

# Switched affine models for describing nonlinear systems

L. Bako, K. Boukharouba, E. Duviella, S. Lecoeuche

<sup>1</sup> Univ Lille Nord de France, F-59000 Lille, France

<sup>2</sup> EMDouai, IA, F-59500 Douai, France

**Abstract:** In this work, a recursive procedure is derived for the identification of switched affine models from input-output data. Starting from some initial values of the parameter vectors that represent the different submodels, the proposed algorithm alternates between data assignment to submodels and parameter update. At each time instant, the discrete state is determined as the index of the submodel that, in term of the prediction error, appears to have most likely generated the regressor vector observed at that instant. Given the estimated discrete state, the associated parameter vector is updated based on recursive least squares. Convergence of the whole procedure although not theoretically proved, seems to be easily achieved when enough rich data are available. Finally performance is tested through some computer simulations and the modeling of an open channel system.

**Keywords:** hybrid systems, piecewise affine systems, system identification, open channel systems.

## 1. INTRODUCTION

Given a mixture of data generated by a set of interacting linear/affines discrete-time submodels, the switched affine regression problem consists in estimating the parameters associated with each of these submodels. This is known to be a challenging identification problem as it involves both data assignment to submodels and parameter estimation. It can be fairly observed that a common weakness to the majority of the existing contributions is the computational complexity. For example, feasibility of the optimization approach reported in [1] is restricted to small size problems due to its important cost. The Minimum Partition into Feasible Subsystems (Min-PFS) solution proposed in [2] is a multi-step greedy algorithm that is likely to be computationally demanding. The algebraic-geometric method [3] embeds the regression data into a higher dimensional space whose dimension increases exponentially with respect to the dimensions of the considered switched system. The clustering-based identification algorithm [4] constructs a neighboring set for each regressor and so, when the number of training data is large, a computational difficulty may arise. In the bayesian approach [5], each parameter vector is in principle regarded as a random variable and therefore substituted in the identification procedure for its probability density function, which in turn is approximated through particle filtering. However this latter approximation comes with a significant increase of computational complexity because fine precision requires the number of particles to be large.

Additionally, most of the published methods for switched systems identification are batch algorithms (see also the survey paper [6] and some more recent techniques presented in [7–9]) except the algebraic algorithm derived in [10] and further extended in [11]. However, this latter method inherits from its batch version appeared in [3], a problem of dimensionality induced by the polynomial embedding. Beyond the elements of complexity that have been observed, note that in general, the computational cost of a batch algorithm depends on the number of training data. Moreover, batch methods are not convenient

for real-time applications where the data need to be processed on-line. It is therefore of interest to develop recursive identification algorithms for hybrid systems.

In this paper, we present a simple recursive approach to the switched affine regression problem. The data are assumed to be sequentially acquired. Then, starting from some initial values, we proceed alternately to data classification and parameter update on-line. At a given time instant, the discrete state is inferred based on the information available up to that time and the parameters are accordingly refined via recursive least squares. In comparison with the batch mode methods mentioned above, it is fair to say that our method, although less effective than some of them, makes it easier to effectively handle higher dimensional data or larger amounts of data. Furthermore, it can be used for on-line identification of hybrid systems which, to the best of our knowledge, has not received much attention yet.

The outline of the paper is as follows. Problem statement is given in Section 2. We describe in Section 3 the proposed algorithm for recursive identification of switched systems with arbitrary switches. In Section 4 we study the particular case when the switched model intends to approximate a nonlinear system. In this scenario the switching mechanism takes a particular form that can also be inferred from data along with the submodels parameters. Finally, some numerical results are shown in Section 5.

## 2. PROBLEM STATEMENT

We consider a switched linear model defined by

$$y(t) = \theta_{\lambda_t}^T x(t) + e_{\lambda_t}(t), \quad (1)$$

where  $x(t) \in \mathbb{R}^n$  is referred to as the regression vector,  $y(t) \in \mathbb{R}$  is designated as the output of the model,  $\lambda_t \in \{1, \dots, s\}$  is the discrete state and  $\theta_{\lambda_t} \in \mathbb{R}^n$  is the associated parameter vector. The sequences of errors  $\{e_j(t)\}$  are assumed to be zero-mean Gaussian, i.i.d random variables. Here, sequences of errors related to two different submodels are not required to have the same variances. When dealing with the input-output

behavior of hybrid dynamical systems, the vector  $x(t)$  takes sometimes the form

$$x(t) = [y(t-1) \cdots y(t-n_a) \\ u(t-1)^\top \cdots u(t-n_b)^\top]^\top, \quad (2)$$

where  $u(t) \in \mathbb{R}^{n_u}$  and  $y(t) \in \mathbb{R}$  are respectively the input and output of the considered system,  $n_a$  and  $n_b$  are its orders. The model (1) is then designated as Switched Auto-Regressive eXogenous (SARX) model.

*Problem:* Given observations  $\{x(t), y(t)\}_{t=1}^N$  generated by a switched linear model of the form (1), with  $x(t)$  defined as in (2), we are interested here in estimating the parameter vectors  $\{\theta_j\}_{j=1}^s$ .

We start by recalling from [12] that the problem of inferring a switched model such as (1) from data, admits multiple solutions so that the identification problem is not well posed. If the structural indexes  $n_a$  and  $n_b$  are not fixed, then we can find for example a trivial switched model consisting of one single submodel with large orders that fits all the dataset. Even if we assign finite and fixed values to  $n_a$  and  $n_b$ , there are still infinitely many switched models that explain the data. For example, it can be verified that there is a switched model with  $s = N$  submodels that can reproduce the data. In order to partially alleviate the identifiability issue, we will assume in this paper that the orders  $n_a$  and  $n_b$  are finite, equal for all submodels and known a priori. It is also assumed that an upper bound  $\bar{s} \ll N$  on the number of submodels is available.

### 3. ALGORITHM DESCRIPTION

In this section we present the main contribution of this paper. From a conceptual viewpoint, the approach to be presented below is comparable to the bayesian learning algorithm of Juloski et al. [5]. The method derived therein was fairly argued by the authors to allow for the incorporation of prior physical knowledge in the identification procedure. We will see that the same argument holds also to some extent for the method suggested in the present paper with much less computational load. In the bayesian approach, each parameter vector  $\theta_j$  is treated as a random variable characterized by its probability density function  $p_{\theta_j}(\cdot)$ . Therefore, instead of estimating directly each  $\theta_j$  as a single vector, one tries to estimate the potentially nonlinear density function  $p_{\theta_j}(\cdot)$ , which is done by approximating it with a particle filtering model. From a computational viewpoint, the particle filtering approximation is equivalent to replacing each parameter vector  $\theta_j$  with a number  $M$  of particles  $\{\theta_{j,k}\}_{k=1}^M$ . Based on the estimated probability functions, the data are sequentially assigned to the competing submodels. However, this idea comes with an important cost in numerical complexity because for the particle filtering approximation to be accurate, the number  $M$  of particles needs to be large.

As an alternative to that method, we derive here a simple recursive algorithm that alternates between discrete state estimation and parameter update via recursive least squares. Unlike the bayesian procedure, only one parameter vector is updated at each time for each submodel. This results in a much less computationally expensive identification scheme.

Given an upper bound  $\bar{s} \geq s$  on the number of submodels, we pose the switched system identification problem as the problem of finding a set  $\{\theta_j\}_{j=1}^{\bar{s}} \subset \mathbb{R}^n$  of parameter vectors and an

associated switching sequence  $\{\lambda_t\}_{t=1}^N \subset \mathcal{S}$  such that the cost function

$$\mathcal{J}(\{\theta_j\}_{j=1}^{\bar{s}}, \{\lambda_t\}_{t=1}^N) = \frac{1}{N} \sum_{t=1}^N (y(t) - \theta_{\lambda_t}^\top x(t))^2 \quad (3)$$

is minimized. Finding an optimal solution to this problem may involve searching over the set  $\mathcal{S}^N = \mathcal{S} \times \cdots \times \mathcal{S}$  of all possible discrete state paths, an ideal path for which there is a set of corresponding parameter vectors that fits the data. However, this is a hard combinatorial problem.

To avoid this problem, we propose a suboptimal recursive procedure. To proceed, we first introduce a few notations. Whenever used in the paper, the *hat* symbol put above a variable will mean the estimate of that variable. For any  $j$ , we denote with  $\hat{\theta}_j(t)$  the estimate of  $\theta_j$  obtained by recursive least squares and based on the observations  $\{x(k), y(k)\}_{k=1}^t$  available up to time  $t$ . Associated with  $\theta_j(t)$  is the matrix  $L_j(t) \in \mathbb{R}^{n \times n}$  whose inverse is the correlation matrix of the regressors generated by mode  $j$ . It is recursively defined by

$$L_j^{-1}(t|\lambda_t = j) = \gamma L_j^{-1}(t-1) + x(t)x(t)^\top \\ L_j^{-1}(t|\lambda_t \neq j) = L_j^{-1}(t-1),$$

where  $\gamma \in ]0, 1[$  is a forgetting factor. Here, the notation such as  $L_j^{-1}(t|\lambda_t = j)$  means the conditional value of  $L_j^{-1}(t)$  given the knowledge that  $\lambda_t = j$ . Note that prior knowledge when available can be incorporated in the choice of the initial values of the matrix  $L_j^{-1}(0)$  (correlation matrix of the regressors generated by submodel  $j$ ) and the vector  $\hat{\theta}_j(0)$ .

If the pair of data  $(x(t), y(t))$  is generated by the submodel indexed by  $j$ , i.e. if  $\lambda_t = j$ , then  $\hat{\theta}_j$  and  $L_j$  are updated by recursive least squares as follows

$$\hat{\theta}_j(t|\lambda_t = j) = \hat{\theta}_j(t-1) + q_j(t)\varepsilon_j(t) \\ L_j(t|\lambda_t = j) = \gamma^{-1}L_j(t-1) - \gamma^{-1}q_j(t)x(t)^\top L_j(t-1), \quad (4)$$

where

$$\varepsilon_j(t) = y(t) - \hat{\theta}_j(t-1)^\top x(t) \\ q_j(t) = \frac{L_j(t-1)x(t)}{\gamma + x(t)^\top L_j(t-1)x(t)} \quad (5)$$

But when the discrete state  $\lambda_t$  is known to be different from the index  $j$ , then there is no update for  $\hat{\theta}_j$  and  $L_j$ .

$$\hat{\theta}_j(t|\lambda_t \neq j) = \hat{\theta}_j(t-1) \\ L_j(t|\lambda_t \neq j) = L_j(t-1). \quad (6)$$

Since  $\lambda_t$  is unknown, we need to replace it by its estimate in the previous equations. The discrete state is therefore estimated as the index of the submodel that most presumably generates the pair of data  $(x(t), y(t))$  in the sense of a certain decision criterion  $\mathcal{D}_j(t)$  that is,

$$\hat{\lambda}_t = \arg \min_{j=1, \dots, \bar{s}} \mathcal{D}_j(t). \quad (7)$$

The criterion  $\mathcal{D}_j(t)$  can be chosen in many different ways. Since no model has been specified for the discrete state dynamics in the SARX (1),  $\mathcal{D}_j(t)$  is taken here as a function of the fitting prior error  $y(t) - \hat{\theta}_j(t-1)^\top x(t)$  or the posterior error  $y(t) - \hat{\theta}_j(t|\lambda_t = j)^\top x(t)$ . This way, two decision functions are studied here:

- The prior error based function

$$\mathcal{D}_j^{\text{prior}}(t) = \frac{\left| \widehat{\theta}_j^\top(t-1)^\top \bar{x}(t) \right|}{\left\| \widehat{\theta}_j(t-1) \right\|_2}, \quad (8)$$

with  $\widehat{\theta}_j(t) = [-1 \ \widehat{\theta}_j(t)^\top]^\top$  and  $\bar{x}(t) = [y(t) \ x(t)^\top]^\top$ ,

- The posterior error based function

$$\mathcal{D}_j^{\text{post}}(t) = \frac{\left| \widehat{\theta}_j^\top(t|\lambda_t = j)^\top \bar{x}(t) \right|}{\left\| \widehat{\theta}_j(t|\lambda_t = j) \right\|_2}. \quad (9)$$

The criteria (8) and (9) can be interpreted as the distances from  $\bar{x}(t)$  to the linear hyperplanes that are normal to the vectors  $\widehat{\theta}_j(t-1)$  and  $\widehat{\theta}_j(t|\lambda_t = j)$  respectively.

Before summarizing the identification algorithm, we need first to make the following two assumptions.

*Assumption 1.* For any time index  $t$ , and for any couple  $(i, j) \in \mathcal{S}^2$  of mode indexes,

$$\frac{\left| \bar{\theta}_i^\top \bar{x}(t) \right|}{\left\| \bar{\theta}_i \right\|} = \frac{\left| \bar{\theta}_j^\top \bar{x}(t) \right|}{\left\| \bar{\theta}_j \right\|} \implies i = j, \quad (10)$$

where  $\bar{\theta}_j = [-1 \ \theta_j^\top]^\top$ .

*Assumption 2.* For any time index  $t$ ,

$$\frac{\left| e_{\lambda_t}(t) \right|}{\left\| \theta_{\lambda_t} \right\|} = \min_{j=1, \dots, s} \frac{\left| \bar{\theta}_j^\top \bar{x}(t) \right|}{\left\| \bar{\theta}_j \right\|}. \quad (11)$$

The purpose of Assumption 1 is obviously to remove any ambiguity in the inference of the discrete state if the true parameters were exactly known. Assumption 2 insures that if the estimates  $\widehat{\theta}_j$  had to converge to the true values then the discrete state will, despite the presence of noise, remain well estimated after the convergence.

---

#### Algorithm 1 Recursive identification of SARX models

---

- Initialization : Set  $L_j(0) = p_0 I$ , with  $p_0 > 1$  and draw different values for  $\widehat{\theta}_j(0)$ ,  $j = 1, \dots, \bar{s}$  at random or following a certain heuristic. Some prior knowledge when available, can also be used in this initialization step.

**For**  $t = 1, \dots, \infty$

- Estimate the discrete state as in (7).
- Update  $\widehat{\theta}_{\lambda_t}(t)$  and  $L_{\lambda_t}(t)$  using Eq. (4).
- The other parameter vectors are maintained unchanged :  
 $\forall j \neq \lambda_t, \left\{ \widehat{\theta}_j(t), L_j(t) \right\} = \left\{ \widehat{\theta}_j(t-1), L_j(t-1) \right\}.$

**EndFor**

---

*Remark 3.* In Algorithm 1, the true number  $s$  of submodels need not be known. Given an upper bound  $\bar{s} \ll N$  on the number of submodels, the algorithm is able to operate with  $\bar{s}$  initially presumed submodels and determine  $s$  at the end, as the number of modes to which a significant number of data have been affected.

By following the identification procedure described above, it is obvious that quality of data assignment depends heavily on the number of competing modes and how distinguishable they are from each other. Recall that at the beginning, neither the parameters nor the discrete state sequence are known. Therefore, there is such a coupling between the tasks of mode

estimation and parameter identification that initialization may play an important role in the convergence. However, since the estimation of the discrete state results from a comparison between some quantities (involving the estimates), for  $\lambda_t$  to be equal to  $\lambda_t$ , it is not necessary that the  $\widehat{\theta}_i(t)$  be equal to the true  $\bar{\theta}_i$ ,  $i = 1, \dots, s$ . In fact, the determination of the discrete state will be correct if for all  $i$ ,  $\exists K < \infty, \forall t \geq K, \lambda_t = i \implies \mathcal{D}_i(t) < \mathcal{D}_j(t), \forall j \neq i$ . And as long as  $\lambda_t$  is correctly estimated, convergence of  $\widehat{\theta}_i$  can be ensured (from RLS identifier properties [13]) if the regressors related to each discrete state obey a certain persistence of excitation condition.

## 4. APPLICATION TO NONLINEAR SYSTEM MODELING

In this section we will approximate the input-output behavior of a smooth dynamical nonlinear system with the switched linear model described previously (this is in fact the main purpose of PieceWise Affine (PWA) models [14]). We will particularly give a simple way to estimate the switching law along with the parameters. The considered nonlinear system is represented by a NARX of the form

$$y(t) = f(x(t)) + \eta(t), \quad (12)$$

where  $x(t)$  is defined as in (2),  $f$  is a continuous nonlinear function defined on an open set  $\mathcal{X} \subset \mathbb{R}^n$ , with  $n = n_a + n_u n_b$ . In (12)  $\eta(t)$  denotes the modeling error which is supposed to be a zero-mean noise.

It is intuitive that trying to capture the global behavior of a nonlinear system such as (12) in a single mathematical model is likely to result in a highly complicated model structure. So, such a model may be difficult to exploit in practice. An alternative is to try to model nonlinear systems as an interaction between a certain number of linear/affine submodels, each of which is related to a region of the system operating space. The main advantage in doing so is that linear models are simple and there exists an abundant theory for controlling and analysing them.

Our strategy amounts to making a linearization around some nominal points by taking the first order term in the Taylor series expansion of the non-linear function  $f$ . Assuming that  $f \in \mathcal{C}^1$ , the first order Taylor series expansion of  $f$  is given as follows

$$f(x) = f(c_j) + (x - c_j)^\top (\nabla f)_{x=c_j} + \varepsilon(c_j, x), \quad (13)$$

where  $\varepsilon(c_j, x)$  accounts for the higher order terms and satisfies  $\lim_{x \rightarrow c_j} \varepsilon(c_j, x) = 0$ . When  $x$  lies in a neighboring ball  $B(c_j, \delta_j) = \{x \in \mathcal{X} : \|x - c_j\| \leq \delta_j\}$  of  $c_j$  (with  $\delta_j > 0$  small), the nonlinear term  $\varepsilon(c_j, x)$  is considered negligible so that  $f$  can be approximated by the linear part of (13). Following this idea,  $\mathcal{X}$  can be decomposed into a set of a large enough number  $s$  of regions  $B(c_j, \delta_j), j = 1, \dots, s$  on which the affine approximation holds. Then, by denoting

$$\theta_j = \left[ (\nabla f)_{x=c_j}^\top f(c_j) - c_j^\top (\nabla f)_{x=c_j} \right]^\top, \quad (14)$$

the function  $\widehat{f}$  defined by

$$\widehat{f}(x) = \theta_j^\top \begin{bmatrix} x \\ 1 \end{bmatrix} \quad \text{if } x \in B(c_j, \delta_j) \quad (15)$$

approaches  $f$ . As such the function  $\widehat{f}$  is not well defined as the set of balls  $B(c_j, \delta_j), j = 1, \dots, s$ , is not necessarily a disjoint cover of  $\mathcal{X}$ . To overcome this problem we redefine  $\delta_j$  as a function of  $x$  in the form  $\delta_j(x) = \min_{i \neq j} \|x - x_i\|$  so that the euclidean ball  $B(c_j, \delta_j)$  becomes

$$\mathcal{X}_j = \{x \in \mathcal{X} : \forall i = 1, \dots, s, \|x - c_j\|_2 \leq \|x - c_i\|_2\}. \quad (16)$$

With this rearrangement, we have  $\cup_{j=1}^s \mathcal{X}_j = \mathcal{X}$  and  $\mathcal{X}_i^o \cap \mathcal{X}_j^o = \emptyset \forall i \neq j$ , where  $\mathcal{X}_i^o$  refers to the interior of  $\mathcal{X}_i$ . One can easily recognize that  $\mathcal{X}_j$  corresponds indeed to a polyhedron (which is more common in the definition of PWA models [4]) of the form

$$\mathcal{X}_j = \{x \in \mathcal{X} : H_j x \leq b_j\}, \quad (17)$$

where

$$H_j = [c_1 - c_j \ \cdots \ c_{j-1} - c_j \ c_{j+1} - c_j \ \cdots \ c_s - c_j]^\top \quad (18)$$

$$b_j = [\beta_{1,j} \ \cdots \ \beta_{j-1,j} \ \beta_{j+1,j} \ \cdots \ \beta_{s,j}]^\top,$$

with the  $\beta_{i,j}$  being defined as  $\beta_{i,j} = \frac{1}{2}(c_i^\top c_i - c_j^\top c_j)$ . Then  $\hat{f}$  is modified to be

$$\hat{f}(x) = \theta_j^\top \begin{bmatrix} x \\ 1 \end{bmatrix} \quad \text{if } x \in \mathcal{X}_j. \quad (19)$$

Thus, Eq. (19) represents a PWA approximation of the nonlinear continuous system defined in (12). Note that in the definition (19), ambiguity may still occur on the boundaries of  $\mathcal{X}_j$ . This issue can be avoided by arbitrarily assigning the points that lie on the common boundaries of any two sets  $\mathcal{X}_i$  and  $\mathcal{X}_j$ , to the one with the smallest index. For now, continuity of function  $f$  in (12) ensures that no numerical consequence is to be feared. However if we want the approximation (19) to also cover the case of discontinuous input-output map  $f$ , it will be necessary to further revise the definition of the regions  $\mathcal{X}_j$  that encodes the switching law. To see why, recall that model (19) encodes roughly the fact that if a  $x$  in the regression space is close enough to the point  $c_j$  then  $f(x)$  should be close to  $f(c_j)$  so that the affine tangent to the surface  $y = f(x)$  at  $c_j$  is considered to be a good approximation of the differentiable map  $f$ . Hence this does not hold when  $f$  is subject to discontinuities. In order to take into account this kind of situations, we may heuristically change the sets  $\mathcal{X}_j$  by incorporating in them an information about the output  $y$ . We may hence replace the sets  $\mathcal{X}_j$  in (19) with new switching sets defined as

$$\tilde{\mathcal{X}}_j = \left\{ x \in \mathcal{X} : \forall i = 1, \dots, s, \left\| \begin{bmatrix} \theta_j^\top \tilde{x} \\ x \end{bmatrix} - \begin{bmatrix} \theta_j^\top \tilde{c}_j \\ c_j \end{bmatrix} \right\|_2 \leq \left\| \begin{bmatrix} \theta_i^\top \tilde{x} \\ x \end{bmatrix} - \begin{bmatrix} \theta_i^\top \tilde{c}_i \\ c_i \end{bmatrix} \right\|_2 \right\}, \quad (20)$$

with  $\tilde{x} = [x^\top \ 1]^\top$  and  $\tilde{c}_i$  being defined likewise.

Now, we use the identification algorithm presented in Section 3 to identify the PWA model (19). This includes estimating the parameter vectors  $\theta_j$  and the points  $c_j$ ,  $j = 1, \dots, s$ , and determining afterwards the guardlines parameters  $H_j$  and  $b_j$  through Eq. (18). For any  $j$ , the point  $c_j$  is recursively estimated as the mean of the regressors vectors  $x(t)$  that are or presumed (based on the decision function) to be relevant to mode  $j$ . Together with the parameters  $\theta_j$  in (4), the operating points  $c_j$  are updated at each estimation step as

$$c_j(t|\lambda_t = j) = \gamma c_j(t-1) + (1-\gamma)x(t). \quad (21)$$

Thus, the above formulation of the PWA regression problem allows us to estimate the parameters along with the switching sets in a simple one-step and recursive manner. This feature of our method is important since the problem of estimating the switching sets is traditionally handled via *batch* support vector machines classification techniques.

## 5. APPLICATIONS

### 5.1 Numerical example

We first apply the proposed identification algorithm to a SISO SARX model composed of three linear submodels of order two. The SARX model is defined by

$$y(t) = \theta_{\lambda_t}^\top [y(t-1) \ y(t-2) \ u(t-1) \ u(t-2)]^\top + e(t) \quad (22)$$

with  $\lambda_t \in \{1, 2, 3\}$  and

$$\begin{aligned} \theta_1 &= [-0.0322 \ 0.8017 \ -1.2878 \ -1.1252]^\top, \\ \theta_2 &= [-0.1921 \ 0.5917 \ 1.1050 \ 0.0316]^\top, \\ \theta_3 &= [1.4746 \ -0.5286 \ -0.4055 \ 0.2547]^\top. \end{aligned} \quad (23)$$

In order to generate the identification data, the excitation input is chosen to be a centered signal with normal distribution and variance unity. The noise  $e(t)$  is a white Gaussian noise and chosen so that the Signal to Noise Ratio (SNR) is equal to 30 dB with respect to the output signal. The switches  $\{\lambda_t\}$  form a periodic signal that switches randomly each 5 samples. With 100 different independent realizations of the input and the noise, we generate 100 data sequences of length 1000 each. Then, starting from random initialization, identification is performed on each data sequence with a forgetting factor  $\gamma = 0.9$ . At each run of the algorithm, the first 500 points are used for identification and the estimated model is validated on the whole sequence of length 1000. We borrow from [15] the criterion

$$\text{FIT} = \left(1 - \frac{\|\hat{\mathbf{y}} - \mathbf{y}\|_2}{\|\mathbf{y} - \bar{\mathbf{y}}\|_2}\right) \times 100\% \quad (24)$$

to measure the fitting error between the true output sequence  $\mathbf{y}$  and the estimated model output sequence  $\hat{\mathbf{y}}$ . In this formula,  $\bar{\mathbf{y}}$  is the mean of the true output sequence. We represent in Figure 1-(a) the distribution (in term of histograms) of the above defined fitting error over 100 independent runs of the algorithm. It can be seen that for all these runs, the performance rate FIT is comprised between 94% and 99% with an average of 97%. This indicates that the algorithm performs very well. Convergence occurs for almost all the 100 runs. In fact the results of Figure 1-(a) are obtained by letting the algorithm pass twice over the identification data sequence. With only one pass, the results get slightly worse (see Figure 1-(b) where we obtain an average of 92% for the criterion FIT).

We also display in Figure 2 the evolution of the decision criterion (9) for mode assignment and in Figure 3 the evolution of the estimated parameter vectors for the three submodels during recursive estimation. It appears from these plots that as convergence starts occurring in the estimates, the decision criterion is driven to zero. One can notice that there is no monotonicity in the decrease of this criterion. The reason for this phenomenon is simply that the estimates do not converge at the same time for all the submodels.

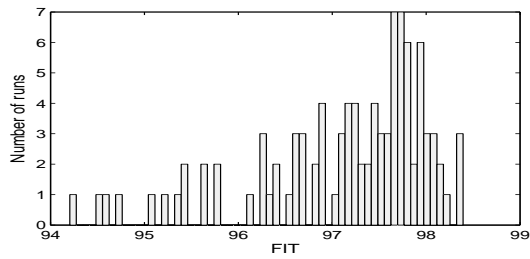
Numerical values of the estimates for one of the 100 simulations are given as

$$\begin{aligned} \hat{\theta}_1 &= [-0.0359 \ 0.8001 \ -1.2976 \ -1.1176]^\top, \\ \hat{\theta}_2 &= [-0.1974 \ 0.5866 \ 1.1106 \ 0.0182]^\top, \\ \hat{\theta}_3 &= [1.4580 \ -0.5092 \ -0.4162 \ 0.2772]^\top. \end{aligned} \quad (25)$$

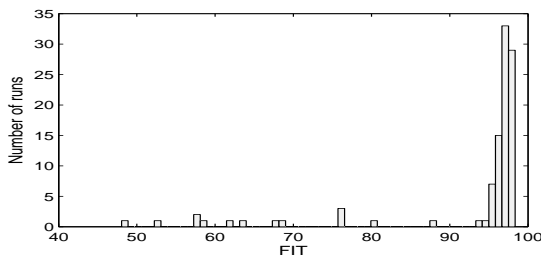
A comparison with the true parameter vectors in (23) reveals correctness of the estimates for all the three submodels. All

these observations lead us to the conclusion that the proposed algorithm is efficient, at least on the considered simulation example.

Other experiments show that similarly as in [11], good convergence properties of our recursive algorithm requires that all the submodels be visited frequently enough. Therefore, in case there is a large minimum dwell time between consecutive switches, there may arise some convergence problems. A solution may then be to start the algorithm with a single submodel and progressively increment the number of submodels when switches are detected (see [16] for details on this idea). If the algorithm is to be performed off-line, it will be also possible to randomize the order in which the data sequence is read.



(a) Two passes (off-line).



(b) One pass (on-line).

Figure 1. Distribution of the criterion FIT over 100 independent runs of the identification algorithm.

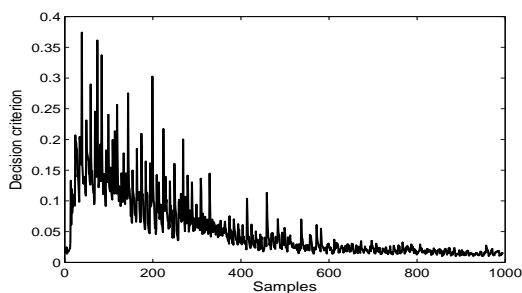


Figure 2. Average of the decision criterion (9) over 100 independent runs of the identification algorithm.

## 5.2 Open channel system

In this subsection, we consider the modeling of an open channel system using the hybrid model approach. We consider a SISO dynamical system [17] consisting of a channel with circular cross-section, radius  $R = 0.9$  m, length  $X = 946$  m and a reach slope  $\alpha = 0.26\%$  as represented in Figure 4. The discharge at any point of the channel is assumed to be bounded below by  $q_{\min} = 0.5$  m<sup>3</sup>/s and above by  $q_{\max} = 5$  m<sup>3</sup>/s.

The problem is to model the water flow between both ends of the channel. Therefore, the input and the output of the system are defined to be respectively the discharge  $u(t)$  at the upstream end and the discharge  $y(t)$  at the downstream end, all expressed in m<sup>3</sup>/s (see Figure 4).

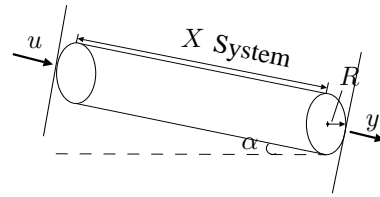


Figure 4. Geometric characteristics of the open channel system.

Because of the large dimensions of open channel systems, real data are sometimes difficult to collect. Hence, it is common practice to resort to numerical simulation tools in order to numerically reproduce in laboratory the behavior of those systems. Given the geometric characteristics (profile, dimensions, etc.) of the channel and given an exciting input sequence  $\{u(t)\}$ , we use the software SIC<sup>1</sup> to simulate the corresponding output sequence  $\{y(t)\}$ . A dataset of 3000 input-output samples is generated. Note that the considered open channel system belongs to the general class of diffusive systems that are known to be subject to time delays. Although some delays have been introduced during the data generation step, we will neglect this aspect in the modeling procedure. By applying the ideas of Section 4, we will also treat the continuous open channel system as a PWARX. We identify successively

- one submodel with orders  $n_a = n_b = 2$ 

$$\hat{\theta}_1 = [0.5544 \ 0.4238 \ 0.0260 \ -0.0043 \ -0.0000]^\top,$$

$$\hat{c}_1 = [0.8285 \ 0.8291 \ 0.8298 \ 0.7896 \ 0.7897]^\top.$$
- two submodels with orders  $n_a = n_b = 2$ 

$$\hat{\theta}_1 = [0.5541 \ 0.4241 \ 0.0260 \ -0.0043 \ 0.0000]^\top,$$

$$\hat{\theta}_2 = [0.0923 \ 0.8546 \ -0.3272 \ 0.4211 \ -0.1152]^\top,$$

$$\hat{c}_1 = [1.9152 \ 1.9033 \ 1.8942 \ 2.4380 \ 2.4319]^\top,$$

$$\hat{c}_2 = [0.8285 \ 0.8291 \ 0.8298 \ 0.7896 \ 0.7897]^\top$$

and plot the estimated output (red curve) together with the true output (blue curve) in Figure 5. The first 1500 points are used for identification and validation is made over all the available data sequence. The high value of the FIT criterion indicates a good correspondence between the model and the system on the considered data.

## 6. CONCLUSION

In this paper, we have presented a simple recursive algorithm for identifying dynamical switched linear systems from input-output measurements. Starting from some initial values (heuristically estimated or randomly drawn from some probability distribution) of the parameter vectors of the different submodels, the discrete state is sequentially and interactively estimated based on prior or posterior knowledge about the modes. At each time instant, the parameters are then updated accordingly. Note that the considered hybrid models are very general in the sense that no condition is imposed on the switching mechanism. In

<sup>1</sup> SIC (Simulation de Canaux d'Irrigation) user's guide and theoretical concepts. CEMAGREF, Montpellier, 1992. <http://canari.montpellier.cemagref.fr/>.

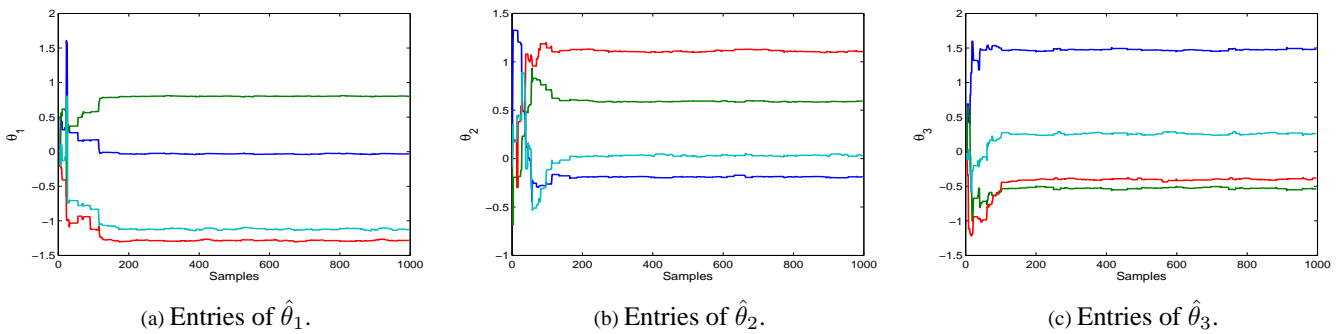


Figure 3. Evolution of the entries of the estimates during recursive estimation for two iterations over the dataset.

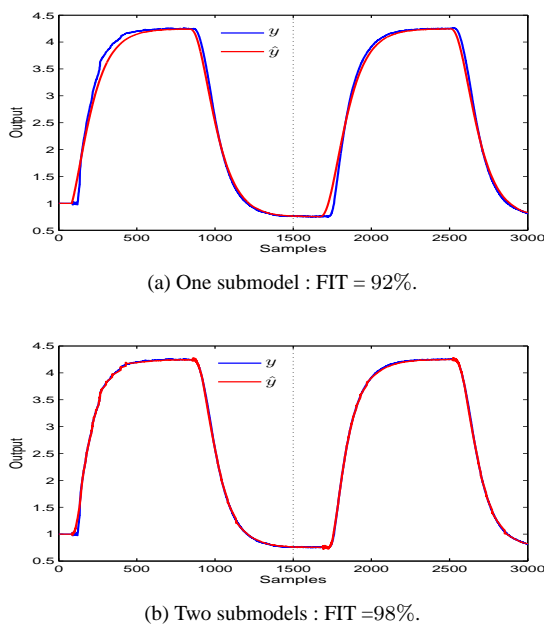


Figure 5. Approximation of the open channel system with switched affine models.

the particular case when the hybrid model intends to approach a nonlinear system, we have shown that the switching law can be inferred from data in a simple and recursive manner. Despite good results obtained on both synthetic and experimental data, lack of convergence analysis may still be criticized. However, one should note that this is in general a hard task in hybrid system identification because of the strong coupling between mode identification and parameter update. We therefore defer this challenge to future consideration.

#### REFERENCES

- [1] J. Roll, A. Bemporad, and L. Ljung, "Identification of piecewise affine systems via mixed-integer programming," *Automatica*, vol. 40, pp. 37–50, 2004.
- [2] A. Bemporad, A. Garulli, S. Paoletti, and A. Vicino, "A bounded-error approach to piecewise affine system identification," *IEEE Transactions on Automatic Control*, vol. 50, pp. 1567–1580, 2005.
- [3] R. Vidal, S. Soatto, Y. Ma, and S. Sastry, "An algebraic geometric approach to the identification of a class of linear hybrid systems," in *Conference on Decision and Control, Maui, Hawaii, USA*, 2003.
- [4] G. Ferrari-Trecate, M. Muselli, D. Liberati, and M. Morari, "A clustering technique for the identification of piecewise affine systems," *Automatica*, vol. 39, pp. 205–217, 2003.
- [5] A. L. Juloski, S. Weiland, and W. Heemels, "A bayesian approach to identification of hybrid systems," *IEEE Transactions on Automatic Control*, vol. 50, pp. 1520–1533, 2005.
- [6] S. Paoletti, A. Juloski, G. Ferrari-Trecate, and R. Vidal., "Identification of hybrid systems: A tutorial.," *European Journal of Control*, vol. 13, pp. 242–260, 2007.
- [7] F. Lauer and G. Bloch, "Switched and piecewise nonlinear hybrid system identification," in *International Conference on Hybrid Systems : Computation and Control, St-Louis, USA*, 2008.
- [8] N. Ozay, M. Sznaier, C. Lagoa, and O. Camps, "A sparsification approach to set membership identification of a class of affine hybrid systems," in *Conference on Decision and Control, Cancun, Mexico*, 2008.
- [9] K. Boukharouba, L. Bako, and S. Lecoeuche., "Identification of piecewise affine systems based on dempster-shafer theory," in *IFAC Symposium on System Identification, Saint Malo, France (To appear)*, 2009.
- [10] Y. Hashambhoy and R. Vidal, "Recursive identification of switched ARX models with unknown number of models and unknown orders," in *Conference on Decision and Control, Seville, Spain*, 2005.
- [11] R. Vidal, "Recursive identification of switched ARX systems," *Automatica*, vol. 44, pp. 2274–2287, 2008.
- [12] R. Vidal, A. Chiuso, and S. Soatto., "Observability and identifiability of jump linear systems," in *Conference on Decision and Control, Las Vegas NV, 2002*, 2002.
- [13] R. M. Johnstone, C. J. R. Johnson, R. R. Bitmead, and B. D. Anderson, "Exponential convergence of recursive least squares with exponential forgetting factor," in *Conference on Decision and Control*, 1982.
- [14] E. D. Sontag, "Nonlinear regulation: The piecewise linear approach," *IEEE Transaction on Automatic Control*, vol. 26, pp. 346–357, 1981.
- [15] L. Ljung, *System Identification Toolbox User's Guide*. 7th ed. Natick, MA: The MathWorks Inc., 2009.
- [16] L. Bako, G. Mercère, and S. Lecoeuche, "Online structured subspace identification with application to switched linear systems," *International Journal of Control*, vol. 82, pp. 1496–1515, 2009.
- [17] E. Duviella, L. Bako, and P. Charbonnaud, "Gaussian and boolean weighted models to represent variable dynamics of open channel systems," in *Conference on Decision and Control, New Orleans LA, USA*, 2007.