A STATE-SPACE APPROACH TO MODELING FUNCTIONAL TIME SERIES APPLICATION TO RAIL SUPERVISION

Allou Samé, Hani El-Assaad

Université Paris-Est, IFSTTAR
COSYS, GRETTIA
Champs-sur-Marne, France

ABSTRACT

This article introduces a state-space model for the dynamic modeling of curve sequences within the framework of railway switches online monitoring. In this context, each curve has the peculiarity of being subject to multiple changes in regime. The proposed model consists of a specific latent variable regression model whose coefficients are supposed to evolve dynamically in the course of time. Its parameters are recursively estimated across a sequence of curves through an online Expectation-Maximization (EM) algorithm. The experimental study conducted on two real power consumption curve sequences from the French high speed network has shown encouraging results.

Index Terms— Time series of functional data, state-space model, Kalman filtering, online Expectation-Maximization (EM) algorithm, condition monitoring

1. INTRODUCTION

Condition monitoring has become a powerful decision-making support for the preventive maintenance of railway infrastructure and rolling stock. It consists in assessing their operating state using condition measurements usually acquired through embedded sensors. In many cases, the acquired data take the form of a sequence of curves, which, in a statistical framework, is also referred to as a series of functional data. In the case of the French high speed lines switch mechanisms, which are considered in this article, each curve represents the electrical power consumption during a switch operation and is possibly made of several regimes.

For assessing the operating state of this system, the classical pattern recognition approach consists in the following two steps: the extraction of a set of relevant patterns from each curve using a segmental regression model, and the supervised classification of the extracted patterns into known operating states or classes [1]. Another approach, which directly performs the classification on the raw curves, has been adopted in [2]. It relies on modeling each class of curves using the previously evoked segmental regression model and on classifying the curves using the Bayes discrimination rule. An extension of this method, which deals with complex shaped classes potentially made of sub-classes, has recently been developed [3]. These (static) approaches require the operating states to be preliminary characterized, which is usually done by using a labeled collection of curves. For each curve of this learning database, the operating state must have been preidentified by an expert. Nevertheless, the collected curves are often unlabelled and besides, the operating states are not all known in advance. One could refer to curve clustering approaches to automatically identify groups of switch operations that show similar dynamic behavior [4]. However, these offline clustering approaches are not designed to analyze a series of curves.

This article proposes a new approach for online monitoring of the railway switch mechanism. For this purpose, a segmental dynamic regression model is introduced in this paper, whose parameters are recursively estimated across a sequence of curves. The proposed model can be formulated as a specific state-space model [5] [6]. From the estimated parameters, some real valued indicators can be extracted and analyzed in the course of time in order to help in monitoring the system.

The paper is organized as follows. Section 2 describes the curve sequences used in our application and section 3 briefly recalls the specific regression model used for the static modeling of our curves. Then, section 4 presents the dynamic model and its recursive parameter estimation technique. In section 4, the proposed methodology is applied on sequences of electrical power consumption curves from successive switch operations on the French railway network.

2. POWER CONSUMPTION CURVES SEQUENCES ACQUIRED FROM SWITCH OPERATIONS

As mentioned in the introduction, the main motivation behind this study was the monitoring of the railways switches that allow trains to change tracks at junctions. A switch operation consists in moving laterally some linked tapering rails (also known as points) into one of two positions. In the case of the

The authors wish to thank M. Marc Antoni of SNCF for the data he provided and for the support provided.

French high speed lines, this operation is generally operated by an electrical motor (380V alternative current).

The monitoring task is performed by temporally analyzing sequences of electrical power consumption curves recorded during switching operations. Each curve is sampled at 100 Hz (100 points are recorded per second) and observed over 5 seconds ($m = 500$ points per curve). Figure 1 shows a sequence of 50 power consumption curves acquired during successive switch operations.

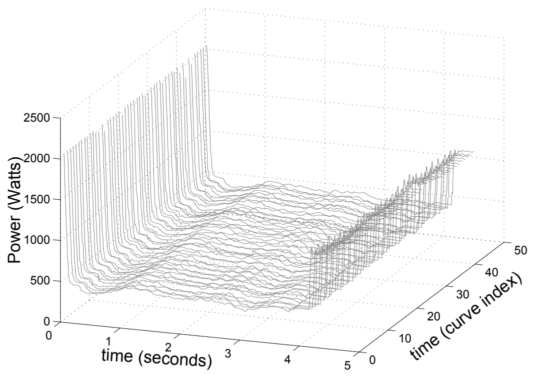

Fig. 1. Example of a sequence of 50 power consumption curves acquired during successive switch operations.

The specificity of the curves to be analyzed in this context is that they are subject to various changes in regime as a result of five successive mechanical movements of the physical components associated with the switch mechanism:

- the starting phase: period between the activation of the motor and the starting of the switch operation itself,
- the points unlocking: phase where the switch points are unlocked, that makes them ready for the translation,
- the points translation: phase corresponding to the translation of the points,
- the points locking: phase where the switch points are locked;
- the friction phase: phase where an additional effort is applied to ensure the locking.

3. STATIC REGRESSIVE REPRESENTATION OF CURVES WITH CHANGES IN REGIME

Let $\mathbf{x}_t$ denote a power consumption curve, where $\mathbf{x}_t = (x_{t1}, \ldots, x_{tn})$ consists of n real values observed over a time grid indexed by the integers $(1, \ldots, n)$. Before introducing the model proposed in this study, let us recall the so-called "Regression model with Hidden Logistic Process" (RHLP) that was used for the modeling and discrimination of curves subject to changes in regime [1]. According to this model, each individual observations x_{tj} of a curve $\mathbf{x}_t$ follows one of K regression models associated to the regimes involved in the curve generation process:

$$x_{tj} = \mathbf{U}'_j \boldsymbol{\beta}_{z_{tj}} + \sigma_{z_{tj}} \varepsilon_{tj}, \qquad (1)$$

where $\epsilon_{tj} \sim \mathcal{N}(0, 1)$ is a random Gaussian noise and $z_{tj} \in \{1, \ldots, K\}$. For each $k \in \{1, \ldots, K\}$, the parameters $\sigma_k \in \mathbb{R}$ and $\boldsymbol{\beta}_k \in \mathbb{R}^{q+1}$ are respectively the noise standard deviation and the coefficient vector of the kth regression model of degree q. The transpose vector $\mathbf{U}'_j$ denotes the vector of regressors $(1, j, j^2, \ldots, j^q)$ associated to $\boldsymbol{\beta}_k$. The assignment of the x_{tj} to the different regression models is specified by the random process denoted by $\mathbf{z}_t = (z_{t1}, \ldots, z_{tj}, \ldots, z_{tn})$. For the process $\mathbf{z}_t$ to define a segmentation into contiguous segments, the variable z_{tj} is supposed to be randomly drawn according to the multinomial distribution $\mathcal{M}(1; \pi_1(j, \boldsymbol{\alpha}), \ldots, \pi_K(j, \boldsymbol{\alpha}))$, where the probabilities $\pi_k(j, \boldsymbol{\alpha})$ are defined as the following logistic functions [1] [2] [4]:

$$\pi_k(j, \boldsymbol{\alpha}) = p(z_{tj} = k) = \frac{\exp(\alpha_{k1} j + \alpha_{k0})}{\sum_{\ell=1}^{K} \exp(\alpha_{\ell 1} j + \alpha_{\ell 0})}, \qquad (2)$$

with $\boldsymbol{\alpha} = (\alpha_{k0}, \alpha_{k1}, k = 1, \ldots, K) \in \mathbb{R}^{2K}$. It should be noticed that the logistic functions $\pi_k(j, \boldsymbol{\alpha})$ verify $0 < \pi_k(j; \boldsymbol{\alpha}) < 1$ and $\sum_{k=1}^{K} \pi_k(j; \boldsymbol{\alpha}) = 1$. It can be shown that the equations defined below can be encompassed into a single mixture of Gaussian probability density functions of x_{ti} defined by:

$$p(x_{tj}; \boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k(j; \boldsymbol{\alpha}) \mathcal{N}(x_{tj}; \mathbf{U}'_j \boldsymbol{\beta}_k, \sigma_k^2), \qquad (3)$$

where $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K, \sigma_1^2, \ldots, \sigma_K^2)$ is the global parameter vector of the model.

The parameters of this model can be estimated from a single curve or from several curves sharing the same characteristics, by maximizing the likelihood function through the Expectation-Maximization (EM) algorithm [7] [1]. From a practical point of view, the parameters estimated from power consumption curves can be used as feature vectors to discriminate between normal operating states and defects [1]. Figure 2 shows the polynomials estimated from a power consumption curve and its associated logistic probabilities.

4. A STATE-SPACE MODEL FOR CURVES SEQUENCES MODELING

The model introduced in this section extends the one described previously, to deal with a time series of power consumption curves. We are interested in modeling the temporal evolution of a curve sequence. Based on this dynamical representation, outliers and changes of behavior in the sequence of operations will be clearly detectable. Let $(\mathbf{x}_1, \ldots, \mathbf{x}_T)$ denote a sequence of T power consumption curves observed over the same time grid, where $\mathbf{x}_t = (x_{t1}, \ldots, x_{tn})$.

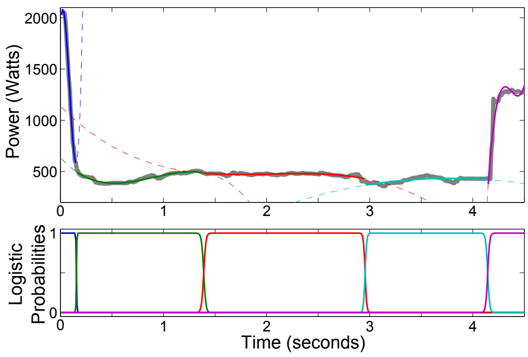

Fig. 2. Regression model estimated from a power consumption curve, with $K = 5$ and $q = 3$ (top) and corresponding logistic probabilities (bottom)

4.1. Model definition

The model that we propose for a sequence of curves is the dynamic extension of the model previously described. It is defined by the pair of equations

$$x_{tj} = \mathbf{U}'_j \boldsymbol{\beta}_{t,z_{tj}} + \sigma_{z_{tj}} \varepsilon_{tj} \quad (4)$$

$$\boldsymbol{\beta}_{tk} = \boldsymbol{\beta}_{t-1,k} + v\, \boldsymbol{\eta}_{tk} \quad (5)$$

where $\boldsymbol{\eta}_t \sim \mathcal{N}(0, \mathbf{I})$ is a Gaussian noise in $\mathbb{R}^{q+1}$, $\mathbf{I}$ being the identity matrix in $\mathbb{R}^{q+1}$. In this dynamic model, the regression coefficient vector sequence $(\boldsymbol{\beta}_{tk})_{t=1,\ldots,T}$, for each segment k, is modeled as a Gaussian random walk, the variance v^2 ensuring a trade-off between model fit and time variation. As for the static model, we suppose that the components $z_{tj}(j = 1, \ldots, n)$ of the random process $\mathbf{z}_t = (z_{t1}, \ldots, z_{tn})$ are independently generated from the multinomial law $\mathcal{M}(1; \pi_1(j, \boldsymbol{\alpha}), \ldots, \pi_K(j, \boldsymbol{\alpha}))$. The parameter vector of this dynamic model can therefore be denoted as $\boldsymbol{\Phi} = (\boldsymbol{\alpha}, v^2, (\boldsymbol{\beta}_{0k}), (\sigma_k^2))$, the variables $\boldsymbol{\beta}_{tk}$ and z_{tj}, for $t \geq 1$, being considered as latent variables.

It can be shown that, conditionally on $\boldsymbol{\beta}_t = (\boldsymbol{\beta}_{tk})_{k=1,\ldots,K}$, the variable x_{tj} is distributed according to a Gaussian mixture density and that, conditionally on $\boldsymbol{\beta}_{t-1,k}$, the variable $\boldsymbol{\beta}_{tk}$ is distributed according to a Gaussian density. These densities are defined as follows:

$$p(x_{tj}|\boldsymbol{\beta}_{tk}; \boldsymbol{\Phi}) = \sum_{k=1}^{K} \pi_k(j; \boldsymbol{\alpha}) \mathcal{N}(x_{tj}; \mathbf{U}'_j \boldsymbol{\beta}_{tk}, \sigma_k^2) \quad (6)$$

$$p(\boldsymbol{\beta}_{tk}|\boldsymbol{\beta}_{t-1,k}; \boldsymbol{\Phi}) = \mathcal{N}(\boldsymbol{\beta}_{tk}; \boldsymbol{\beta}_{t-1,k}, v^2\mathbf{I}). \quad (7)$$

4.2. Recursive EM algorithm

Let $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$ denote the sequence of observed curves, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_T)$ the latent regression coefficients and $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_T)$ the latent segments.

In case the complete curve sequence $\mathbf{x}$ is available before the analysis, the usual way to estimate the dynamic model parameters consists in maximizing the log-likelihood criterion $\log p(\mathbf{x}; \boldsymbol{\Phi})$ via the Expectation-Maximization (EM) algorithm [7] [8]. Given the maximum likelihood estimate $\widehat{\boldsymbol{\Phi}}$ of $\boldsymbol{\Phi}$, the latent variables can be estimated by their posterior expectation $E(\boldsymbol{\beta}|\mathbf{x}; \widehat{\boldsymbol{\Phi}})$ and $E(\mathbf{z}|\mathbf{x}; \widehat{\boldsymbol{\Phi}})$. Unfortunately, the E-step of this algorithm is intractable because it requires successive integrations whose number grows exponentially with T. To get an exact inference learning algorithm, we suggest, instead of maximizing the classical log-likelihood, to maximize the criterion

$$\mathcal{L}(\boldsymbol{\Phi}) = \max_{\boldsymbol{\beta}} \log p(\mathbf{x}, \boldsymbol{\beta}; \boldsymbol{\Phi}), \quad (8)$$

which is equivalent to maximize the criterion

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\Phi}) = \log p(\mathbf{x}, \boldsymbol{\beta}; \boldsymbol{\Phi}), \quad (9)$$

with respect to $(\boldsymbol{\beta}, \boldsymbol{\Phi})$. A specific EM algorithm can be derived to solve this optimization problem.

However, for our purposes, the estimation must be performed concurrently with data acquisition. In this case, a recursive version of the EM algorithm is used to estimate both the parameters and latent regression coefficients. Given starting values $(\boldsymbol{\beta}_0, \boldsymbol{\Phi}^{(0)})$, it consists, while new curves $\mathbf{x}_{t+1}$ are forthcoming, in computing the new estimates $(\boldsymbol{\beta}_{t+1}, \boldsymbol{\Phi}^{(t+1)})$, by maximizing the auxiliary criterion defined by:

$$Q_{t+1}(\boldsymbol{\beta}_{t+1}, \boldsymbol{\Phi}) = Q_t(\boldsymbol{\beta}_t, \boldsymbol{\Phi}) + \sum_{j,k} \tau_{t+1,j,k} \log \pi_k \mathcal{N}(x_{t+1,j}; \mathbf{U}'_j \boldsymbol{\beta}_{tk}) + \sum_{k=1}^{K} \log \mathcal{N}(\boldsymbol{\beta}_{t+1,k}; \boldsymbol{\beta}_{tk}, v^2), \quad (10)$$

where $Q_0 = 0$ and $\tau_{t+1,j,k}$ is the posterior probability defined by equation (11). The resulting recursive EM algorithm is then defined by the following steps.

Initialization

- Compute the initial logistic probabilities coefficient $\boldsymbol{\alpha}^{(0)}$, regression coefficients $(\boldsymbol{\beta}_{0k})$ and variances $\left(\sigma_k^{2^{(0)}}\right)$ by identifying the "static model" described by equations (1) and (3) on the five first curves, and set $v^{2^{(0)}} = 1$.
- Set $t = 0$.

E-Step (Expectation)

For each new curve $\mathbf{x}_{t+1}$, compute for $j = 1 \ldots, n$ and for $k = 1, \ldots, K$, the posterior probability that $x_{t+1,j}$ originates from the kth regression model, given the previous parameters and regression coefficients:

$$\tau_{t+1,j,k} = \frac{\pi_k(j; \boldsymbol{\alpha}^{(t)}) \mathcal{N}(x_{t+1,j}\,; \mathbf{U}'_j \boldsymbol{\beta}_{tk}, \sigma_k^{2^{(t)}})}{\sum_{\ell=1}^{K} \pi_\ell(j; \boldsymbol{\alpha}^{(t)}) \mathcal{N}(x_{t+1,j}\,; \mathbf{U}'_j \boldsymbol{\beta}_{t\ell}, \sigma_\ell^{2^{(t)}})} \quad (11)$$

M-Step (Maximization)

- Update the regression coefficients using Kalman filtering recursions:

$$\boldsymbol{\beta}_{t+1,k} = \boldsymbol{\beta}_{tk} + \\ = \mathbf{K}_{t+1,k} \sum_{j=1}^{n} \tau_{t+1,j,k}\, \mathbf{U}_j \left(x_{t+1,j} - \mathbf{U}_j' \boldsymbol{\beta}_{tk}\right) \quad (12)$$

where the Kalman gain $\mathbf{K}_{t+1,k}$ is defined by

$$\mathbf{K}_{t+1,k} = \Big[\big(\sum_{j=1}^{n} \tau_{t+1,j,k}\, \mathbf{U}_j \mathbf{U}_j'\big) + \big(\sigma_k^{2^{(t)}}/v^{2^{(t)}}\big)\mathbf{I}\Big]^{-1}.$$

- Update the variances:

$$v^{2^{(t+1)}} = \mathbf{S}_{t+1}/(t+1) \quad (13)$$
$$\sigma_k^{2^{(t+1)}} = \mathbf{S}_{t+1,k}/\mathbf{R}_{t+1,k}, \quad (14)$$

with

$$\mathbf{S}_{t+1} = \mathbf{S}_t + \sum_k \left\|\boldsymbol{\beta}_{t+1,k} - \boldsymbol{\beta}_{tk}\right\|^2$$
$$\mathbf{S}_{t+1,k} = \mathbf{S}_{tk} + \sum_j \tau_{t+1,j,k}\ \left(x_{t+1,j} - \mathbf{U}_j' \boldsymbol{\beta}_{t+1,k}\right)^2$$
$$\mathbf{R}_{t+1,k} = \mathbf{R}_{tk} + \sum_j \tau_{t+1,j,k}\ ,$$

where $\mathbf{S}_0 = \mathbf{S}_{0,k} = \mathbf{R}_{0,k} = 0$.

- Update the logistic regression coefficients:

$$\boldsymbol{\alpha}^{(t+1)} = \boldsymbol{\alpha}^{(t)} - [\mathbf{H}(\boldsymbol{\alpha}^{(t)})]^{-1}\, \mathbf{g}(\boldsymbol{\alpha}^{(t)}), \quad (15)$$

where $\mathbf{g}(\boldsymbol{\alpha}^{(t)})$ and $\mathbf{H}(\boldsymbol{\alpha}^{(t)})$ are respectively the vector and the matrix defined by:

$$\mathbf{g}(\boldsymbol{\alpha}^{(t)}) = \big(\mathbf{g}_k(\boldsymbol{\alpha}^{(t)})\big)_{1\leq k\leq K-1}$$
$$\mathbf{H}(\boldsymbol{\alpha}^{(t)}) = \big(\mathbf{H}_{k\ell}(\boldsymbol{\alpha}^{(t)})\big)_{1\leq k,\ell\leq K-1}$$

with

$$\mathbf{g}_k(\boldsymbol{\alpha}^{(t)}) = \sum_j \big(\tau_{t+1,j,k} - \pi_k(j;\boldsymbol{\alpha}^{(t)})\big),$$
$$\mathbf{H}_{k\ell}(\boldsymbol{\alpha}^{(t)}) = -(t+1) \times \sum_j \pi_k(j;\boldsymbol{\alpha}^{(t)})\big(\delta_{k\ell} - \pi_\ell(j;\boldsymbol{\alpha}^{(t)})\big)$$

where $\delta_{k\ell} = 1$ if $k = \ell$ and 0 otherwise.

- Set $t = t + 1$.

5. APPLICATION TO THE MONITORING OF SWITCH OPERATIONS

Given a curve sequence, a mixture of $K = 5$ polynomials of order $p = 3$ are recursively estimated according to the algorithm described in the previous section. Concurrently with parameters estimation, a single indicator defined by

$$\boldsymbol{\mu}_k(t) = \frac{\sum_{j=1}^{n} \pi_k(j;\boldsymbol{\alpha}^{(t)})\mathbf{U}_j'\boldsymbol{\beta}_{tk}}{\sum_{j=1}^{n} \pi_k(j;\boldsymbol{\alpha}^{(t)})} \quad (16)$$

is extracted from each phase $k \in \{1,\ldots,5\}$, at each time t. This numerical indicator can be interpreted as the average value of the kth polynomial curve over the kth segment.

Our descriptive condition monitoring approach then consists in visualizing and analyzing the five series of indicators $(\boldsymbol{\mu}_1(t))_t, \ldots, (\boldsymbol{\mu}_5(t))_t$. These indicators charts extracted from the estimated parameters provide a descriptive tool for visualizing more easily the dynamic behind the curves regimes. Based on this representation, atypical curves (outliers) and changes in the behavior of the curve sequence can be clearly detectable.

This strategy has been applied on two curve sequences (see figure 3) recorded from two different point machines. Their characteristics are the following:

- curve sequence A:
  - time period : from June 01 2011 to July 31 2011,
  - length of the curve sequence: $T = 872$,
  - average number of operations per day: 15.
- curve sequence B:
  - time period : from June 01 2011 to July 31 2011,
  - length of the curve sequence: $T = 1817$,
  - average number of operations per day: 30.

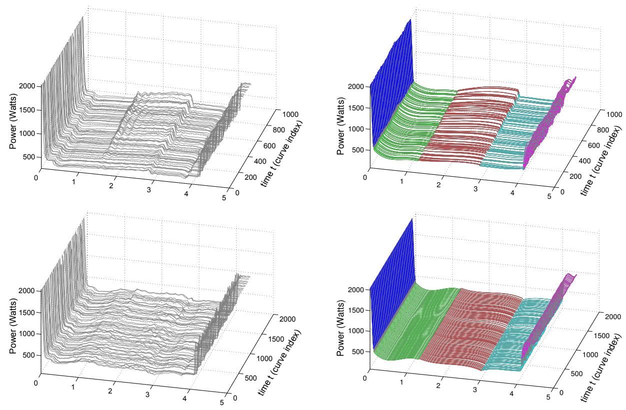

Fig. 3. Curve sequences A (top-left) and B (bottom-left) and their estimation (top-right and bottom-right); each estimated curve is the sum of estimated polynomial regression curves weighted by the logistic probabilities.

Figure 4 shows the series of indicators corresponding to the unlocking $(\boldsymbol{\mu}_2(t))$, translation $(\boldsymbol{\mu}_3(t))$ and locking $\boldsymbol{\mu}_4(t)$ phases. The series corresponding to the starting $(\boldsymbol{\mu}_1(t))$ and the friction $(\boldsymbol{\mu}_5(t))$ phases, which vary only slightly, are not displayed. For the curve sequence A, a slow degradation is revealed during the translation phase, which can be attributed to a lubrication defect. This point is illustrated by the figure 5(a), which clearly shows the increasing power consumption during the translation phase. For curve sequence B, a one-time anomaly (outlier B1) is observed during the unlocking phase. The three charts also reveal a slight global change (change point B2) in the behavior of the switch operations. To illustrate the observations made from the time series of indicators, figure 5(b) shows the atypical curve together with normal curves, and figure 5(c) displays 50 curves before the change point B2 and 50 curves after this one.

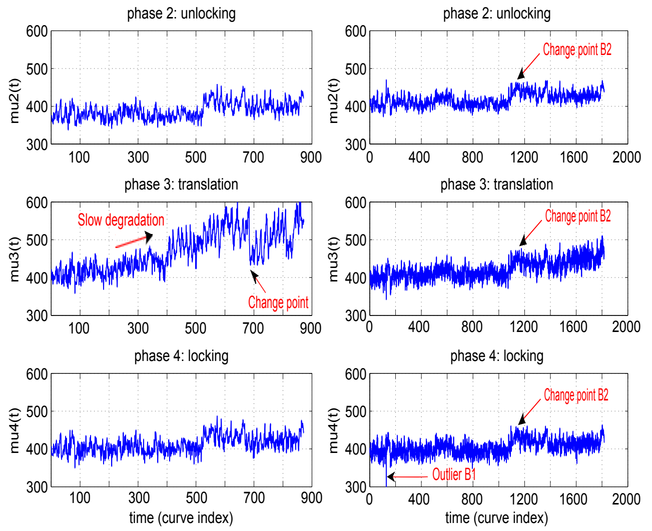

Fig. 4. Temporal evolution of indicators $\boldsymbol{\mu}_k(t)$ associated to the unlocking, translation and locking phases of a switch operation, obtained for curve sequences A (left) and B (right).

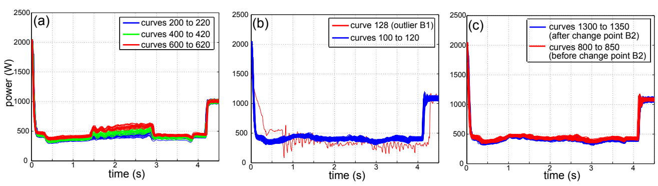

Fig. 5. Curves illustrating the singular points identified from the temporal analysis of the indicators $\boldsymbol{\mu}_k(t)$ (see figure 4).

6. CONCLUSION

A dynamic probabilistic modeling of curve sequences, dedicated to the online monitoring of railway switch operations, has been introduced in this paper. It involves modeling the curves using conjointly several regression models whose temporal evolution is tracked across a sequence of curves. The model parameters are identified online using a recursive variant of the Expectation-Maximization algorithm whose M-step involves Kalman filtering recursions.

The experimental study conducted on two curve sequences acquired from switch operations on the French high speed network has shown encouraging results in terms of characterization of the temporal evolution of curves. The indicators charts extracted from the estimated parameters provides a descriptive tool for visualizing more easily the dynamic behind the curves regimes that can be augmented by detection thresholds tuned adequately according to preventive maintenance strategies. As the curve sequences to be studied change dynamically with time, suitable comparison methods were not available. One of the prospects of this work will be to numerically evaluate the proposed approach using simulated curves with known regression coefficients.

The proposed approach is sufficiently generic to be applied on other curve sequences. In this case, the problem of assessing the optimal number of segments and polynomial order could be addressed by using classical model selection criteria such as the Bayesian Information (BIC) criterion [9].

REFERENCES

[1] F. Chamroukhi, A. Samé, G. Govaert, and P. Aknin, "Time series modeling by a regression approach based on a latent process," *Neural Networks*, vol. 22, pp. 593–602, 2009.

[2] F. Chamroukhi, A. Samé, G. Govaert, and P. Aknin, "A hidden process regression model for functional data description. application to curve discrimination," *Neurocomputing*, vol. 73, pp. 1210–1221, 2010.

[3] F. Chamroukhi, H. Glotin, and A. Samé, "Model-based functional mixture discriminant analysis with hidden process regression for curve classification," *Neurocomputing*, vol. 112, pp. 153–163, 2013.

[4] A. Samé, F. Chamroukhi, G. Govaert, and P. Aknin, "Model-based clustering and segmentation of time series with changes in regime," *Adv. Data Analysis and Classification*, vol. 5, no. 4, pp. 301–321, 2011.

[5] R. H. Shumway and D. S. Stoffer, *Time series analysis and its applications*, Springer, 2011.

[6] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[7] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society B*, vol. 39, pp. 1–38, 1977.

[8] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, Wiley, New York, 2008.

[9] G. Schwarz, "Estimating the number of components in a finite mixture model," *Annals of Statistics*, vol. 6, pp. 461–464, 1978.