. The second model is then written as

$$\psi_{21}(x_1 - 3x_2 - 4x_3) = 0$$

Finally, the identified switched model is given by

$$x_1 - 2x_2 + x_3 = 0$$
$$x_1 - 3x_2 - 4x_3 = 0$$

3.3 General algorithm

For the general switched regression model of (8), the estimation of the local regression model parameters is done in five steps:

- Step 0: Collect the input/output data and build the regressors $x_1^{n_1} \ldots x_d^{n_d}$.
- Step 1: Compute the coefficients $p_{n_1 \ldots n_d}$ from (11). The number of coefficients $p_{n_1 \ldots n_d}$ to be estimated is equal to $n_p = \binom{s+d-1}{s}$.
- Step 2: For each local regression model, compute the ratios (ψ_{1i}/ψ_{1j}), (ψ_{2i}/ψ_{2j}), $\ldots$, (ψ_{si}/ψ_{sj}), with $i = 1, \ldots, d-1$, $j = j+1$ and for $i = d$, $j = 1$, as the roots of the polynomial equations (28) in α

$$c_s\alpha^s + c_{s-1}\alpha^{s-1} + \cdots + c_1\alpha + c_0 = 0 \tag{28}$$

The coefficients $c_s, \ldots, c_0$ of the polynomial equations (28) are obtained from the previously calculated coefficients

Table 1 Test of the possible solutions

α_{12}	−1/2	−1/3	−1/2	−1/3	−1/2	−1/3	−1/2	−1/3
α_{23}	−2	−2	−2	−2	3/4	3/4	3/4	3/4
α_{31}	−4	−4	1	1	−4	−4	1	1
τ	−5	−11/3	0	−1/3	1/2	0	−11/8	−5/4

$p_{n_1 \ldots n_d}$ and given by

$$
\begin{aligned}
c_s &= (-1)^0 p_{\underbrace{0\ldots0n_i}_{i}\, n_j 0\ldots0} && n_i = 0,\ n_j = s \\
c_{s-1} &= (-1) p_{\underbrace{0\ldots0n_i}_{i}\, n_j 0\ldots0}, && n_i = s - n_j,\ n_j = s-1 \\
&\vdots && \vdots \\
c_l &= (-1)^{(s-l)} p_{\underbrace{0\ldots0n_i}_{i}\, n_j 0\ldots0}, && n_i = s - n_j,\ n_j = l \\
&\vdots && \vdots \\
c_1 &= (-1)^{(s-1)} p_{\underbrace{0\ldots0n_i}_{i}\, n_j 0\ldots0}, && n_i = 1 - n_j,\ n_j = 1 \\
c_0 &= (-1)^{s} p_{\underbrace{0\ldots0n_i}_{i}\, n_j 0\ldots0}, && n_i = s,\ n_j = 0
\end{aligned}
\tag{29}
$$

- Step 3: Use the constraint

$$\tau = \left(\prod_{j=1}^{d-1} (\psi_{ij}/\psi_{i(j+1)}) \right) (\psi_{id}/\psi_{i1}) - 1 = 0 \tag{30}$$

to select the solutions corresponding to the parameters of the local models.

- Step 4: Extract the local model parameters from the selected ratios.

The derivation of the coefficients $c_s, \ldots, c_0$ of the polynomial equations (28) is straightforward. In fact, one can easily verify that while identifying the relations between the coefficients $p_{n_1 \ldots n_d}$ of the HDC equation and the local models parameters ψ_{ij}, $i = 1, \ldots, s$, $j = 1, \ldots, d$, for $n_i \neq 0$ and $n_j \neq 0$, the coefficients $p_{\underbrace{0\ldots0n_i}_{i}\, n_j 0\ldots0}$ as given by (29) are the sum of terms that are the product of n_i parameters $\Psi_{i_1 i}$, $i_1 \in S_1$, $S_1 \subset \{1, 2, \ldots, s\}$, and n_j parameters $\Psi_{i_2 j}$, $i_2 \in S_2$, $S_2 \subset \{1, 2, \ldots, s\}$, with $S_1 \cap S_2 = \emptyset$. For $n_i = 0$ (respectively, $n_j \neq 0$), the coefficients $p_{\underbrace{0\ldots0n_i}_{i}\, n_j 0\ldots0}$ are simply equal to $\Psi_{1(i+1)} \Psi_{2(i+1)} \ldots \Psi_{s(i+1)}$ (respectively, $\Psi_{1i} \Psi_{2i} \ldots \Psi_{si}$). From there, the sums of the product from one by one until s by s of the ratios of step 2 can always be retrieved from the coefficients $p_{\underbrace{0\ldots0n_i}_{i}\, n_j 0\ldots0}$. Finally, the polynomial equations (28) is obtained by means of the Viète's formulas [26], which relate the roots of a polynomial to its coefficients.

3.4 On the presence of measurement noise

The presence of measurement noise in the input/output identification dataset generally introduces some estimation errors for the coefficients $p_{n_1 \ldots n_K}$, leading to the violation of the constraint $\tau = 0$. This situation is handled by the usage of a user-defined threshold or by selecting the configurations corresponding to the s smallest absolute values of τ. For example, let us consider the polynomial equation (27) of the previous numerical example with a variation in its coefficients resulting from the influence of measurement noise on the estimation of the coefficients $p_{n_1 \ldots n_d}$

$$x_1^2 - 5.3 x_1 x_2 - 3.1 x_1 x_3 + 6.5 x_2^2 + 5.2 x_2 x_3 - 4.3 x_3^2 = 0 \tag{31}$$

By using the coefficients $p_{n_1 \ldots n_d}$ in (31), the resolution of the polynomial (23)–(25) gives the following solutions

$$
\begin{aligned}
\alpha_{12} &= -0.519 \text{ or } -0.297 \\
\alpha_{23} &= -1.977 \text{ or } 0.763 \\
\alpha_{31} &= -4.133 \text{ or } 1.042
\end{aligned}
$$

The resulting values of τ are shown in Table 2. All the values of τ are different from zero. However, by identifying the configurations corresponding to the two smallest absolute values of τ (grey cells in Table 2), the acceptable solutions are $\psi_{11}/\psi_{12} = -0.519$, $\psi_{12}/\psi_{13} = -1.977$, $\psi_{13}/\psi_{11} = 1.042$ and $\psi_{21}/\psi_{22} = -0.297$, $\psi_{22}/\psi_{23} = 0.763$, $\psi_{23}/\psi_{21} = -4.133$, which can be compared with the values previously obtained in the numeric example.

Remark 1: It is obvious that the effectiveness of the identification procedure largely depends on how representative the identification dataset is in regard to the operating regimes of the system. This property is linked to the characterisation of the set of sufficiently exciting input for switched operating regime systems. Although this issue still remains opened, some answers can be found in [24, 27].

Remark 2: In the presence of outliers, the estimated coefficients $p_{n_1 \ldots n_d}$ can become less accurate and eventually

Table 2 Test of the possible solutions

α_{12}	−0.519	−0.297	−0.519	−0.297	−0.519	−0.297	−0.519	−0.297
α_{23}	−1.977	−1.977	−1.977	−1.977	0.763	0.763	0.763	0.763
α_{31}	−4.133	−4.133	1.042	1.042	−4.133	−4.133	1.042	1.042
τ	−5.239	−3.424	0.068	−0.389	0.637	−0.064	−1.412	−1.236

render inoperative the roots selection procedure. This scenario fuels the development of a robust approach.

Remark 3: When the model structure is not known a priori, the results of the estimation procedure could be used to detect structural errors in the model. Indeed, the parameters $p_{\underbrace{0\ldots0}_{i}n_i n_j 0\ldots0}$, $n_i = s$, $n_j = 0$, are generally different from zero. The presence of parameters $p_{\underbrace{0\ldots0}_{i}n_i n_j 0\ldots0}$, $n_i = s$, $n_j = 0$, with a value equal to zero (noise-free case) or close to zero (case of noisy measurements) is an indication of an over-sized regression vector. For instance, in the two modes switched system example, if the value of the parameter p_{002} is equal to zero, one can suspect an over-sized regression vector as ψ_{13} or ψ_{23} will be equal to zero. The model dimension should then be reduced. An under-sized regression vector is indicated by an impossibility to solve some of the polynomial (28) as they are never satistified. Future work will consider how to use these tools for extending the proposed method to the case where the number of operating modes and the model structure are unknown.

4 Robust estimation procedure

In the parameter estimation procedure presented in Section 3, the first stage of the method for the characterisation of a system with multiple operating modes has consisted of estimating from the HDC equation, the parameters $p_{n_1\ldots n_K}$ of a single model characterising all the operating modes. In general, the least squares method is inefficient for this estimation in the presence of outliers in the identification dataset. In this section, a robust technique that tolerates the presence of outliers in the identification dataset is proposed. This robust approach allows us to render the whole identification procedure insensitive to the presence of outliers.

4.1 Principle of the robust approach

In a deterministic scenario (noise-free scenario), the HDC equation is given by (10) and it is used to retrieve the coefficients $p_{n_1\ldots n_K}$. In the more realistic situation where different sources of errors are considered, the HDC equation is modified into the form (32)

$$\prod_{i=1}^{s}(\psi_i^T X_k) = \varepsilon_k \tag{32}$$

where ε_k is the resulting equation error, which accounts for the presence of different sources of errors. By using the polynomial form (11) of the HDC equation, (32) can be written as

$$\sum_{n_1\ldots n_d} p_{n_1\ldots n_d} x_1^{n_1}\ldots x_d^{n_d} = \varepsilon_k \tag{33}$$

For sake of clarity in the following development, let us denote the coefficients $p_{n_1\ldots n_K}$ and the variables $x_1^{n_1}\ldots x_d^{n_d}$, respectively, by b_i, $i = 1, \ldots, n_p$ and z_{ik}, $i = 1, \ldots, n_p$ with $n_p = \binom{s+d-1}{s}$. Equation (33) is then simply rewritten as

$$\sum_{i=1}^{n_p} b_i z_{ik} = \varepsilon_k \tag{34}$$

The variable ε_k is assumed to be a realisation of a random variable that can accommodate two types of errors. The first type involves errors of low magnitude and centred around zero; the second type relates to large magnitude errors, essentially of accidental nature, and that are not necessarily of zero-mean values. Given the presence of these two types of error, the probability density function (pdf) of the variable ε_k is chosen in an additive form depending on a blending factor λ. Hence, for the kth measurement, the pdf of ε_k is written as [28]

$$p(\varepsilon_k) = \lambda p_1(\varepsilon_k) + (1-\lambda)p_2(\varepsilon_k) \tag{35}$$

The partial pdf $p_1(\cdot)$ and $p_2(\cdot)$ are given by

$$p_1(\varepsilon_k) = \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{1}{2}\left(\frac{\varepsilon_k}{\sigma_1}\right)^2\right) \tag{36}$$

$$p_2(\varepsilon_k) = \frac{1}{\sqrt{2\pi}\sigma_2}\exp\left(-\frac{1}{2}\left(\frac{\varepsilon_k}{\sigma_2}\right)^2\right) \tag{37}$$

with $\sigma_2 \gg \sigma_1$. The parameters λ, σ_1 and σ_2 characterise the shape of the pdf $p(\cdot)$ known as a contaminated distribution function. The blending factor λ accounts for the proportion of outliers present in the dataset. An empirical value of 0.7–0.9 seems to be well suited. In addition, the numerical results obtained throughout the various simulations have shown little sensitivity to this choice.

The partial pdfs $p_1(\cdot)$ and $p_2(\cdot)$ represent random measurement errors of low magnitude and accidental errors of large magnitude, respectively. The standard deviations σ_1 and σ_2 are used to distinguish these two pdfs. The main idea is to maintain a ratio of about 5–10 between σ_1 and σ_2. Note that this type of contaminated distribution functions has been used in the area of data validation [29] where the variables of interest for the estimation procedure are the system's variables. To refine the selection of the tuning parameters λ, σ_1 and σ_2, in a situation where the number of measurements is sufficient, it is also possible to consider them as parameters to be estimated.

By assuming the statistical independence of the errors ε_k, $k = 1, \ldots, N$, the log-likelihood function of the

identification dataset can be written as

$$\mathcal{V} = \log\left(\prod_{k=1}^{N} p(\varepsilon_k)\right) \tag{38}$$

which is equivalent to

$$\mathcal{V} = \sum_{k=1}^{N} \log\left(\lambda p_1(\varepsilon_k) + (1-\lambda)p_2(\varepsilon_k)\right) \tag{39}$$

$$\mathcal{V} = \sum_{k=1}^{N} \log\left(\lambda p_1\left(\sum_{i=1}^{n_p} b_i z_{ik}\right) + (1-\lambda)p_2\left(\sum_{i=1}^{n_p} b_i z_{ik}\right)\right) \tag{40}$$

The estimation of the parameters b_i, $i = 1, \ldots, n_p$ (which stand for the coefficients $p_{n_1 \ldots n_K}$) is then formulated as the maximisation of the log-likelihood function (39) in regard to these parameters. For a given parameter b_{i_0}, $i_0 \in \{1, \ldots, n_p\}$, using (35)–(37) and (40), the partial derivative of the log-likelihood function with respect to b_{i_0} is

$$\frac{\partial \mathcal{V}}{\partial b_{i_0}} = \sum_{k=1}^{N}\left(\omega_k\left(\sum_{i=1}^{n_p} b_i z_{ik}\right) z_{i_0 k}\right) \tag{41}$$

$$\omega_k = \frac{[p_1(\varepsilon_k)/\sigma_1^2] + (1-\lambda)[p_2(\varepsilon_k)/\sigma_2^2]}{\lambda p_1(\varepsilon_k) + (1-\lambda)p_2(\varepsilon_k)} \tag{42}$$

From the definitions of the extended regression and parameters vectors in (8), one has $b_1 = 1$ as this coefficient is the one corresponding to the term $z_1 = y_k^s$ in the HDC equation. From there, (41) can be put in the form

$$\frac{\partial \mathcal{V}}{\partial b_{i_0}} = \sum_{k=1}^{N}\left(\omega_k\left(z_{1k} + \sum_{i=2}^{n_p} b_i z_{ik}\right) z_{i_0} k\right) \tag{43}$$

which is equivalently written in the following matrix form

$$Z^T \Omega (Zb - z) = 0 \tag{44}$$

with

$$Z = \begin{pmatrix} z_{21} & z_{31} & \cdots & z_{n_p 1} \\ z_{22} & z_{32} & \cdots & z_{n_p 2} \\ \vdots & \vdots & \ddots & \vdots \\ z_{2N} & z_{3N} & \cdots & z_{n_p N} \end{pmatrix} \tag{45}$$

$$z = (z_{11} \ldots z_{1N})^T \tag{46}$$

$$\Omega = \mathrm{diag}(\omega_1, \ldots \omega_N) \tag{47}$$

$$b = (b_2 \ldots b_{n_p})^T \tag{48}$$

The parameters b_i, $i = 1, \ldots, n_p$ are therefore given by

$$\begin{cases} b_1 = 1 \\ b = (Z^T \Omega Z)^{-1} Z^T \Omega z \end{cases} \tag{49}$$

Note that in (49), the vector b depends on the diagonal matrix Ω, which is a function of $\varepsilon_{(\cdot)}$ as shown by (42) and (47). Therefore (49) is clearly a non-linear equation with respect to b. An iterative algorithm is proposed for the calculation of b from (49):

- Step 0: Start with an initial guess $b = b^0$.
- Step 1: Compute the model errors from (34): $\varepsilon_k = \sum_{i=1}^{n_p} b_i^0 z_{ik}$, $k = 1 \ldots, N$.
- Step 2: Calculate the pdfs $p_1(\cdot)$, $p_2(\cdot)$ and $p(\cdot)$ from (36), (37) and (35).
- Step 3: Compute ω_k, $k = 1, \ldots, N$ from (42) and the matrix Ω from (47).
- Step 4: Calculate b as the eigenvector corresponding to the smallest eigenvalue of $Z^T \Omega Z$.
- Step 5: Return to step 1 until $\|b^l - b^{l-1}\| < \eta$ where $\|\cdot\|$ is a vector norm, l is the iteration step and η is the solution convergence tolerance.

Note that for $\lambda = 1$, the proposed estimation procedure is exactly the ordinary least square estimation technique. For $\lambda \in (0, 1)$, the procedure performs a weighted least square estimation, the weights being given by (42). This weighting plays a fundamental role in the estimation procedure because it makes it partially insensitive to the presence of outliers that is reflected by large magnitude values of ε_k. Fig. 6 highlights the importance of this weighting function. For $\sigma_1 = 1$, $\sigma_2 = 10$, $\lambda = 0.8$ and $\varepsilon_k \in [-6, 6]$, the pdf $p(\cdot)$ (35) and the weighting function $\omega_{(\cdot)}$ (42) have been plotted for $\lambda = 0.8$ (upper graphics) and $\lambda = 1$ (lower graphics). The graphics clearly shows that for $\lambda = 1$, the weights ω_k are constant while, for $\lambda = 0.8$, these weights decrease when the absolute value of ε_k increases. The estimation method rejects outliers by assigning small weights to the corresponding data points during the estimation.

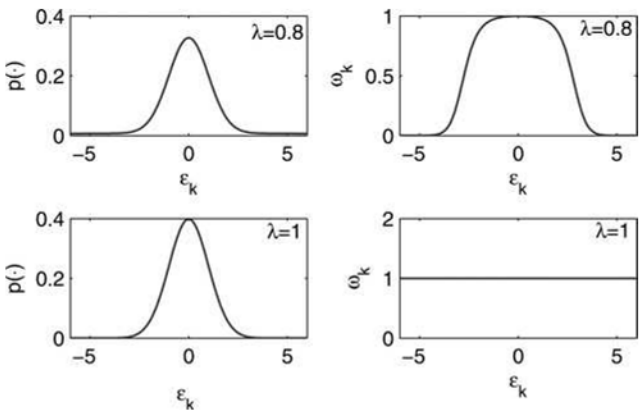

Figure 6 *pdf p(·) and weighting function $\omega_{(\cdot)}$ for $\lambda = 0.8$ and 1*

The initial guesses of the parameters σ_1, σ_2 and λ could be selected by seeking their values that minimise the log-likelihood function (38). In practice, the obtained improvement is not quite significant and a few trials are generally sufficient to obtain satisfactory values for σ_1, σ_2 and λ.

The approach proposed here for the parameters estimation of a switched regression model can generally be viewed as an alternative to the approach presented in [8, 24]. The performances of the two approaches are the same in a deterministic situation with the advantage of a recursive implementation for the one in [8, 24]. Recursive implementation of the approach presented here will be considered in future work. Both approaches take into account the presence of measurement noise by resorting to different tuning parameters. Robustness to outliers is integrated to the approach proposed in this paper.

5 Illustrative examples

5.1 Example 1

Consider the following switched regression model [17]

$$y_k = \begin{cases} \theta_1^T \varphi_k + e_k & \text{if } 4y_{k-1} - u_{k-1} + 10 < 0 \\ \theta_2^T \varphi_k + e_k & \text{if } 4y_{k-1} - u_{k-1} + 10 \geq 0 \\ & \text{and } 5y_{k-1} + u_{k-1} - 6 \leq 0 \\ \theta_3^T \varphi_k + e_k & \text{if } 5y_{k-1} + u_{k-1} - 6 > 0 \end{cases} \quad (50)$$

where $e_{(\cdot)}$ is a measurement noise, $\varphi_k = [y_{k-1} \quad u_{k-1} \quad 1]^T$, $\theta_1^T = [-0.4 \quad 1 \quad 1.5]$, $\theta_2^T = [0.5 \quad -1 \quad -0.5]$ and $\theta_3^T = [-0.3 \quad 0.5 \quad -1.7]$. The focus here is on the estimation of the model parameters θ_1, θ_2 and θ_3. Let us point out that, once the model parameters θ_1, θ_2 and θ_3 have been estimated, the estimation of the parameters of the switching mechanism, defined by the inequalities in model (50), can still be performed by resorting to two-class or multi-class linear separation techniques [30, 31]. Practical applications of the model structure presented in (50) can be found in [10, 17, 24, 32].

The input signal $u_{(\cdot)}$ is generated as the realisations of an uniformly distributed random variable on the interval $[-5, 5]$; the measurement noise $e(\cdot)$ is simulated as an uniform distribution on the interval $[-0.1, 0.1]$. The generated dataset constitutes of 200 data points and is corrupted by outliers in a proportion of approximatively 10% of the total number of data points. The presence of outliers can be appreciated in Fig. 7, which shows the available identification data points in contrast to the hyperplanes representing the three operating modes. In Fig. 7, the hyperplanes are presented without taking into account their respective domain of validity. The estimation of the model parameters θ_1, θ_2 and θ_3 is done using the technique presented in Section 3. The step of the estimation of the coefficients $p_{n_1} \ldots p_{n_d}$ is done in two different manners with a TLS estimation technique [33] and the robust procedure (RP) presented in Section 4. The tuning parameters for the RP have been arbitrarily set at $\lambda = 0.8$, $\sigma_1 = 2$ and $\sigma_2 = 20$. Table 3 presents the parameters estimation errors obtained without ($\Delta\theta_i^{\text{TLS}}$) and with the robust approach ($\Delta\theta_i^{\text{RP}}$). The estimation error $\Delta\theta_i^{(\cdot)}$ for a parameter θ_i is calculated as the difference between the true parameter θ_i and the estimated one $\hat{\theta}_i^{(\cdot)}$. The improvement introduced by the use of the robust approach can be clearly noticed while comparing the different estimation errors. Let us mention that during the conducted simulations, it has been observed that the proposed method fails to recover the model parameters when the number of outliers is relatively large (approximately greater than 30% of the number identification data points). Fig. 8 shows the evolution of the weighting function of (42) for this simulation. The zero values of this function correspond to the measurements detected as outliers during the

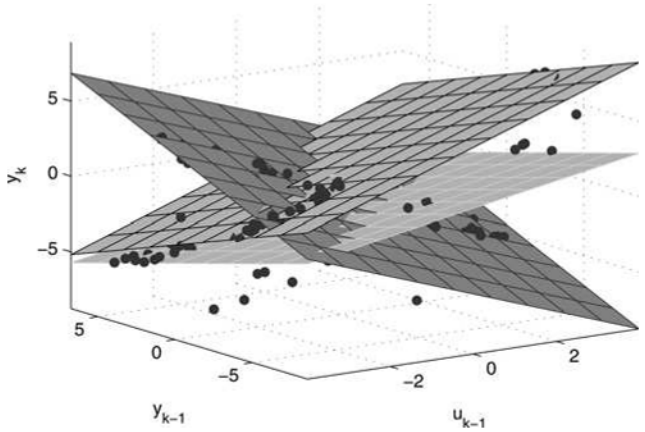

Figure 7 *Hyperplanes and data points (with outliers)*

Table 3 Parameters estimation errors with and without the robust approach

$\Delta\theta_1^{\text{TLS}}$	$\Delta\theta_1^{\text{RP}}$	$\Delta\theta_2^{\text{TLS}}$	$\Delta\theta_2^{\text{RP}}$	$\Delta\theta_3^{\text{TLS}}$	$\Delta\theta_3^{\text{RP}}$
0.325	0.006	2.889	−0.002	−0.082	−0.006
−0.913	−0.074	−1.301	−0.098	−0.529	−0.028
−1.315	−0.206	4.010	−0.154	−6.591	0.079

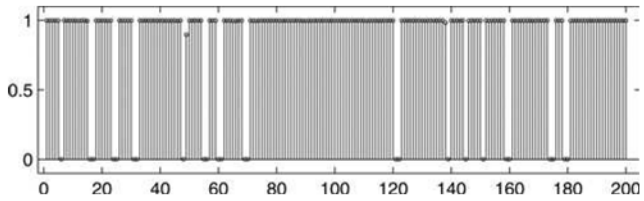

Figure 8 *Weights*

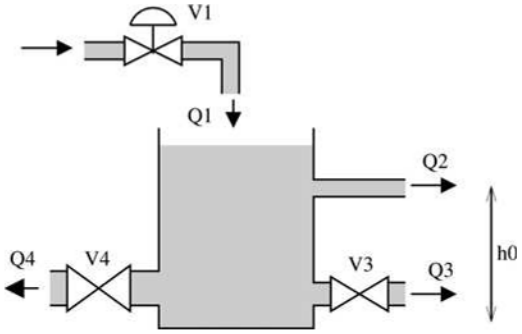

Figure 9 *Separation unit*

identification process and that are then automatically discarded from the identification procedure. In a fault detection perspective, the weighting function can be used to detect and locate outliers.

5.2 Example 2

Consider the simplified separation unit example presented in Fig. 9. The feed of the unit, which is a mixture of different products from upstream process, is introduced through the control valve V1. The valve V1 is controlled in such a way that the level of the mixture inside the unit is always above the reference level $h0$. The feed mixture is separated into two distinct phases by the means of gravity separation. The light phase, which is the final product, leaves the unit through the middling stream with a flow rate $Q2$. The heavy phase settles at the bottom of the vessel and exits the units through either the valves V3 or V4 for additional processing in downstream processes. The valves V3 and V4 are ON−OFF valves that are used to produce two different types of products. Indeed, opening either the valve V3 or the valve V4 changes the composition of the light and heavy phases as the residence time of the different particles present in the feed is more or less reduced. So, when the valve V3 is opened, the unit produces a product of grade A and when the valve V4 is opened, the unit's product is of grade B. The valves V3 and V4 are never opened simultaneously. The operating conditions leading to grade A and B products are, respectively, characterised by (51) and (52)

$$S\dot{h} = Q1 - K_2\sqrt{h - h0} - K_3\sqrt{h} \tag{51}$$

$$S\dot{h} = Q1 - K_2\sqrt{h - h0} - K_4\sqrt{h} \tag{52}$$

where h is the level of the product in the vessel, S is the surface area of the vessel, K_2, K_3 and K_4 are constants that depend on the gravity constant, the coefficients of discharge of the outlets, and their cross-sectional areas. An identification dataset has been obtained for this process by a finite discretisation of (53) and (54) with a sampling time of one second

$$S(h_k - h_{k-1}) = Q1_{k-1} - K_2\sqrt{h_{k-1} - h0} - K_3\sqrt{h_{k-1}} \tag{53}$$

$$S(h_k - h_{k-1}) = Q1_{k-1} - K_2\sqrt{h_{k-1} - h0} - K_4\sqrt{h_{k-1}} \tag{54}$$

The dataset is composed of 500 data points in an approximate proportion of 60% for the operating regime corresponding to a product of grade A and 40% for the operating regime corresponding to a product of grade B. The measurements of the level h are corrupted with a Gaussian white noise of variance 0.5. The dataset is also corrupted with outliers in a proportion of 10%. The numerical values of the physical parameters are: $h0 = 10$, $S = 2$, $K_2 = 1.063$, $K_3 = 0.620$ and $K_4 = 1.187$.

The proposed identification procedure is applied with $\lambda = 0.9$, $\sigma_1 = 4$ and $\sigma_2 = 20$. The extended regression vector in this case is $X_k = [-h_k \quad h_{k-1} \quad \sqrt{h_{k-1}} \quad \sqrt{h_{k-1} - h0} \quad Q1_{k-1}]$. Note that the model parameters to be estimated are $\psi_1 = [-1 \quad 1 \quad -K_3/S \quad -K_2/S \quad 1/S]^T$ and $\psi_2 = [-1 \quad 1 \quad -K_4/S \quad -K_2/S \quad 1/S]^T$. The estimated parameters $\hat{\psi}_1$ and $\hat{\psi}_2$ are shown in Table 4. One can noticed that the results of the identification procedure are overall good.

Table 4 True and estimated parameters

ψ_1	−1.000	1.000	−0.310	−0.532	0.500
$\hat{\psi}_1$	−1.000	0.998	−0.308	−0.536	0.491
ψ_2	−1.000	1.000	−0.594	−0.532	0.500
$\hat{\psi}_2$	−1.000	0.998	−0.607	−0.535	0.505

6 Conclusion

In this paper, an estimation procedure has been proposed for the identification of switched regression models. The proposed method provides a solution to the parameter estimation as a root selection problem from the resolution of polynomial equations. The issue of robustness to outliers in the input/output dataset is also addressed. The usage of a robust technique, based on contaminated distributions, renders the procedure readily usable in the quite realistic and practical situation in which the identification dataset is corrupted by the presence of outliers. To expand the scope of this approach, the problem of the simultaneous determination of the number of operating modes and the structure of the models associated with these different modes should be considered. An open issue is the design of sufficiently rich excitation signal for identification.

7 References

[1] HUDSON D.J.: 'Fitting segmented curves whose join points have to be estimated', *J. Am. Stat. Soc.*, 1966, **61**, pp. 1097–1129

[2] QUANDT R.E.: 'The estimation of the parameters of a linear regression system obeying two separate regimes', *J. Am. Stat. Assoc.*, 1958, **53**, (284), pp. 873–880

[3] GOLDFELD S.M., QUANDT R.E.: 'A Markov model for switching regression', *J. Econometrics*, 1973, **1**, pp. 3–16

[4] BREIMAN L.: 'Hinging hyperplanes for regression, classification and function approximation', *IEEE Trans. Inf. Theory*, 1993, **39**, (3), pp. 999–1013

[5] PUCAR P., SJÖBERG J.: 'On the hinge-finding algorithm for hinging hyperplanes', *IEEE Trans. Inf. Theory*, 1998, **44**, (3), pp. 1310–1319

[6] KIM C.J., PIGER J.M., STARTZ R.: 'Estimation of Markov regime-switching regression model with endogenous switching'. Working Papers 2003-015, Federal Reserve Bank of St. Louis, 2004

[7] HOCINE A., CHADLI M., MAQUIN D., RAGOT J.: 'A discrete-time sliding window observer for Markovian switching system'. Proc. 45th IEEE Conf. Decision and Control, 2006

[8] VIDAL R., SOATTO S., MA Y., SASTRY S.: 'An algebraic approach to the identification of a class of linear hybrid systems'. Proc. 42nd IEEE Conf. on Control and Decision, 2003, pp. 167–172

[9] SONTAG E.D.: 'Nonlinear regulation: the piecewise linear approach', *IEEE Trans. Autom. Control*, 1981, **26**, (2), pp. 346–358

[10] FERRARI-TRECATE G., MUSELLI M., LIBERATI D., MORARI M.: 'A clustering technique for the identification of piecewise affine systems', *Automatica*, 2003, **39**, (2), pp. 205–217

[11] BEMPORAD A., MORARI M.: 'Control of systems integrating logic, dynamics and constraints', *Automatica*, 1999, **35**, (3), pp. 407–427

[12] ROLL J., BEMPORAD A., LJUNG L.: 'Identification of piecewise affine systems via mixed-integer programming', *Automatica*, 2004, **40**, (1), pp. 37–50

[13] DEMPSTER A.P., LAIRD N.M., RUBIN D.B.: 'Maximum likelihood from incomplete data via EM algorithm', *J. R. Stat. Soc.*, 1977, **39**, pp. 1–38

[14] ROUSSEUW P.J.: 'Silhouettes: a graphical aid to the interpretation and validation of cluster analysis', *J. Comput. Appl. Math.*, 1987, **20**, pp. 53–65

[15] RAGOT J., MOUROT G., MAQUIN D.: 'Parameter estimation of switching piecewise linear systems'. Proc. 42nd Conf. on Decision and Control, 2003, pp. 5783–5788

[16] JULOSKI A., WEILAND S., HEEMELS M.: 'A Bayesian approach to identification of hybrid systems'. Proc. 43rd IEEE Conf. on Decision and Control, 2004, pp. 13–19

[17] BEMPORAD A., GARULLI A., PAOLETTI S., VICINO A.: 'A bounded-error approach to piecewise affine system identification', *IEEE Trans. Autom. Control*, 2005, **50**, (10), pp. 1567–1580

[18] NAKADA H., TAKABA K., KATAYAMA T.: 'Identification of piecewise affine systems based on statistical clustering technique', *Automatica*, 2005, **41**, (5), pp. 905–913

[19] VIDAL R., MA Y., SASTRY S.: 'Generalized principal component analysis (GPCA)', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2005, **27**, (12), pp. 1945–1959

[20] PAOLETTI S.: 'Identification of piecewise affine models'. PhD dissertation, Department of Information Engineering, University of Siena Siena, Italy, 2004

[21] DOMLAN E.A., RAGOT J., MAQUIN D.: 'Systèmes à commutation: recherche du mode actif, identification de la loi de commutation', *J. Eur. Syst. Autom.*, 2007, **41**, (1), pp. 7–30

[22] JOHNSON N.L., KEMP A.W., KOTZ S.: 'Univariate discrete distributions' (Wiley, 1992, 2nd edn.)

[23] HWANG I., BALAKRISHNAN H., TOMLIN C.: 'State estimation for hybrid systems: applications to air traffic tracking', *IEE Proc. Control Theory Appl.*, 2006, **153**, (5), pp. 556–566

[24] VIDAL R.: 'Recursive identification of switched ARX systems', *Automatica*, 2008, **44**, (9), pp. 2274–2287

[25] LENNARTSON B., TITTUS M., EGARDT S., PETTERSSON B.: 'Hybrid systems in process control', *IEEE Control Syst. Mag.*, 1996, **16**, (5), pp. 45–56

[26] RAHMAN Q.I., SCHMEISSER G.: 'Analytic theory of polynomials', (*London Mathematical Society Monographs Series*), (Oxford University Press, 2002)

[27] VIDAL R., ANDERSON B.D.O.: 'Recursive identification of switched ARX hybrid model: exponential convergence and persistence of excitation'. Proc. IEEE Conf. on Decision and Control, 2004, pp. 27–32

[28] ALHAJ-DIBO M., MAQUIN D., RAGOT J.: 'Data reconciliation: a robust approach using a contaminated distribution', *Control Eng. Pract.*, 2008, **16**, (2), pp. 159–170

[29] RAGOT J., CHADLI M., MAQUIN D.: 'Mass balance equilibration: a bilinear case with a robust approach using contaminated distribution', *AIChE J.*, 2005, **51**, (5), pp. 1569–1575

[30] BENNETT K.P., MANGASARIAN O.L.: 'Multicategory discrimination via linear programming', *Optim. Meth. Softw.*, 1994, **3**, pp. 27–39

[31] BREDENSTEINER E.J., BENNETT K.P.: 'Multicategory classification by support vector machines', *Comput. Optim. Appl.*, 1999, **12**, pp. 53–79

[32] JULOSKI A., HEEMELS M., FERRARI-TRECATE G.: 'Data-based hybrid modeling of the component placement process in pick-and-place machines', *Control Eng. Pract.*, 2004, **12**, (10), pp. 1241–1252

[33] VAN HUFFEL S., VANDEWALLE J.: 'The total least squares problem: computational aspects and analysis' (*Society for Industrial and Applied Mathematics*, 1991)