Applied Soft Computing 47 (2016) 271–280

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Identification of switching linear systems using self-organizing models with application to silicon prediction in hot metal

John-Eric Saxén, Henrik Saxén, Hannu T. Toivonen*

Faculty of Science and Engineering, Åbo Akademi University, FI-20500 Åbo (Turku), Finland

ARTICLE INFO

Article history:
Received 10 December 2015
Received in revised form 15 May 2016
Accepted 30 May 2016
Available online 4 June 2016

Keywords:
Hybrid systems
System identification
Self-organizing maps
Switched systems
Silicon content

ABSTRACT

An approach based on self-organizing models is presented for identification of piecewise linear switching systems, whose dynamics switch between a number of modes. The proposed method is based on a formulation of the identification problem as a generalized plane-clustering problem, which is solved using self-organizing maps. The method does not assume that the system modes depend on the state, but the mode switches may occur in an arbitrary and unknown manner. The procedure does not require knowledge of the number of modes or system orders, and it can be used for both on-line and off-line identification. Numerical examples illustrate that the procedure identifies switching systems correctly. The identification method is also applied to data from an industrial blast furnace for modeling and prediction of the silicon content of the hot metal. A switched linear model is demonstrated to capture different dynamics of the process and an analysis of the results reveals how mode switching models gradual and rapid changes in the output. The resulting models are finally shown to provide insight into factors that govern the silicon content in the blast furnace in different states.

1. Introduction

Switching systems are dynamical systems with the feature that they can switch between a number of modes with different dynamical properties [16,27]. An important class of switching systems consists of hybrid systems, whose continuous dynamics depend on discrete-valued logical variables. Systems with multiple operating regimes can also be modeled as switching systems, where the modes are associated with the various operating conditions. In this case the mode is usually known or is a function of known variables. In more general cases, the mode switchings may, however, be random, or they may depend on variables which are unknown.

Black-box identification of switching systems using input–output data is an important problem which has been addressed in several studies [5,10,30,22,18,12,28,31,2,6]. A special problem in switching system identification is the fact that the times of the mode switches may not be known. In these cases the switching times between the various modes should be identified simultaneously with the individual models, which makes the identification of switching systems significantly more demanding than standard system identification. Therefore, many studies have addressed the simpler problem where it is assumed that the modes depend on some measured variables.

A special class of switching systems consists of systems where the modes depend on the state. Identification of such systems can be roughly decomposed into two parts. First, clustering techniques are used to determine regions where the various system modes are active, followed by identification of the individual models using standard techniques. In [21], a two-stage approach was applied to the identification of nonlinear systems, using self-organizing maps for input data clustering, and modeling the system using locally valid linear models. In [18] statistical clustering of the input data was applied to identification of piecewise affine systems. The fact that the clustering and model identification steps interact and cannot be solved independently of each other has been addressed in [5,10], where piecewise affine models were identified using combinations of clustering, classification and linear identification methods. In [12] a Bayesian approach was applied to identification of hybrid systems, where the model parameters are described as random variables. In [6] an identification method for a class of linear switched systems was developed based on a discrete particle swarm optimization formulation. The authors demonstrated the feasibility of the approach on two simulated examples, one being a switching ARX model and the other a switched mode power supply system.

The approaches which rely on data clustering to determine the regions where individual modes are active are not suitable for

* Corresponding author. Tel.: +358 2 2154676; fax: +358 2 2515557.
E-mail addresses: josaxen@abo.fi (J.-E. Saxén), hsaxen@abo.fi (H. Saxén), htoivone@abo.fi (H.T. Toivonen).

http://dx.doi.org/10.1016/j.asoc.2016.05.048
1568-4946/

systems where the switching dynamics do not depend on the state, but is random or depend on some unknown variable. Identification of switching systems with arbitrary switching dynamics can be formulated as a convex linear or quadratic mixed integer programming problem [22]. The practical applicability of this approach is, however, restricted by the high complexity of the mixed integer optimization problems involved. Therefore, various alterative identification methods have been proposed.

In [30], identification of switched ARX systems was addressed by minimizing the sum of the product of the squared prediction errors of individual models. By construction, in the noise-free case this objective function is minimized by the correct models associated with the various modes. A nice feature of this approach is that the identification can be performed recursively [11,31]. The procedure was further modified in [8] for robustness to outliers in the data. The method involves, however, the solution of a computationally demanding nonconvex optimization problem, for which convergence to the globally optimal solution can in general not be guaranteed. Other approaches include subspace identification to identify switching systems with a minimum dwell time in each mode [1], solution by a modified k-plane clustering algorithm [7], and a sparse optimization approach inspired by the field of compressive sensing [2].

In [31], a recursive identification method for switched ARX systems was presented. The switching system was characterized in terms of a single lifted model in a higher-dimensional space. The lifted model, which is derived using the product of the prediction errors of the individual models, has a linear in parameters representation in terms of hybrid model parameters, which have a one-to-one correspondence with the ARX model parameters. A recursive identification method can therefore be obtained by applying standard recursive identification to identify the hybrid model parameters, from which the ARX model parameters can be reconstructed.

Self-organizing maps have been introduced to produce low-dimensional representations of high-dimensional data using unsupervised learning [14]. The technique has been applied to a number of problems, including clustering [13], processing of temporal data and time-series [4], and identification and control of dynamical systems [3]. For signal segmentation and classification, [15] introduced an approach based on a self-organizing map of linear models. It was assumed that the signal segments can be characterized by autoregressive (AR) models, and a self-organizing map [14] was applied to detect the models which describe the signal in the various segments. The method does not assume that the different modes depend on the state, but the modes are found by a combination of Kohonen's self-organizing map and adaptive least-squares to estimate the AR model parameters.

In this paper we study identification of switching systems with arbitrary switching dynamics. The proposed approach is based on the observation that while the system modes cannot be found by data clustering, which is the case when the modes depend on the state, the identification problem can instead be formulated as a kind of modified plane-clustering problem [7]. This fact has been exploited previously in [28] to solve the identification problem by k-plane clustering. Here, the resulting plane-clustering problem will be solved using self-organizing models, and in this respect the approach studied here is closely related to the method introduced in [15]. The proposed method does not require that the number of modes be known, and it can also be formulated for the case with unknown model orders. The technique is easy to implement and it can be applied in both batch mode and in an on-line manner.

In the present study, the proposed method will also be applied to predicting the silicon content of the hot metal produced in a blast furnace. This problem has been tackled by black-box modeling in the literature (e.g., [20,32,25]) but only with moderate success due to the complex nature of the problem. The results of the switching system modeling and identification provide insight into the process dynamics and also a means by which more light can be shed on the variables that govern the silicon content in hot metal in different states of the blast furnace.

The paper is organized as follows. In Section 2, the identification problem is formulated, and it is shown that it can be stated as a generalized plane clustering problem. Section 3 contains the main contribution of the paper, consisting of an identification method using self-organizing models. Finally, in Section 4 the performance of the procedure is studied using both simulated examples and industrial data.

2. Problem formulation

We consider a dynamical system whose behavior switches between a number of modes. It is assumed that the system behavior is described by the linear time-dependent discrete-time model

$$y(k) = \varphi(k)^T\theta(k) + v(k) \tag{1}$$

where $y(k)$ is the output, $\varphi(k) \in R^n$ is a suitably defined state vector, $v(k)$ is a zero-mean disturbance with variance $Ev(k)^2 = \sigma(k)^2$, and $\theta(k) \in R^n$ is a parameter vector. It will be assumed that the system dynamics switch between a number of modes, so that the system parameters $\theta(k)$ at each time instant k belong to a finite set corresponding to the various modes, i.e.,

$$(\theta(k), \sigma(k)) \in \{(\theta^{(i)}, \sigma^{(i)}), i \in \{1, 2, \ldots, N_{\text{mode}}\}\} \tag{2}$$

where N_{mode} is the number of modes.

An important special case of switching system models of the form (1) consists of switched ARX models [31],

$$\begin{aligned} y(k) - a_1(k)y(k-1) - \cdots - a_{n_a}(k)y(k-n_a) \\ = b_1(k)u(k-1) + \cdots + b_{n_b}(k)u(k-n_b) + v(k) \end{aligned} \tag{3}$$

where $u(k)$ is the input, and for which the state vector is given by

$$\varphi(k) = [\, y(k-1) \quad \cdots \quad y(k-n_a) \quad u(k-1) \quad \cdots \quad u(k-n_b) \,]^T \tag{4}$$

and the parameter vector is defined as

$$\theta(k) = [\, a_1(k) \quad \cdots \quad a_{n_a}(k) \quad b_1(k) \quad \cdots \quad b_{n_b}(k) \,]^T \tag{5}$$

The system identification problem studied in this paper can be stated as follows. Assume that a number of input–output data pairs $(\varphi(k), y(k))$, $k = 1, \ldots, N$ from the switching system have been observed. The problem is then to identify the number of different modes N_{mode} needed to model the system behavior during the identification experiment, the parameters $\theta^{(i)}$ of the individual modes, and the time regions during which the system has been in the various modes. In a least-squares setting, the problem can be formulated quantitatively as finding the parameter vectors $\hat{\theta}^{(i)}$ which minimize the sum of squares of the smallest prediction errors, or

$$\min_{\{\hat{\theta}^{(i)}\}} \sum_{k=1}^{N} \min_i (y(k) - \varphi(k)^T\hat{\theta}^{(i)})^2 \tag{6}$$

It is assumed that the input–output sequence is the only information available, and it is thus not known when the system behavior switches from one mode to another, nor is it required that the number N_{mode} of modes is known. However, as the minimum value of the sum of squared prediction errors in (6) is a monotonically decreasing function of the number of modes, the proper number of modes should in practice be determined by statistical methods.

Remark 2.1. The optimization problem (6) can be compared to methods based on clustering of the state vectors. A subset of the linear switching system (1) consists of piecewise affine systems,

for which it is assumed that the system mode is a function of the state [5,10],

$$\theta(k) = \theta^{(i)}, \quad \text{if} \quad \varphi(k) \in \mathcal{X}_i$$

where $\{\mathcal{X}_i\}_{i=1}^{s}$ is a partition (usually assumed polyhedral) of R^n. Piecewise affine systems described by (1) where the system modes are defined by a partition of the state can be identified by applying clustering techniques to classify the data pairs $(y(k), \varphi(k))$ in order to determine the subsets $\mathcal{X}_i$ and the time instants associated with the various modes, followed by identification of the submodels [18]. Various methods can be applied to solve the clustering problem. In [18] a statistical clustering method based on a Gaussian mixture model was used. In a least-squares setting, the classification stage consists of the minimization problem

$$\min_{\{\mu^{(i)}\}} \sum_{k=1}^{N} \min_i \|\mu^{(i)} - z(k)\|^2 \tag{7}$$

where $\mu^{(i)}$ are the cluster centers, $z(k) = [\, y(k) \quad \varphi(k)^T \,]^T$, and $\|\cdot\|$ is the Euclidian vector norm. For future reference, we may observe that self-organizing maps [14] can be applied to find the cluster centers by adapting a set of vectors $\hat{\mu}^{(i)}$ iteratively in the direction of the negative gradients of the quadratic terms in (7), according to

$$\begin{aligned} \hat{\mu}^{(i)}(k+1) &= \hat{\mu}^{(i)}(k) - \frac{1}{2}\gamma_k \mathrm{d}\|\hat{\mu}^{(i)}(k) - z(k)\|^2 / \mathrm{d}\hat{\mu}^{(i)}(k) \\ &= \hat{\mu}^{(i)}(k) + \gamma_k(z(k) - \hat{\mu}^{(i)}(k)) \end{aligned} \tag{8}$$

where γ_k is a nonnegative steplength parameter.

Remark 2.2. Observe that the optimization problems (6) and (7) have a similar general structure. Indeed, the system identification problem defined by (6) can be considered as a generalized clustering problem. More precisely, it is closely related to plane clustering [7], where the cluster plane associated with the vector $\tilde{\theta}^{(i)}$ is defined as

$$\mathcal{P}(\tilde{\theta}^{(i)}) = \{x | x^T \tilde{\theta}^{(i)} = 0\} \tag{9}$$

The plane clustering problem consists of finding the planes $\mathcal{P}(\tilde{\theta}^{(i)})$ described by (9) and the associated vectors $\tilde{\theta}^{(i)}$ such that the sum of the squares of each data point $x(k)$ to the nearest plane is minimized. This corresponds to the minimization problem [7]

$$\min_{\{\tilde{\theta}^{(i)}\}} \sum_{k=1}^{N} \min_i (\mathrm{dist}(x(k), \mathcal{P}_i))^2 \tag{10}$$

where the 2-norm distance between the point $x(k)$ and the plane $\mathcal{P}_i$ is [7]

$$\mathrm{dist}(x(k), \mathcal{P}_i) = |x(k)^T \tilde{\theta}^{(i)}| / \|\tilde{\theta}^{(i)}\|$$

This form of plane clustering problem has been introduced in [7], where a k-plane clustering algorithm was presented as a generalization to the standard k-means method. By making the identifications $x(k) = [\, y(k) \quad -\varphi(k)^T \,]^T$ and $\tilde{\theta}^{(i)} = [\, 1 \quad (\hat{\theta}^{(i)})^T \,]^T$ the identification problem (6) takes the form

$$\min_{\{\tilde{\theta}^{(i)}\}} \sum_{k=1}^{N} \min_i |x(k)^T \tilde{\theta}^{(i)}|^2 \tag{11}$$

which is essentially equivalent to the plane clustering problem defined by (10), apart for the normalizing factor $\|\tilde{\theta}^{(i)}\|$.

3. Identification method

In this section we present a solution to the switching system identification problem stated in Section 2. The identification problem defined by (6) will be solved using a method which can be considered as a modification of Kohonen's self-organizing maps [14]. The procedure consists of applying the iterative algorithm (8) to minimize (6) instead of (7), but is otherwise similar as the standard procedure. Using the fact that the derivative of the squared prediction errors in (6) is

$$\frac{\mathrm{d}}{\mathrm{d}\theta}(y(k) - \varphi(k)^T\theta)^2 = -2\varphi(k)(y(k) - \varphi(k)^T\theta) \tag{12}$$

we obtain the following self-organizing model algorithm.

Off-line identification algorithm:

1. *Initialization.* Select the model structure and the maximum number of modes, M. Set iteration number $p = 0$ and choose a set of initial parameter vectors $\hat{\theta}_j^p, j = 1, 2, \ldots, M$.
2. *Initialize iteration.* Set $p \leftarrow p+1$, time index $k = 0$ and $\hat{\theta}_j^p(k) = \hat{\theta}_j^p, j = 1, 2, \ldots, M$.
3. *Matching.* Set $k \leftarrow k+1$. For input–output data pair $\varphi(k)$, $y(k)$, compute the squared model prediction errors

$$q_j(k) = (y(k) - \varphi(k)^T \hat{\theta}_j^p(k))^2, \quad j = 1, 2, \ldots, M \tag{13}$$

and let j^* be the index of the best-matching parameter vector $\hat{\theta}_{j^*}^p(k)$ which solves

$$\min_j q_j(k), \quad j = 1, 2, \ldots, M$$

4. *Updating.* Adjust the parameters according to

$$\hat{\theta}_j^p(k+1) = \hat{\theta}_j^p(k) + \eta^p \Lambda_{j,j^*}^p \varphi(k)(y(k) - \varphi(k)^T \hat{\theta}_j^p(k)) \tag{14}$$

where η^p is a learning parameter and Λ_{j,j^*}^p is a neighborhood function.
5. *Continuation?* If $k < N$, continue from step 3; otherwise go to step 6.
6. *Test for convergence.* If parameters have converged, go to step 7. Otherwise, set $\hat{\theta}_j^p = \hat{\theta}_j^p(N), j = 1, 2, \ldots, M$, and continue from step 2.
7. *Determine individual modes.* Determine the efficient number of modes and associated model parameters by applying clustering to the parameter vectors $\hat{\theta}_j^p$.

The procedure can be applied to on-line identification by repeating the matching and updating steps 3 and 4 recursively,

$$\hat{\theta}_j(k+1) = \hat{\theta}_j(k) + \eta(k)\Lambda_{j,j^*}(k)\varphi(k)(y(k) - \varphi(k)^T\hat{\theta}_j(k)) \tag{15}$$

The neighborhood function $\Lambda_{j,j^*}(k)$ is a key design variable in determining algorithm performance and convergence of the parameters. Its purpose is to adjust those parameter vectors which give the best match for the data pair $(\varphi(k), y(k))$ in the descent direction of the cost $q_j(k)$, whereas the rest of the parameter vectors should not be influenced. In order to achieve good convergence properties, it should be a non-increasing function of the squared prediction error $q_j(k)$. The simplest form of neighborhood function is a step function which includes all parameters that give a prediction error less than a defined maximum prediction error, i.e.,

$$\Lambda_{j,j^*} = \begin{cases} 1, & \text{if} \quad q_j \le c q_{j^*} \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

where $c \geq 1$. This approach, however, ignores the effect of distances within the neighborhood. A typical neighborhood function which takes the distances into account is a Gaussian function [21]

$$\Lambda_{j,j^*} = \begin{cases} \exp(-0.5q_j/d^2), & \text{if } q_j \leq cq_{j^*} \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

where d is the neighborhood width parameter. The values of design parameters c and d in (16) and (17) reflect the magnitudes of the prediction errors considered acceptable for a model to describe a given system mode.

3.1. Convergence analysis

The proposed identification method has convergence properties similar to those of the standard self-organizing map [14,3]. The convergence can be studied by considering the average of the squared prediction error in (6). Assume that the system is stationary and the various modes have probabilities

$$p_i = E[(\theta(k), \sigma(k)) = (\theta^{(i)}, \sigma^{(i)})], \quad i = 1, \ldots, N_{\text{mode}}$$

For a set of parameter estimates $\{\hat{\theta}_j\}$, the prediction errors and the neighborhood functions Λ_{j,j^*} then have well defined probability distributions, and we can introduce the average cost

$$\begin{aligned} J(\{\hat{\theta}_j\}) &= E\left[\sum_{j=1}^{M} \Lambda_{j,j^*} \|e_j(k)\|_2^2\right] \\ &= \sum_{i=1}^{N_{\text{mode}}} E\left[\sum_{j=1}^{M} \Lambda_{j,j^*} \|e_j(k)\|_2^2 \;\middle|\; (\theta(k), \sigma(k)) = (\theta^{(i)}, \sigma^{(i)})\right] p_i \end{aligned} \tag{18}$$

where

$$e_j(k) = y(k) - \varphi^{\mathrm{T}}(k)\hat{\theta}_j, \quad j = 1, \ldots, M$$

The stochastic approximation method to minimize $J(\cdot)$ when the probabilities p_i are unknown consists of considering samples of $J(\cdot)$,

$$J_k(\{\hat{\theta}_j(k)\}) = \sum_{j=1}^{M} \Lambda_{j,j^*} \|y(k) - \varphi^{\mathrm{T}}(k)\hat{\theta}_j(k)\|_2^2 \tag{19}$$

and adjusting the parameters in the direction of the negative gradient of $J_k(\cdot)$,

$$\hat{\theta}_j(k+1) = \hat{\theta}_j(k) - \frac{1}{2}\eta(k)\frac{\mathrm{d}J_k(\{\hat{\theta}_j\})}{\mathrm{d}\hat{\theta}_j} \tag{20}$$

This is identical with the updating formula (14).

The stochastic approximation algorithm converges to a local minimum of $J(\{\hat{\theta}_j\})$ if

$$\sum_{k=1}^{\infty} \eta(k) = \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \eta^2(k) < \infty$$

i.e., the learning parameter $\eta(k)$ should decrease as the models become more accurate, but sufficiently slowly to allow convergence. A common choice is [21]

$$\eta(k) = \frac{1}{\alpha + \beta k} \tag{21}$$

where $\alpha, \beta \geq 0$, or in batch mode,

$$\eta^p(k) = \frac{1}{\alpha + \beta p}$$

For proper choices of the neighborhood function Λ_{j,j^*}, such as (16) with a sufficiently small radius, the local minima are achieved by the true mode parameters $\theta^{(i)}$. Convergence to the true parameters as $N \to \infty$ requires in addition that the model structures are correct, and that each mode satisfies a persistent excitation condition, i.e.,

$$\lim_{N\to\infty} \frac{1}{N_i} \sum_{k_i} \varphi(k_i)\varphi^{\mathrm{T}}(k_i) > 0 \quad \text{and} \quad \lim_{N\to\infty} \frac{N_i}{N} > 0$$

where k_i and N_i denote the time instants and the total number of time instants, respectively, that the system is in mode i.

The convergence points of the individual parameter vectors depend on their initial values. In order to ensure that all system modes are found, the initial parameter vectors should therefore cover the parameter space sufficiently well.

3.2. Unknown model orders

The identification method presented above assumes that the model orders are known. When this is not case, the numbers n_a, n_b of parameters of the various models should be identified as well. One approach is to apply sparse modeling techniques, in which the number of model parameters are first overestimated, and pruned in the identification process. This can be achieved by replacing the cost function (19) by

$$V_k(\{\hat{\theta}_j(k)\}; \{W_j(k)\}) = J_k(\{\hat{\theta}_j(k)\}) + \sum_{j=1}^{M} \Lambda_{j,j^*} \hat{\theta}_j^{\mathrm{T}}(k) W_j(k) \hat{\theta}_j(k) \tag{22}$$

where $W_j(k)$ are diagonal weighting matrices

$$W_j(k) = \mathrm{diag}(w_{j,1}(k), \ldots, w_{j,n_j})$$

where the elements are non-negative, and n_j is the dimension of $\hat{\theta}_j$. An update algorithm for minimization of $V(\cdot;\cdot)$ can be constructed in analogy with (20).

The purpose of the parameter weighting in the cost function is to make the estimates of redundant parameters of an overparameterized model equal to zero. For this purpose the parameter weights should be properly chosen. This is, however, in general not possible a priori as the parameters whose actual values are zero are not known. A widely used method for pruning model parameters is to apply iteratively reweighted least squares [9], where the parameter weights are given as functions of the previous estimates,

$$w_{j,m}(k) = \frac{\lambda_{j,m}}{\epsilon + |\hat{\theta}_{j,m}(k-1)|} \tag{23}$$

where $\epsilon > 0$ is a small parameter. The purpose of the reweighting scheme is to adjust the weights in such a way that the estimates of any redundant parameters eventually become zero. To secure that the correct parameters are pruned, the weights $\lambda_{j,m} \geq 0$ can be selected to put a larger weight on parameters a_{m_a}, b_{m_b} which are associated with higher-order models.

4. Examples

In this section numerical examples are presented to illustrate the behavior of the proposed identification method. In Section 4.1, on-line identification of a simulated system is considered, while switching models of an industrial process are identified in Section 4.2.

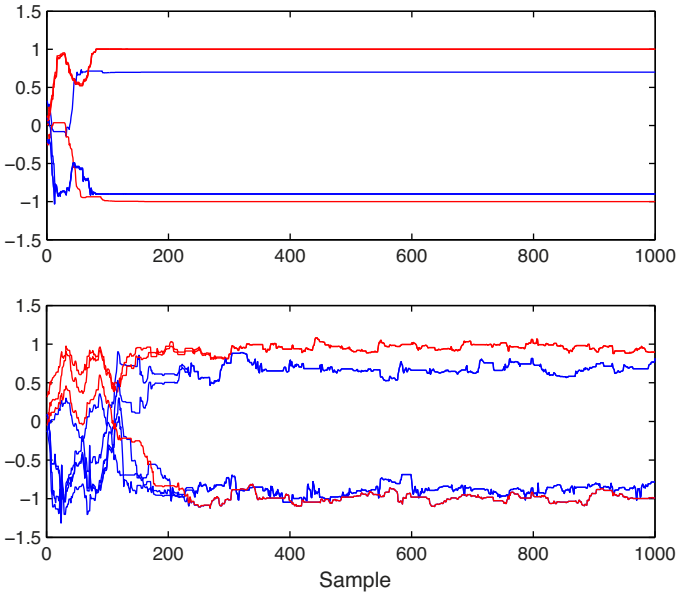

Fig. 1. Estimated parameters $\hat{a}^{(j)}(k)$ (blue) and $\hat{b}^{(j)}(k)$ (red) in Example 1. Noise-free and noisy case ($\sigma = 0.3$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

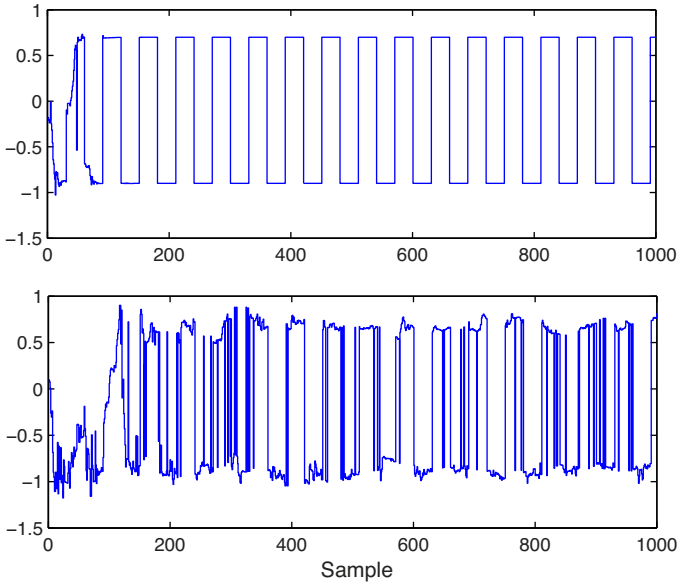

Fig. 2. Estimated parameter $\hat{a}(k)$ of winning modes in Example 1. Noise-free and noisy case ($\sigma = 0.3$).

4.1. Simulated examples

In the following examples, we consider recursive identification of the simulated first-order switched ARX system [11,31]

$$y(k) = a(k)y(k-1) + b(k)u(k-1) + v(k) \tag{24}$$

The system has two modes, in which the parameters take the values $a^{(1)} = -0.9$, $b^{(1)} = 1$ and $a^{(2)} = 0.7$, $b^{(2)} = -1$, respectively.

The system is simulated with varying switching times between the two modes. The input $u(k)$ was taken to be normally distributed white noise with zero mean and unit variance.

The on-line version of the identification method presented in Section 3 was applied to recursively identify the switching system (24). The neighborhood function was taken as Gaussian (17) with a bounded c-value and the learning parameter $\eta(k)$ as a decreasing function of time according to (21). The initial parameter vector values were selected at random.

In the first example, periodic mode switching is studied, while in the second example the mode switches are random. In the third example the approach in Section 3.2 is applied to the case when the model orders are unknown.

The simulation results are comparable to the ones presented in [31] where the same switching system has been studied.

Example 1. In this example, the system is simulated with periodic switching of period 30 samples. The number of models was overestimated as $M = 4$, while the correct number of models is $M = 2$. The learning parameters in (21) were taken as $\alpha = 9$, $\beta = 0.01$. The experiment is repeated for both noise-free and noisy cases. In the latter case, the disturbance $v(k)$ is normally distributed white noise with standard deviation $\sigma = 0.3$. Fig. 1 shows the evolution of the model parameters for the noise-free and noisy cases, while Fig. 2 shows the estimated parameter $\hat{a}(k)$ of the winning mode.

It is seen that the parameters of the individual models converge to the true values, models 1 and 4 converging to mode 1 and models 2 and 3 to mode 2. The parameters of the winning mode, taken as the parameter vector which minimizes the quadratic cost (13), are seen to be correctly identified after approximately 100 samples (Fig. 2) in the noise-free case. In the noisy case, the identified parameters are close to the correct values after approximately 200 samples, but there are some misclassified modes. This cannot be completely avoided, because in the presence of noise even using the correct system models may occasionally result in misclassification.

Table 1
Parameter estimates in Example 2. Mean values and root mean square errors based on 100 simulations with different noise standard deviations σ.

	$\sigma = 0$	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.3$
Mode 1				
$\hat{a}$	-0.900 ± 0.000	-0.899 ± 0.018	-0.897 ± 0.035	-0.893 ± 0.049
$\hat{b}$	1.000 ± 0.000	0.998 ± 0.015	1.002 ± 0.036	1.000 ± 0.052
Mode 2				
$\hat{a}$	0.700 ± 0.000	0.701 ± 0.018	0.694 ± 0.035	0.699 ± 0.054
$\hat{b}$	-1.000 ± 0.000	-1.000 ± 0.015	-1.009 ± 0.034	-0.999 ± 0.051

Example 2. In this example, the system and model order are as above, but the switching is modeled as a Markov process. This case demonstrates both the identification of a randomly switching system and how the identification method adapts to a high switching frequency. The number of models was overestimated as $M = 4$ and the same learning parameters were used as in Example 1. The Markov process was defined as a first-order Markov chain with the transition matrix

$$P = \begin{pmatrix} 0.520 & 0.480 \\ 0.576 & 0.424 \end{pmatrix}$$

where $P_{ij} = E[\theta(k+1) = \theta^{(j)} | \theta(k) = \theta^{(i)}]$. Fig. 3 shows the evolution of the model parameters for the noise-free and noisy cases. It is seen that the convergence rate is similar to that in Example 1. Table 1 shows the mean parameter estimates and their root mean square errors from 100 simulations for various noise variances.

Example 3. Finally, we consider the case when the system order is unknown. The model is therefore overparameterized using the second-order ARX model

$$y(k) = a_1(k)y(k-1) + a_2(k)y(k-2) + b_1(k)u(k-1) + b_2(k)u(k-2) + v(k) \tag{25}$$

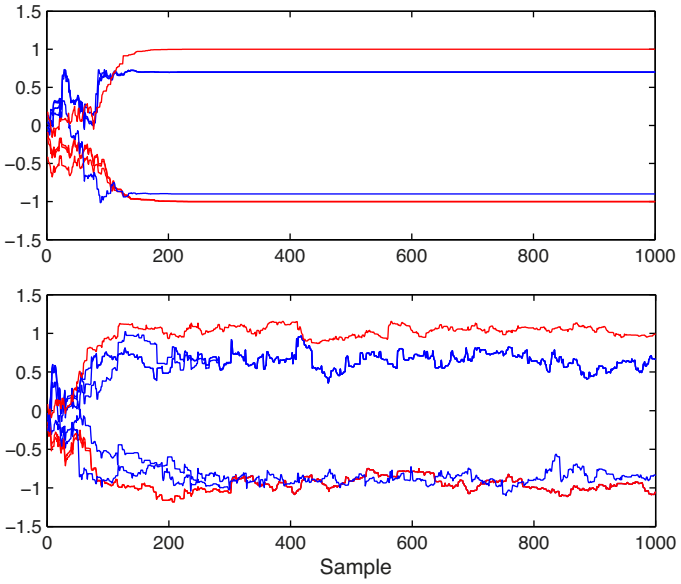

Fig. 3. Estimated parameters $\hat{a}^{(j)}(k)$ (blue) and $\hat{b}^{(j)}(k)$ (red) in Example 2. Noise-free and noisy case ($\sigma = 0.3$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Iterative reweighted least squares based on (22) was applied to prune model parameters. In the reweighting scheme (23) the parameters were taken as $\lambda = 0$, $\epsilon = 0.2$ for $a_1(k)$, $b_1(k)$ and $\lambda = 0.5$, $\epsilon = 0.2$ for $a_2(k)$, $b_2(k)$. In analogy with Example 1, the system is simulated with a constant switching period of 30 samples and the number of models is now assumed correctly as $M = 2$. The learning parameters in (21) were taken as $\alpha = 12$, $\beta = 0.01$.

Fig. 4 shows the evolution of the model parameters for the noise-free and noisy cases. Although convergence is slower than in Example 1, the model parameters converge to the correct ones, and the model structures are correctly identified. Table 2 shows the mean parameter estimates and their root mean square errors from 100 simulations for various noise variances.

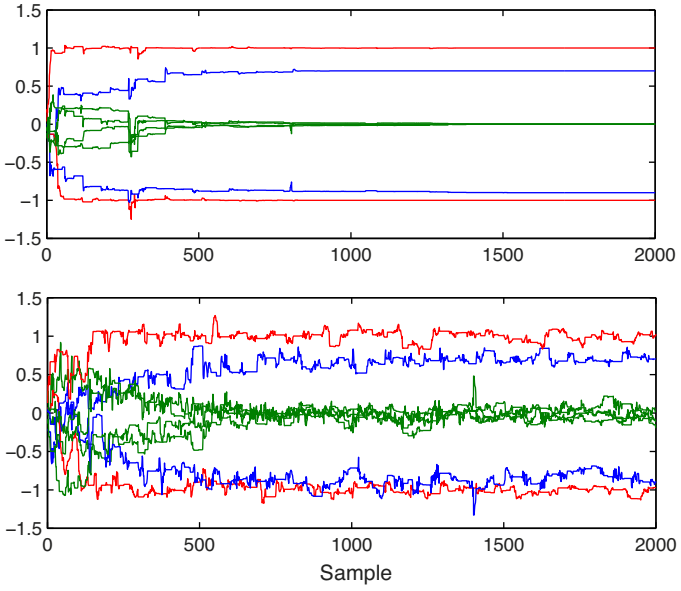

Fig. 4. Estimated parameters $\hat{a}_1^{(j)}(k)$ (blue), $\hat{b}_1^{(j)}(k)$ (red) and $\hat{a}_2^{(j)}(k)$, $\hat{b}_2^{(j)}(k)$ (green) in Example 3. Noise-free and noisy case ($\sigma = 0.3$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Parameter estimates in Example 3. Mean values and root mean square errors based on 100 simulations with different noise standard deviations σ.

	$\sigma = 0$	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.3$
Mode 1				
$\hat{a}_1$	−0.897 ± 0.022	−0.823 ± 0.051	−0.873 ± 0.044	−0.840 ± 0.062
$\hat{a}_2$	0.0026 ± 0.021	0.0113 ± 0.042	0.0234 ± 0.045	0.0328 ± 0.065
$\hat{b}_1$	1.000 ± 0.003	1.002 ± 0.022	0.997 ± 0.037	1.005 ± 0.056
$\hat{b}_2$	−0.0032 ± 0.021	−0.0129 ± 0.046	−0.0303 ± 0.045	−0.0641 ± 0.068
Mode 2				
$\hat{a}_1$	0.690 ± 0.070	0.676 ± 0.108	0.674 ± 0.087	0.677 ± 0.055
$\hat{a}_2$	0.0073 ± 0.050	0.0109 ± 0.079	0.0160 ± 0.080	0.0009 ± 0.060
$\hat{b}_1$	−1.000 ± 0.001	−1.000 ± 0.016	−1.003 ± 0.032	−1.005 ± 0.059
$\hat{b}_2$	−0.0103 ± 0.072	−0.0229 ± 0.105	−0.0317 ± 0.095	−0.0242 ± 0.060

4.2. Identification of switching system model of a metallurgical process

Finally, application of the proposed procedure will be demonstrated on data from an industrial metallurgical process. In the ironmaking blast furnace, which produces molten iron (hot metal) for further refining into steel, the silicon (Si) content of the hot metal is an important process variable that reflects the product quality and also the thermal status of the process [25]. Predicting the silicon content has been used as a benchmark problem in many modeling efforts (e.g., [20,32]).

The process dynamics may vary due to a number of factors, such as changes in raw materials, different manual control strategies used by the operators, gas and burden distribution, etc. Furthermore, the level around which the silicon content varies changes with the internal conditions (e.g., thermal level of the high-temperature region). Obviously, these aspects should be taken into account in the process modeling, so it seems natural to describe the process dynamics with multiple models, as suggested in a recent review paper on the topic [25]. To the best of our knowledge, switching models for predicting the silicon content have only been used once earlier [19].

Inspired by the findings in [24], hourly mean values of three process variables in an industrial blast furnace, namely a gas resistance index, the specific blast volume and the heat losses from the tuyeres (air supply nozzles) are taken as inputs for a model of the hot metal silicon content. This is in agreement with process knowledge, as these variables mainly reflect the conditions in the lower part of the furnace, where the hot metal assumes its final state (composition and temperature). The data used for identification consist of a set of 797 consecutive data points with one output y, the silicon content, and the three input variables mentioned above, x_1, x_2 and x_3. These variables were normalized between 0 and 1 and are illustrated in Fig. 5. Previous studies [24] have demonstrated that the process could be described by a model where the output $y(k)$ is taken as a function of recent values of the inputs. To tackle the problem of unknown time lags the three variables were included at time lags of one, two and three hours, i.e., $x_i(k-l)$, $i = 1, 2, 3$, $l = 1, 2, 3$. In order to assess whether the changing process dynamics can be captured by a switching system, the switching linear model

$$\hat{y}(k) = \sum_{i=1}^{3} \sum_{l=1}^{L_i} \hat{\theta}_{il}(k) x_i(k-l) \tag{26}$$

was used, where $\hat{y}(k)$ is the predicted output (silicon content).

In this example, the on-line version of the procedure described in Section 3 was applied on the data shown in Fig. 5, with $M = 3$ models. After training the switching model on the 697 first data points, the performance on the remaining 100 data points was studied. As the model here was used for one-step ahead prediction rather than estimation, the value of the output at time k is unknown at the time

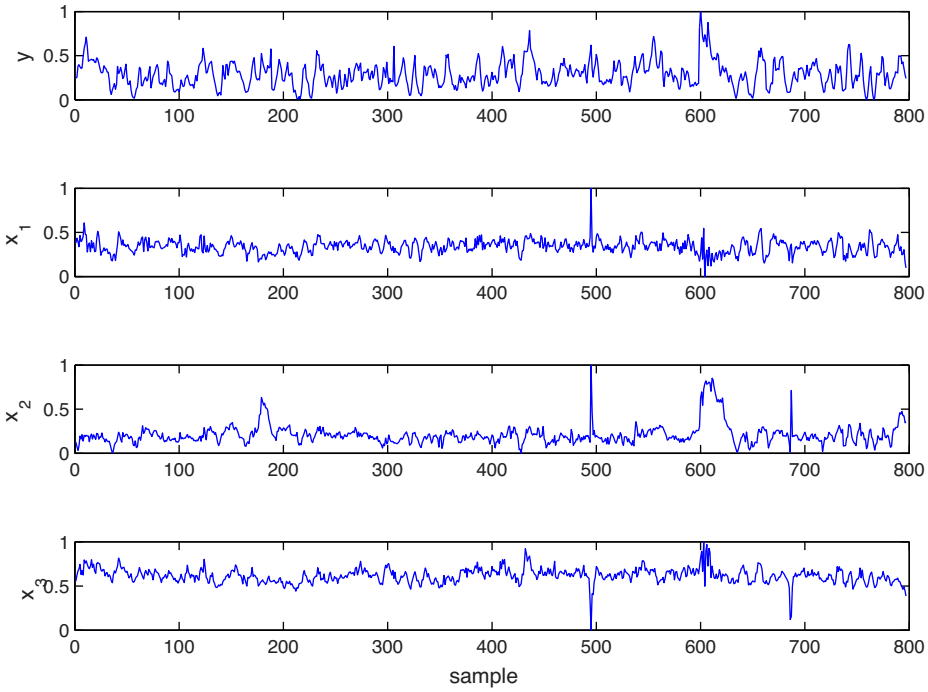

Fig. 5. Output $y(k)$ (silicon content) and input variables $x_1(k)$ (gas resistance index), $x_2(k)$ (specific blast volume) and $x_3(k)$ (tuyere heat loss) in blast furnace example.

of making the prediction, $\hat{y}(k)$. Therefore, in carrying out the prediction the mode was chosen as the one selected by the method in the updating of the parameters at time $k-1$. The neighborhood function was chosen Gaussian (17) and the initial value of the learning parameter was chosen as $\eta(0) = 0.3$.

The top panel of Fig. 6 shows the (scaled) silicon content (solid line) and the model predictions (dashed line) for the test set. It is seen that the model can follow the main trends of the signal, but experiences problems in predicting abrupt changes and often lags behind. This problem is common for all kinds of models of the silicon content in the blast furnace reported in the literature [25]. However, at the end of the test set the model predicts the behavior quite accurately without delay. For the purpose of illustration, the bottom panel of Fig. 6 illustrates the simultaneous evolution of one model parameter, ($\hat{\theta}_{11}$), which expresses the dependence between the resistance index at a time lag of one hour and the silicon content. From this panel the three modes of operation (seen as distinct levels in the parameter, and indicated by numerals 1–3) a can be seen. It may be observed that mode 3 occurs less frequently than modes 1 and 2.

Fig. 7 illustrates by solid lines the predictions of the three modes on the test set. The open circles connected by the dashed lines depict the one-step ahead prediction by the mode applied by the method. The figure gives an idea of how switching occurs (on this particular data set). It may be noted that the three modes provide very similar output patterns, but the levels of the outputs differ: This could reflect medium-term changes in the heat level of the process [17] encountered in the training set. It is, furthermore,

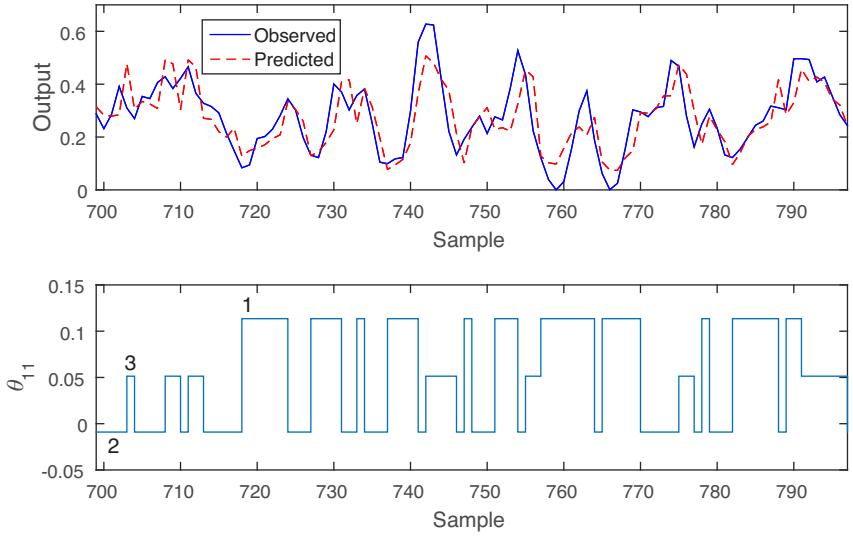

Fig. 6. Top panel: True (solid line) silicon content for the last 100 observations of the data set illustrated in Fig. 5, and one-step ahead prediction (dashed line) of a three-mode switching model. Bottom panel: Simultaneous evolution of one parameter of the switching model, showing mode transitions as abrupt changes in the parameter value. The numerals at the line denotes the mode numbers.

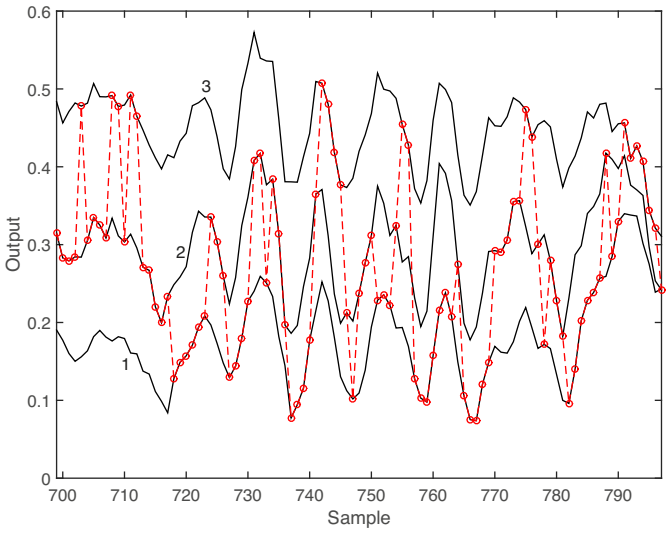

Fig. 7. Silicon content predicted by the three modes of the model (numbered by small numerals at the lines) as well as the prediction by the best-model selected by the method (open circles connected by dashed lines).

seen that the main transitions occur between neighbor modes, i.e., between modes 1 and 2 or between modes 2 and 3. However, on two occasions in the test set (at samples 757 and 790), a direct switching between modes 1 and 3 occurs, and in some situations (samples 740–742, 745–747 and 776–778) a switching from one extreme mode to the other with only one application of the intermediate mode is seen. Another interesting observation is that it seems that mode 1 is usually applied to describe an increase of the silicon content from low values, while mode 3, by contrast, is used to describe a decrease from high values. Yet another noteworthy observation is that even though the three models implement changes on different silicon content levels, the highest value of mode 1 is greater than the lowest value of mode 3, as seen in the end of the test set. In summary, the switching model has been able to capture level changes in the outputs and, furthermore, changes in the process dynamics that will be discussed next.

In order to further analyze the three modes of operation applied to describe the hot metal silicon content in the blast furnace in question, the values of the parameters (after their estimation on the training set) are depicted graphically in Fig. 8. The order of appearance of the parameters is defined in the lowest subpanel. From these results, the following may be concluded: mode 1 is primarily based on the second input, i.e., the specific blast volume, which is associated with the efficiency of the reduction process: A low value of the specific blast volume implies a high production rate, a low level of the coke reserve, which gives a low silicon content. Therefore, the fact that all parameters are clearly positive is in agreement with process knowledge. Furthermore, mode 1 stresses the importance of this input with longer time lags, which may reflect the inertia of the high-temperature region. As for the other to inputs in mode 1, the resistance index (Input 1) shows a positive, but rather weak correlation, while the overall effect of the third input, the tuyere heat loss, is almost negligible as the sum of the parameters is practically zero. Mode 2, in turn, is characterized by an influence of all inputs. Even though the effect of the specific blast volume is strongest like in mode 1 (with a parameter sum of 0.51), the importance now decreases with the time lag, so recent values of this input are more relevant. The gas resistance (Input 1) also plays a strong role, with a time lag of (mainly) 2. The effect of the tuyere heat loss is positive for lags 1 and 2, but the negative parameter value at lag 3 decreases the effect and leads to a more complicated dynamic response (as will be discussed later). The sum of the parameters is still slightly positive (0.11). As for mode 3, even though all parameters but one are positive, it reflects clearly different dynamics than modes 1 and 2. The tuyere heat loss (Input 3) is the most influential variable (parameter sum 0.67), and the impact gets stronger with decreasing time lag, so the most recent value of the heat loss plays a key role for the output of this mode. For the first input, a more complex dynamic is implemented, as the parameters for lags 2 and 3 are of the same absolute size, but of opposite sign. Thus, an increase in the gas resistance will first yield an increase in the silicon content after two hours, then the effect is cancelled to be followed by a decrease. This complex behavior

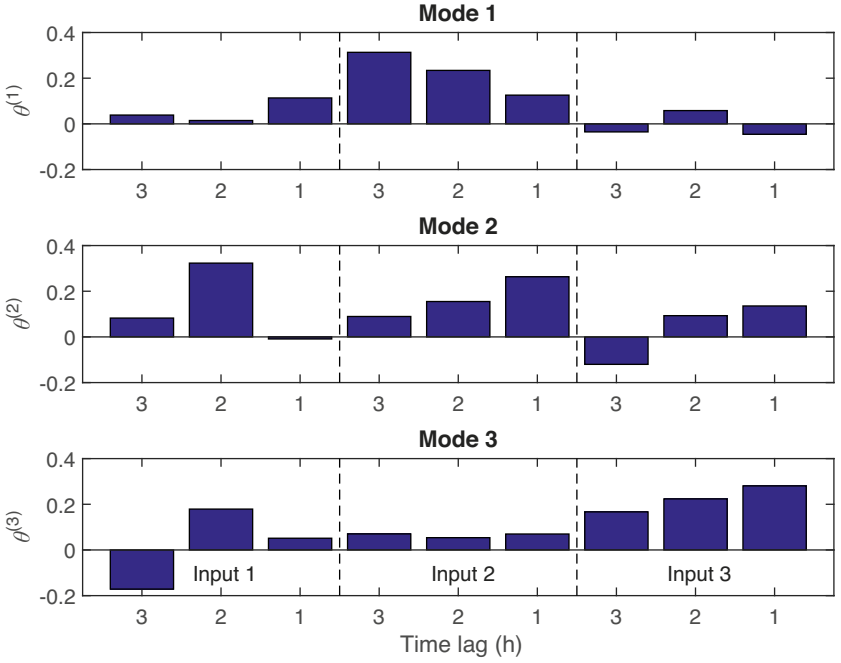

Fig. 8. Parameter values of Eq. (26) for the three modes after the training phase, arranged according to the definition in the bottom panel.

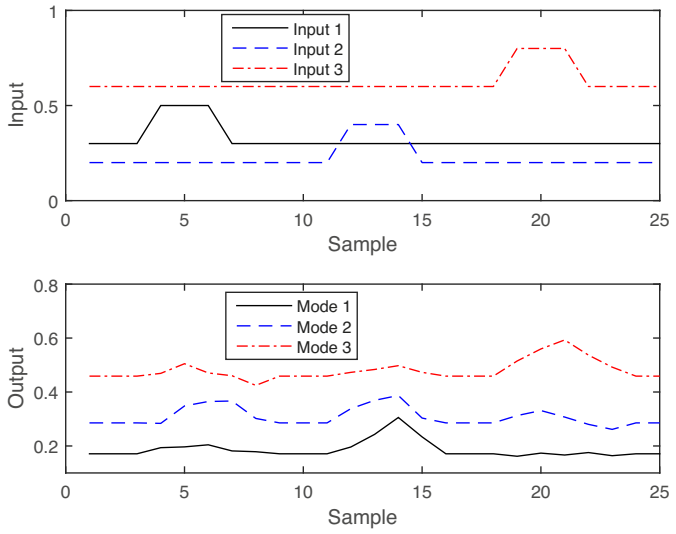

Fig. 9. Artificial set of inputs (top panel) for illustration of the dynamics of the output predicted by the three modes (bottom panel).

reflects the immediate and longer-term effect of a gas resistance change: As the gas resistance increases, the hot metal production rate decreases, which leads to a higher silicon content. However, as the gas resistance continues to stay high, reflecting a nonuniform gas flow, the reduction and heat transfer efficiency of the furnace drops and hence the silicon content decreases. It is a well-known fact that the hot metal temperature and chemistry may show corresponding inverse responses after changes in the fuel injection rates [26]. Finally, the effect of the specific blast volume (Input 2) is positive, but minor.

To shed further light on these findings, an artificial input sequence was generated where one input in turn was perturbed by raising the value by 0.2 from its base level during three consecutive samples, keeping the other inputs on their base levels. The base levels of the three inputs were selected as 0.3, 0.2 and 0.6, mimicking the mean values of these variables in the test set (cf. last 100 observations of x_i in Fig. 5). Fig. 9 illustrates the inputs (upper panel) and the outputs of the three modes (lower panel) based on the model parameters in Fig. 8. By analyzing the lower panel, the following may be concluded: First, the different levels of the output in the three modes is clearly seen. As for the dynamic responses, mode 1 (solid line) is mainly affected by changes in Input 2, as noted in conjunction with Fig. 8, and the effect increases with time. In mode 2 (dashed line), the response to the two first inputs is similar to that of a first-order system, where the former is lagged by two and the latter by one time step. The different signs of the parameters of Input 3, in turn, give rise to a somewhat more complex response. Finally, mode 3 (dash-dotted line) reacts mainly on changes in the third input.

Summarizing the findings, we may also conclude that modes 1 and 2 show a strong positive correlation with the second input, which means that an increase (decrease) in the specific blast volume would increase (decrease) the silicon content unless it is not already high. The specific blast volume essentially expresses the ratio of the gas and hot metal flow, and a high ratio means that more SiO gas (generated in the combustion of coke in the raceways in front of the tuyeres) is available for reaction into Si dissolved in the hot metal [29]. Furthermore, also considering the findings presented in Fig. 7, a recovery of the silicon content from very low levels is reflected in an increase in the specific blast volume. This may be due to control actions by the blast furnace operators: as the silicon content drops low, the operators charge extra coke, which decreases the production rate and, hence, increases the specific blast volume when the extra coke has descended to the lower part of the furnace. Changes in the first input (gas resistance) mainly affects the silicon content when it is on an intermediate level. Therefore, fluctuations in the silicon content around the set point are reflected in gas resistances changes [23], so controlling the gas resistance could stabilize the silicon content. Finally, the third input, the tuyere heat loss, reflects changes in the silicon content when the heat level is high. As noted earlier, a decrease of the silicon content from high levels is captured by mode 3. Thus, a decrease in the tuyere heat loss is a good indication that the silicon content, after being high, will decrease. Summing up, these observations demonstrate that the findings from the multiple-mode model can help understanding the complex dynamics of the silicon content of the hot metal in the blast furnace.

5. Conclusions

An identification method for switching systems has been presented. The method does not assume that the system mode depends on the state, but may change in an arbitrary way. In the proposed procedure the identification problem is formulated as a generalized plane clustering problem. The resulting problem is solved using a method which can be considered as a modification of self-organizing maps. One advantage of the proposed identification method is that in analogy with standard self-organizing maps, it is simple to implement, and it can be applied both off line or recursively to on-line identification. Numerical simulations show that the proposed technique efficiently identifies the correct modes and parameters.

On a set of industrial data taken from an ironmaking blast furnace, the method was illustrated to provide reasonable predictions of the output variable, the hot metal silicon content, thus demonstrating the potential of the technique for real-time applications. Furthermore, the switching between the modes was analyzed and the interpretation provided by the modes was discussed. The results of the analysis with multiple models demonstrated that the modes describe different dynamics and also different operation levels of the silicon content, further revealing interesting findings on the way in which the input variables affected the output in different states. Despite the black-box nature of the model, a detailed analysis of the arising modes can provide additional information about the factors governing the silicon content in the hot metal. These findings may prove useful, e.g., in the design of strategies for automatic control of the silicon content.

References

[1] L. Bako, G. Mercère, S. Lecœuche, On-line structured identification of switching systems with possibly varying orders, Int. J. Control 82 (2009) 1496–1515.
[2] L. Bako, Identification of switched linear systems via sparse optimization, Automatica 47 (2011) 668–677.
[3] G.A. Barreto, A.F.R. Araújo, Identification and control of dynamical systems using the self-organizing map, IEEE Trans. Neural Netw. 15 (2004) 1244–1259.
[4] G.A. Barreto, Time series prediction with the self-organizing map: a review, in: B. Hammer, P. Hitzler (Eds.), Perspectives of Neural-symbolic Integration, Springer, Berlin, Heidelberg, 2007, pp. 135–158.
[5] A. Bemporad, A. Garulli, S. Paoletti, A. Viconi, A greedy approach to identification of piecewise affine models, in: Hybrid Systems: Computation and Control, volume 2623 of Lecture Notes in Computer Science, Springer-Verlag, 2003, pp. 97–112.
[6] S. Boubaker, M. Djemai, N. Manamanni, F. M'Sahli, Active modes and switching instants identification for linear switched systems based on discrete particle swarm optimization, Appl. Soft Comput. 14 (2014) 482–488.
[7] P.S. Bradley, O.L. Mangasarian, *k*-Plane clustering, J. Glob. Optim. 16 (2000) 23–32.
[8] E. Domlan, B. Huang, J. Ragot, D. Maquin, Robust identification of switched regression models, IET Control Theory Appl. 3 (2009) 1578–1590.
[9] M. Elad, Sparse and Redundant Representations, Springer, New York, 2010.
[10] G. Ferrari-Trecate, M. Muselli, D. Liberati, M. Morari, A clustering technique for the identification of piecewise affine systems, Automatica 39 (2003) 205–217.

[11] Y. Hashambhoy, R. Vidal, Recursive identification of switched ARX models with unknown number of models and unknown orders, in: Proceedings of the 44th IEEE Conference on Control and Decision, Seville, 2005, pp. 6115–6121.
[12] A. Lj. Juloski, S. Weiland, W.P.M.H. Heemels, A Bayesian approach to identification of hybrid systems, IEEE Trans. Autom. Control 50 (2005) 1520–1532.
[13] M.Y. Kiang, Extending the Kohonen self-organizing map networks for clustering analysis, Comput. Stat. Data Anal. 38 (2001) 161–180.
[14] T. Kohonen, Self-organizing Maps, Springer, Berlin, Heidelberg, 1995.
[15] J. Lampinen, E. Oja, Self-organizing maps for spatial and temporal AR models, in: Proceedings of the 6th Scandinavian Conference on Image Analysis, Oulu, 1989, pp. 120–127.
[16] D. Liberzon, Switching in Systems and Control, Birkhauser, 2003.
[17] K. Matsuda, N. Nagai, M. Konishi, K. Kadoguchi, T. Yabata, Forecasting system for decreasing heat levels in blast furnace, Kobelco Technol. Rev. 5 (1989) 1–4.
[18] H. Nakada, K. Takaba, T. Katayama, Identification of piecewise affine systems based on statistical clustering technique, Automatica 41 (2005) 905–913.
[19] A. Nurkkala, F. Pettersson, H. Saxén, Blast furnace dynamics using multiple autoregressive models with exogenous inputs, ISIJ Int. 33 (2012) 1764–1771.
[20] M.S. Phadke, S.M. Wu, Identification of multiinput–multioutput transfer function and noise model of a blast furnace from closed-loop data, IEEE Trans. Autom. Control 19 (1974) 944–951.
[21] J. Principe, L. Wang, M. Motter, Local dynamic modeling with self-organizing maps and application to nonlinear system identification and control, Proc. IEEE 86 (1998) 2240–2258.
[22] J. Roll, A. Bemporad, L. Ljung, Identification of piecewise affine systems via mixed-integer programming, Automatica 40 (2004) 37–50.
[23] H. Saxén, Short-term prediction of silicon content in pig iron, Can. Metall. Q. 33 (1994) 319–326.
[24] H. Saxén, F. Pettersson, Nonlinear prediction of the hot metal silicon content in the blast furnace, ISIJ Int. 47 (2007) 1732–1737.
[25] H. Saxén, C.H. Gao, Z. Gao, Data-driven time discrete models for dynamic prediction of the hot metal silicon in the blast furnace – a review, IEEE Trans. Ind. Inf. 9 (2013) 2213–2225.
[26] C. Staib, N. Jusseau, J. Vigliengo, J.C. Cochery, Blast furnace dynamic behavior and automatic control, Ironmak. Proc. 26 (1967) 66–83.
[27] Z. Sun, S.S. Ge, Switched Linear Systems: Control and Design, Springer, 2005.
[28] M. Tabatabaei-Pour, K. Salahshoor, B. Moshiri, A modified *k*-plane clustering algorithm for identification of hybrid systems, in: Proceedings of the 6th World Congress on Intelligent Control and Automation, Dalian, China, 2006, pp. 1333–1337.
[29] N. Tsuchiya, M. Tokuda, M. Ohtani, Transfer of silicon from gas-phase to molten iron in the blast furnace, Metall. Trans. B 7B (1976) 315–320.
[30] R. Vidal, S. Soatto, Y. May, S. Sastry, An algebraic approach to the identification of linear hybrid systems, in: Proceedings of the 42nd IEEE Conference on Control and Decision, Hawaii, 2003, pp. 167–172.
[31] R. Vidal, Recursive identification of switched ARX systems, Automatica 44 (2008) 2274–2287.
[32] M. Waller, H. Saxén, On the development of predictive models with applications to a metallurgical process, Ind. Eng. Chem. Res. 39 (2000) 982–988.