



**HAL**  
open science

## Bayesian curve fitting for transient signals by using smooth transition regression models

Matthieu Sanquer, Florent Chatelain, Mabrouka El-Guedri, Nadine Martin

► **To cite this version:**

Matthieu Sanquer, Florent Chatelain, Mabrouka El-Guedri, Nadine Martin. Bayesian curve fitting for transient signals by using smooth transition regression models. CM 2010 - MFPT 2010 - 7th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, Jun 2010, Stratford-upon-Avon, United Kingdom. pp.n.c. hal-00511097

**HAL Id: hal-00511097**

**<https://hal.science/hal-00511097>**

Submitted on 23 Aug 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Bayesian curve fitting by using smooth transition regression models

Journal:	<i>CM 2010 and MFPT 2010</i>
Manuscript ID:	CM-MFPT-0192-2010.R1
Topic:	Structured Sessions
Date Submitted by the Author:	16-Apr-2010
Complete List of Authors:	Sanquer, Matthieu; Gipsa-lab, DIS Chatelain, Florent; GIPSA-lab, DIS
Keywords:	Condition monitoring, Signal processing, Statistics



# Bayesian curve fitting for transient signals by using smooth transition regression models

Matthieu Sanquer\*, Florent Chatelain\*, Mabrouka El-Guedri†, Nadine Martin\*

\*University of Grenoble, GIPSA-lab, 961 rue de la Houille Blanche, BP 46, 38402 Saint Martin d'Hères, France

† Electricité de France – R&D, 4 Quai Watier, 78401 Chatou, France

matthieu.sanquer@gipsa-lab.grenoble-inp.fr, florent.chatelain@grenoble-inp.fr, mabrouka.el-guedri@edf.fr,  
nadine.martin@grenoble-inp.fr

## Abstract

This communication addresses the problem of fitting time series with smooth transition regression models. These models are of interest to characterize transient signals in the context of system monitoring and diagnosis. Within this modelling, time series are segmented by sequences of piecewise constant polynomial regression models. Moreover, smooth transitions between each segment are obtained by introducing some smooth, monotonically increasing parametric transition functions. It allows one to give a synthetic representation of signals composed by smooth transitions between different regimes. Estimation of the parameters of these models appears to be an ill-posed problem. Direct optimization algorithms are not robust enough with regard to the initial parameters guess. Therefore, to achieve parameters estimation, we introduce a Bayesian framework. Appropriate priors for the unknown model parameters are introduced to penalize a data-driven criterion built from the likelihood of the observations. As the resulting posterior probability distributions do not admit closed-form analytical expressions, Markov Chain Monte Carlo (MCMC) sampling methods are derived to obtain the standard Bayesian estimators of the model parameters. Results are shown for synthetic and real appliance load monitoring data.

# Bayesian curve fitting by using smooth transition regression models

Matthieu Sanquer\*, Florent Chatelain\*, Mabrouka El-Guedri†, Nadine Martin\*

\*University of Grenoble, GIPSA-lab, 961 rue de la Houille Blanche  
BP 46, 38402 Saint Martin d'Hères, France

† Electricité de France – R&D, 4 Quai Watier, 78401 Chatou, France

matthieu.sanquer@gipsa-lab.grenoble-inp.fr, florent.chatelain@grenoble-inp.fr  
mabrouka.el-guedri@edf.fr, nadine.martin@grenoble-inp.fr

## Abstract

This communication addresses the problem of fitting time series with smooth transition regression models. These models are of interest to characterize transient signals in the context of system monitoring and diagnosis. Within this modelling, time series are segmented by sequences of piecewise constant polynomial regression models. Moreover, smooth transitions between each segment are obtained by introducing some smooth, monotonically increasing parametric transition functions. It allows one to give a synthetic representation of signals composed by smooth transitions between different regimes. However, the estimation of the parameters is an ill-posed problem. Direct optimization algorithms are not robust enough with regard to the initial parameters guess. Therefore, the parameter estimation is achieved in a Bayesian framework. Appropriate priors for the unknown model parameters are introduced to penalize a data-driven criterion built from the likelihood of the observations. As the resulting posterior probability distributions does not admit closed-form analytical expressions, Markov Chain Monte Carlo (MCMC) sampling methods are derived to obtain the standard Bayesian estimators of the model parameters. Results are given for synthetic signals and for some household end-uses measurements.

## 1 Introduction

Residential non-intrusive load monitoring (NILM) consists in extracting information about the turned-on electrical appliances from the global electrical load. It aims at detecting and characterizing the active appliances plugged on the electrical house network. Such information is useful for electrical utilities to best manage energy quality or energy saving problems. From a customer point of view, it would allow one to get a better understanding of its electrical consumption. Finally, an

other application could be remote monitoring and surveillance of domestic electric network.

The electric appliances characterization problem can be decomposed into two steps: 1) feature extraction from signal and 2) learning of the parameters for the different classes corresponding to the different electrical appliances. This second step is usually performed by using a labeled database of signals. The overall detection and classification performance will depend on both, the quality of the feature extraction and the quality of the learning. First works on non-intrusive electric load monitoring were performed in [7]. The electric appliances characterization task was achieved by detecting step changes in signal, filtering transients and then clustering similar changes. The used descriptors are derived from the active and reactive power. In [8], *a priori* knowledge about the steady-states and the transients to be detected are used for supervised classification of appliances via a matched filter [8]. More recently, a global modelling of load curves by a hidden Markov model has been proposed [3]. This model is based on the residual of a variance analysis with respect to some exogenous variables such as the customer contract, the hour and the date,... Finally in [4], a point process framework is proposed to model the space-heating load by a linear mixture of square periodic signals with slowly time-varying duty cycles.

In this work, we choose to concentrate on the feature extraction problem. A general Bayesian method Our approach does not require a physical model of the system and simply assumes that the observed transients are fully characterized by the current envelope. This envelope is then segmented into sequences of piecewise constant polynomial, with smooth transitions between each of them. This model is based on the properties of electrical transients which are composed of smooth transitions between some quasi-constant current level corresponding to steady states. Thus, this segmentation framework allows one to reduce the electrical transient to a few number of parameters, leading to a synthetic characterization. However, the estimation of these parameters appears to be an ill-posed problem. Indeed, direct optimization algorithms are not robust enough with regard to the initial parameters guess. Parameter estimation is achieved in a Bayesian framework. Appropriate priors on the unknown model parameters are introduced to penalize a data-driven criterion built from the likelihood of the observations. As the resulting posterior probability distributions does not admit closed-form analytical expressions, Markov Chain Monte Carlo (MCMC) sampling methods are derived to obtain the standard Bayesian estimators of the model parameters.

Note that a quite similar model has been used in the case study of a railway infrastructure. More precisely, a piecewise polynomial model using a hidden switch process is introduced in [1]. The proposed model needs to know the number of transitions. Moreover, the hidden switch process appears to be not flexible enough to model the smooth transitions observed in electrical transients.

Note also that some works have previously been done about the estimation of regression models embedded in smooth transition processes. In particular, the estimation

problem of a two-regime regression model with smooth sigmoidal transition using a Bayesian framework is addressed in [9]. This paper aims at extending such a framework to a hierarchical Bayesian model with an unknown number of changes, and with any kind of transition functions.

This paper is organized as follows. Section 2 presents the Bayesian model used to fit the transient current envelope. In Section 3, a Gibbs sampling method is derived to estimate the model parameters. The application to electrical transient segmentation is discussed in Section 4. Conclusions and future works are finally reported in Section 5.

## 2 Smooth transition polynomial regression model

### 2.1 Model and notations

An observed signal  $\mathbf{x}$  of  $n$  samples  $(x_i)_{i=1,\dots,n}$  acquired at time instant  $(t_i)_{i=1,\dots,n}$  is considered. The transition function associated with the  $k^{\text{th}}$  rupture is denoted as  $\pi_{\tau_k, \lambda_k}(t)$ , for all  $t = t_1, \dots, t_n$ . These transition functions are supposed monotonically varying from 0 to 1. They are fully parameterized by a location parameter  $\tau_k$  and a smoothing parameter  $\lambda_k > 0$ . In this work, these parameters are defined as following

$$\tau_k \in \{t_1, \dots, t_n\} \quad \text{such that} \quad \pi_{\tau_k, \lambda_k}(\tau_k) = \frac{1}{2} \quad (1)$$

$$\lambda_k \in \mathbb{R}_+^* \quad \text{such that} \quad \left. \frac{d\pi_{\tau_k, \lambda_k}(t)}{dt} \right|_{\tau_k} = \frac{1}{4\lambda_k}. \quad (2)$$

Small values of  $\lambda_k$  indicates an abrupt change, while as  $\lambda_k$  increases, the rupture becomes smoother. Definition (2) comes from the natural parameterization of the sigmoid transition function

$$\pi_{\tau_k, \lambda_k}(t) = \frac{1}{1 + e^{-\frac{t - \tau_k}{\lambda_k}}}. \quad (3)$$

The proposed framework can be extended to any family of functions matching the previous constraints. For the sake of simplicity, the notation  $\pi_k(t)$  will stand for  $\pi_{\tau_k, \lambda_k}(t)$ .

Let  $K$  be the number of segments, i.e the number of different polynomial models. The rupture configuration is then given by the parameters  $(\tau_k, \lambda_k)$  for  $k = 1, \dots, (K - 1)$ . For each segments,  $p_k(t)$  denotes a polynomial of order  $q$  whose coefficients are  $\boldsymbol{\beta}_k = (\beta_{k,0}, \beta_{k,1}, \dots, \beta_{k,q})$ :

$$p_k(t) = \beta_{k,0} + \beta_{k,1}t + \beta_{k,2}t^2 + \dots + \beta_{k,q}t^q. \quad (4)$$

Throughout this paper, the polynomial order  $q$  is assumed to be fixed and known. Then, the signal mean on segment  $k = 2, \dots, K - 1$  expresses as

$$m_k(t) = [1 - \pi_{k-1}(t)] p_{k-1}(t) + [\pi_{k-1}(t) - \pi_k(t)] p_k(t) + \pi_k(t) p_{k+1}(t), \quad (5)$$



where  $I_k$  is the identity matrix of size  $(\tau_k - \tau_{k-1})$ ,  $\Pi_k^l$  is the diagonal matrix generated by the vector  $(\pi_k(\tau_{l-1}), \dots, \pi_k(\tau_{l+1}))^T$ , and  $T_k$  is defined as

$$T_k = \begin{bmatrix} 1 & \tau_{k-1} & \tau_{k-1}^2 & \cdots & \tau_{k-1}^q \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (\tau_{k-1} - 1) & (\tau_{k-1} - 1)^2 & \cdots & (\tau_{k-1} - 1)^q \end{bmatrix}. \quad (15)$$

The expression of the signal mean  $\mathbf{m}$  is finally obtained as

$$\mathbf{m} = Z\boldsymbol{\beta}. \quad (16)$$

Note that in the case of step transition function (i.e, when  $\lambda \rightarrow 0$ ), the matrix  $Z$  becomes block diagonal as in [10],[2]. Using smooth transition function, upper and lower diagonal blocks appear as the mean  $m_k$  depends also on  $p_{k-1}$  and  $p_{k+1}$ .

Equation (16) underlines the linear dependence between  $\mathbf{m}$  and  $\boldsymbol{\beta}$ . Consequently the maximum likelihood estimator of the polynomial coefficient  $\hat{\boldsymbol{\beta}}_{ML}$  conditionally to  $(\boldsymbol{\tau}, \boldsymbol{\lambda})$  is the classical linear least mean square estimator

$$\hat{\boldsymbol{\beta}}_{ML} = (Z^T Z)^{-1} Z^T \mathbf{x}, \quad (17)$$

where  $\mathbf{x} = (x_1, \dots, x_n)^T$  denotes the the observed time series.

Finally, the matrix expression of the likelihood yields

$$\mathcal{L}(\mathbf{x}|\theta) = \left( \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \right) e^{-\frac{1}{2\sigma^2}(\mathbf{x}-Z\boldsymbol{\beta})^T(\mathbf{x}-Z\boldsymbol{\beta})}. \quad (18)$$

If one deals with a fixed model order  $K$ , the estimation of the parameter  $\theta$  can simply be performed by iterative maximization of the likelihood function (18). However, this method is not robust with regard to the initial guess, and is unsuitable if the estimation of  $K$  must be performed jointly with the other parameters. In this case, some priors defined on  $K$  have to be defined to regularize this ill-posed problem. This paper proposes to use a hierarchical Bayesian framework to achieve the parameter estimation.

#### 2.4 Hierarchical Bayesian Model

Inspired by [2], a standard reparameterization of the model is done by introducing indicator variables  $r_i$  for  $i = 1, \dots, n$  such as:

$$r_i = \begin{cases} 1 & \text{if } \exists \tau_k = t_i, \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

Then the parameter vector  $\mathbf{r} = [r_1, \dots, r_n]^T$  replaces the parameters  $(\boldsymbol{\tau}, K)$  and  $\theta$  becomes :

$$\theta = (\mathbf{r}, \boldsymbol{\lambda}, \boldsymbol{\beta}, \sigma^2)^T. \quad (20)$$

By convention, first and last indicators variables are fixed to one, i.e.  $r_1 = r_n = 1$ . Thus,  $K(\mathbf{r}) = (\sum_{i=1}^n r_i) - 1$



## 2.5 Parameter Prior

In a Bayesian framework, a prior distribution must be set on each parameter. In order to keep a data-driven estimation of parameters, this prior should not be too informative.

**Indicators variables** The changes in the model are detected via the indicators variables  $r_i$ , for  $i = 1, \dots, n$ . These variables  $r_i$  are chosen to be mutually independent and distributed according to a Bernoulli distribution of parameter  $p$ :

$$r_i \sim \text{Bernoulli}(p) \quad \forall i = 1, \dots, n. \quad (21)$$

**Variance and polynomial coefficients** The  $\boldsymbol{\beta}$  and  $\sigma^2$  prior distribution are chosen among a so-called *conjugate prior* distribution with respect to the model likelihood (18). This particular choice of distribution is very convenient as it allows one to integrate these parameters in the posterior distribution.

The conjugate prior for the variance parameter with respect to the likelihood is an inverse-gamma distribution. That is:

$$\sigma^2 | (\nu, \rho) \sim \mathcal{IG} \left( \frac{\nu}{2}, \frac{\rho}{2} \right), \quad (22)$$

where  $\nu$  is a shape hyperparameter and  $\rho$  a scale hyperparameter. In the following,  $\nu$  is set to  $\nu = 2$ . Similarly, the conjugate prior for the polynomial coefficients  $\boldsymbol{\beta}$  is a Gaussian distribution:

$$\boldsymbol{\beta} | (\mathbf{r}, \boldsymbol{\lambda}, \sigma^2, \delta_0^2) \sim \mathcal{N} \left( 0, \delta_0^2 \sigma^2 (Z^T Z)^{-1} \right), \quad (23)$$

A consequence of this specific parameterization of the Gaussian distribution is that the posterior distribution is invariant from the choice of the polynomial time basis (15). Moreover, it is interesting to see that the parameter  $\delta_0^2$  corresponds to the signal-to-noise ratio since  $\delta_0^2 = \frac{E[\mathbf{m}^T \mathbf{m}]}{n\sigma^2}$ .

**Smoothing parameters** As the dependence of the likelihood with  $\boldsymbol{\lambda}$  is non-linear and non-standard, one cannot find a conjugate-prior distribution for these parameters. Then, regardless to the chosen prior distribution,  $\boldsymbol{\lambda}$  cannot be integrate in the posterior distribution. Few prior information is available for this parameter. It is defined on  $\mathbb{R}^+$ , and the probability it takes large values should be low. Indeed, as mention in [9], if  $\lambda_i \rightarrow \infty$ , an identification problem for the polynomial  $p_i$  appears.

In order to tackle this problem, a higher bound on  $\boldsymbol{\lambda}$ , denoted as  $\lambda_{max}$ , is introduced. This bound is defined by the higher acceptable value of  $\boldsymbol{\lambda}$ , i.e in the situation where there is only one rupture on the middle of the signal:  $\pi_{\frac{n}{2}, \lambda_{max}}(0) = \alpha$ ,  $\alpha$  being some arbitrary low threshold (in this work,  $\alpha = 0.01$ ).

Symmetrically, a lower bound is computed such that  $\pi_{i+1, \lambda_{min}}(i) = \alpha$ . Finally a truncated inverse-gamma prior is chosen in order to favor abrupt changes:

$$\lambda_i | (\nu', \rho') \sim \mathcal{IG}_{[\lambda_{min} \lambda_{max}]} \left( \frac{\nu'}{2}, \frac{\rho'}{2} \right) \quad \forall i = 1, \dots, n, \quad (24)$$

with  $\nu' = 2$ . It is important to note that a smoothing parameter  $\lambda_i$  is introduced for each indicator variable  $r_i$  regardless of the value of  $r_i$ , i.e regardless of the presence of a rupture.

## 2.6 Hyperparameter prior

Let  $\phi$  be the vector of the prior distribution parameters:  $\phi = (p, \rho, \delta_0^2, \rho')^T$ .

Rather than using fixed deterministic values for the prior parameters, these parameters are supposed to be distributed according to some hyperprior distributions. That is, a second level of Bayesian modelling is introduced leading to a *hierarchical* Bayesian model. Since no specific knowledge about how hyperparameters should be distributed is available, non-informative prior distributions are introduced. One advantage of such a hierarchical model is that the overall model is more robust than standard Bayesian models [10].

For the parameter  $p$ , a conjugate beta distribution is chosen:

$$p | (\Omega, \Omega') \sim \beta(\Omega, \Omega'). \quad (25)$$

Setting  $\Omega = \Omega' = 1$  ensures that the prior is a uniform distribution.

For the parameter  $\rho$ , a non-informative improper Jeffrey's prior is used:

$$f(\rho) = \frac{1}{\rho} \mathbb{I}_{\mathbb{R}^+}(\rho). \quad (26)$$

Finally, for  $\rho'$ , a vague conjugate gamma distribution is used :

$$\rho' \sim \mathcal{G}(\chi, \psi) \quad (27)$$

with  $\chi = 1$  and a large value of  $\psi$  (basically,  $\psi = 100$ ).

A prior distribution could also be introduced for the hyperparameter  $\delta_0^2$ . However, a fixed deterministic value is preferred in our application. Indeed,  $\delta_0^2$  represents the signal-to-noise ratio (SNR) of the observed signal. Therefore, its value controls the precision of the signal segmentation. Thus,  $\delta_0^2$  must be fixed to the desired SNR to be reached in our segmentation algorithm.

## 2.7 Joint Posterior Distribution

Within this hierarchical Bayesian framework, the posterior distribution of the parameters  $\theta$  becomes:

$$f(\theta|\mathbf{x}) = \int f(\theta, \phi)d\phi \propto \int \mathcal{L}(\mathbf{x}|\theta)f(\theta|\phi)f(\phi)d\phi, \quad (28)$$

where  $f(\phi)$  expresses as:

$$f(\phi) \propto \left(p^{\Omega-1}(1-p)^{\Omega'-1}\right) \left(\frac{1}{\rho}\right) \left((\rho')^{\chi-1}e^{-\frac{\rho'}{\psi}}\right), \quad (29)$$

and  $f(\theta|\phi)$  is:

$$\begin{aligned} f(\theta|\phi) &\propto \left(p^{K(\mathbf{r})}(1-p)^{n-K(\mathbf{r})}\right) \left(\frac{\rho^{\frac{\nu}{2}}e^{-\frac{\rho}{2\sigma^2}}}{(\sigma^2)^{1+\frac{\nu}{2}}}\right) \\ &\times \left(\prod_{i=2}^{n-1} \frac{\rho^{\frac{\nu'}{2}}e^{-\frac{\rho'}{2\lambda_i}}}{\lambda_i^{1+\frac{\nu'}{2}}}\mathbb{I}_{[\lambda_{min}\lambda_{max}]}(\lambda_i)\right) \left(\frac{e^{-\frac{1}{2\delta_0^2\sigma^2}\boldsymbol{\beta}^T Z^T Z \boldsymbol{\beta}}}{(2\pi\delta_0^2\sigma^2)^{\frac{K(\mathbf{r})(q+1)}{2}}|Z^T Z|^{-\frac{1}{2}}}\right). \end{aligned} \quad (30)$$

## 2.8 Nuisance parameter integration

The choice of conjugate prior allows one to integrate out the parameters  $\sigma^2$  and  $\boldsymbol{\beta}$ , and the hyperparameters  $p$  and  $\rho'$  in the expression of the joint posterior distribution  $f(\theta, \phi|\mathbf{x})$ .

After integration of these parameters, the following marginal posterior is obtained:

$$\begin{aligned} f(\mathbf{r}, \boldsymbol{\lambda}, \rho, \delta_0^2|\mathbf{x}) &\propto \\ &\left(\frac{\prod_{i=2}^{n-1} \lambda_i^{-(1+\frac{\nu'}{2})}}{\left[\frac{1}{\psi} + \sum_{i=1}^{n-1} \frac{1}{2\lambda_i}\right]^{(\chi+(n-2)\frac{\nu'}{2})}}\right) \left(\frac{\Gamma(K(\mathbf{r})+\Omega)\Gamma(n-2-K(\mathbf{r})+\Omega')(\delta_0^2+1)^{-\frac{K(\mathbf{r})(q+1)}{2}}}{\left(\frac{\rho}{2} + \frac{1}{2}\left[\mathbf{x}^T \mathbf{x} - \frac{\delta_0^2}{1+\delta_0^2}(\mathbf{x}^T Z[Z^T Z]^{-1}Z^T \mathbf{x})\right]\right)^{\frac{n+\nu}{2}}}\right) \end{aligned} \quad (31)$$

Obviously, this posterior distribution is too complex to permit closed-form calculation of standard Bayesian estimators (namely, the maximum a posteriori (MAP) or the minimum mean square error (MMSE) estimators) for the parameters of interest  $(\mathbf{r}, \boldsymbol{\lambda})$ . A classical solution to tackle this problem is to use a Markov Chain Monte-Carlo (MCMC) method to generate samples asymptotically distributed according to the posterior distribution (31). Estimators can then be derived from their empirical posterior distribution.

As mention in [2], the advantage of choosing indicators variable  $\mathbf{r}$  instead of natural parameters  $(\boldsymbol{\tau}, K)$  is to constrain the parameter vector of the problem to live in a fixed dimensional space. Then the simple Gibbs sampler can be used to generate samples instead of more complex Reversible Jump Markov Chain Monte-Carlo samplers [6]. Note, however, that, using this parameterization, the number of variables to sample is increased since the couple  $(r_i, \lambda_i)$  must be sampled for each time instant  $i = 1, \dots, n$ , regardless if there is a rupture.

### 3 Gibbs sampler for parameters estimation

The Gibbs sampler is a MCMC sampling strategy introduced in [5]. It consists of successive scans of the parameter vector. For each scan, each parameter is sampled according to its posterior distribution conditionally on all other variables. By this mean, all the simulation steps are univariate although the global problem is multivariate. So, in order to perform Gibbs sampling, one needs to be able to sample from each of these univariate conditional posterior distributions.

#### 3.1 Conditional Posterior distribution

The following conditional posterior distributions are derived from the expression of  $f(\theta, \phi)$ . For the parameter  $\sigma^2$ , the posterior is obtain after integration of  $\beta$  over  $\mathbb{R}^{q+1}$ :

$$\rho|\sigma^2 \sim \mathcal{G}(1, 2\sigma^2), \quad (32)$$

$$\beta|\mathbf{r}, \boldsymbol{\lambda}, \mathbf{x}, \delta_0^2 \sim \mathcal{N}\left(\frac{\delta_0^2}{1 + \delta_0^2}(Z^T Z)^{-1} Z^T \mathbf{x}, \frac{\sigma^2 \delta_0^2}{1 + \delta_0^2}(Z^T Z)^{-1}\right), \quad (33)$$

$$\sigma^2|\mathbf{r}, \boldsymbol{\lambda}, \rho, \delta_0^2, \mathbf{x} \sim \mathcal{IG}\left(\frac{n + \nu}{2}, \frac{\rho}{2} + \frac{1}{2}\left[\mathbf{x}^T \mathbf{x} - \frac{\delta_0^2}{1 + \delta_0^2}(\mathbf{x}^T Z[Z^T Z]^{-1} Z^T \mathbf{x})\right]\right) \quad (34)$$

**Indicators variables conditional posterior** Let  $\tilde{\mathbf{r}}_i$  be the vector  $\mathbf{r}$  where the  $i$ th element has been removed:  $\tilde{\mathbf{r}}_i = (r_1, \dots, r_{i-1}, r_{i+1}, \dots, r_n)^T$ . Although the distribution of  $r_i|\tilde{\mathbf{r}}_i, \boldsymbol{\lambda}, \delta_0^2, \rho$  is not in a standard form, its posterior distribution can be easily obtained since  $r_i$  is a binary variable. Indeed, to generate samples distributed according to this posterior distribution, one just need to compute the value of the joint posterior distribution (31) for  $r_i = 0$  and  $r_i = 1$  with all other parameter fixed. After normalization  $r_i$  is sampled from a Bernoulli distribution:

$$\begin{aligned} f(r_i = 0) &= f(r_i = 0, \tilde{\mathbf{r}}_i, \boldsymbol{\lambda}, \rho, \delta_0^2|\mathbf{x}), \\ f(r_i = 1) &= f(r_i = 1, \tilde{\mathbf{r}}_i, \boldsymbol{\lambda}, \rho, \delta_0^2|\mathbf{x}), \\ p_{r_i} &= \frac{f(r_i=1)}{f(r_i=0)+f(r_i=1)}, \\ r_i &\sim \text{Bernoulli}(p_{r_i}). \end{aligned} \quad (35)$$

**Smoothing parameter conditional posterior** The last distribution needed for completing a Gibbs scan is the conditional posterior of each parameter  $\lambda_i$ . Unfortunately, these parameters are not distributed according to a standard probability distribution. To tackle this problem, a Metropolis-Hastings within Gibbs algorithm is used. The Metropolis-Hastings algorithm is a special case of acceptance-rejection algorithm. It is used to generate samples from a distribution whose probability density function is known analytically, up to a constant. Metropolis-Hastings algorithm requires a proposition distribution  $q(\lambda)$  hoped to be close enough from real distribution  $f(\lambda)$ . Then, a sample  $\tilde{\lambda}$  is generated from this proposition distribution and is accepted or not according to the following the procedure:

$$\lambda_i^m = \begin{cases} \tilde{\lambda} & \text{if } u < \frac{f(\tilde{\lambda}|\boldsymbol{\lambda}/\lambda_i^{m-1}, \mathbf{r}, \delta_0^2, \rho)q(\lambda_i^{m-1}|\boldsymbol{\lambda}/\lambda_i^{m-1}, \mathbf{r}, \delta_0^2, \rho)}{f(\lambda_i^{m-1}|\boldsymbol{\lambda}/\lambda_i^{m-1}, \mathbf{r}, \delta_0^2, \rho)q(\tilde{\lambda}|\boldsymbol{\lambda}/\lambda_i^{m-1}, \mathbf{r}, \delta_0^2, \rho)} \\ \lambda_i^{m-1} & \text{otherwise,} \end{cases} \quad (36)$$

where  $u \sim \mathcal{U}(0, 1)$ ,  $\lambda_i^m$  is the value of the smoothing parameter at time instant  $t_i$  and Gibbs scan  $m$ . This procedure ensures that the sample  $\lambda_i^m$  is asymptotically distributed according to  $f(\lambda_i | \boldsymbol{\lambda} / \lambda_i, \mathbf{r}, \delta_0^2, \rho)$ .

In order to increase the efficiency of the Metropolis-Hastings step, one need to choose  $q(\lambda)$  as close as possible from the conditional posterior distribution of  $\lambda_i$ . A Gaussian distribution is used, with mean  $\lambda_p$  and variance  $\sigma_p^2$ . These values are computed at each time instant  $i$ , and for each rupture configuration  $\mathbf{r}$ , by performing a local regression of the normalized observations (between 0 and 1) with respect to the transition function  $\pi_{i,\lambda}$ . The regression estimation is done via a linear least mean square procedure and the the smoothness parameter of the fitted transition function parameter gives  $\lambda_p$ , whereas the variance of the residual gives  $s_p^2$ .

### 3.2 Gibbs sampling algorithm

Finally, the Gibbs algorithm can be decomposed as follow:

1. for each time index  $i = 1, \dots, n$  of the time series  $\mathbf{x}$ 
  - compute proposition distribution parameter  $\mu_p$  and  $\sigma_p^2$
  - sample  $\tilde{\lambda} \sim \mathcal{N}(\mu_p, \sigma_p^2)$
  - accept/reject  $\tilde{\lambda}$  following (36)
  - sample  $r_i$  following (35)
2. sample  $\sigma^2$  following (34)
3. sample  $\boldsymbol{\beta}$  following (33)
4. sample  $\rho$  following (32)

## 4 Simulations

### 4.1 Electrical transient data

The algorithm presented in this paper has been tested on both synthetic and real electrical transient data. The ‘‘ON’’ event transients correspond to the observed signal while an electrical appliance is turned-on on a domestic electrical network. They are characterized by periodic, slow-varying envelope current. The fundamental frequency is dependent of the utility frequency of the country electrical network and is 50 Hz in France. As this frequency is known, the envelope can be easily extracted from the signal. These envelopes are then segmented by the Gibbs algorithm. Figures 1 and 2 show respectively real data from a refrigerator ON event transient and a Hoover ON transient event.

### 4.2 Synthetic Data

The proposed segmentation algorithm has been first tested on a synthetic signal built according to the studied model. The values of the parameters were chosen to reflect those of the real refrigerator signals displayed in Fig. 1. Figure 3 displays

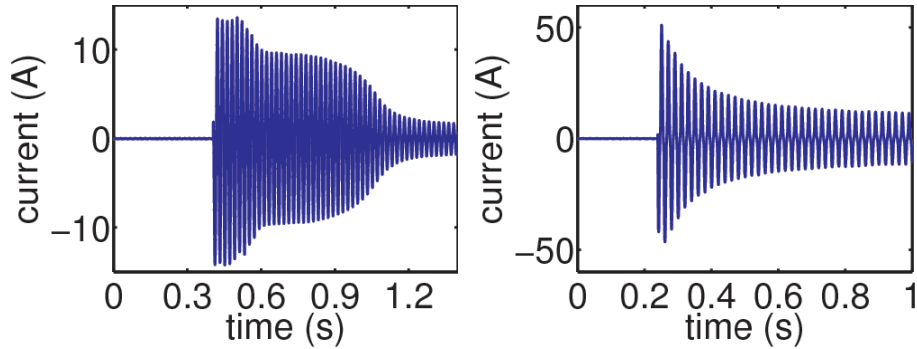


Figure 1: Refrigerator ON event transient      Figure 2: Hoover ON event transient

the marginal posteriors of the parameters of interest. These empirical posteriors are obtained from the samples generated by the MCMC method. The number of Gibbs scans that has been performed in these simulations is  $N = 10000$ . A fixed SNR parameter was chosen:  $\delta_0^2 = 50$ . The true parameter value, denoted by a tilde, the MAP and MMSE estimates of the parameters are reported in Table 1. Despite a small bias on the smoothness parameters, one can see that the obtained Bayesian estimates are close to the true values. In particular, the Bayesian estimates of the rupture locations are in good agreement with the true values.

### 4.3 Real Data

Simulations on real electrical transient data has been performed. These real data are gathered by Electricité de France (EDF) in some houses. The data used in this paper are the transient observed for only one appliance. The experimental setup is the same that the one introduced for synthetic data ( $N = 10000$  Gibbs scans,  $\delta_0^2 = 50$ ). The maximum likelihood (ML) values of the parameters are used as reference to appreciate the estimation results. These ML values are obtained by a direct numerical maximization of the likelihood for a fixed number of segments. The results of the algorithm applied to a real refrigerator transient signal are displayed and reported in Figure 4 and Table 2 respectively. Similarly, the results of the algorithm applied to a real hoover transient signal are displayed and reported in Figure 5 and Table 3 respectively. In the last simulation the following exponential transition function is used:

$$\pi_{\tau,\lambda} = \max \left( 1 - \frac{1}{2} e^{-\frac{(x-\tau)}{2\lambda}}, 0 \right). \quad (37)$$

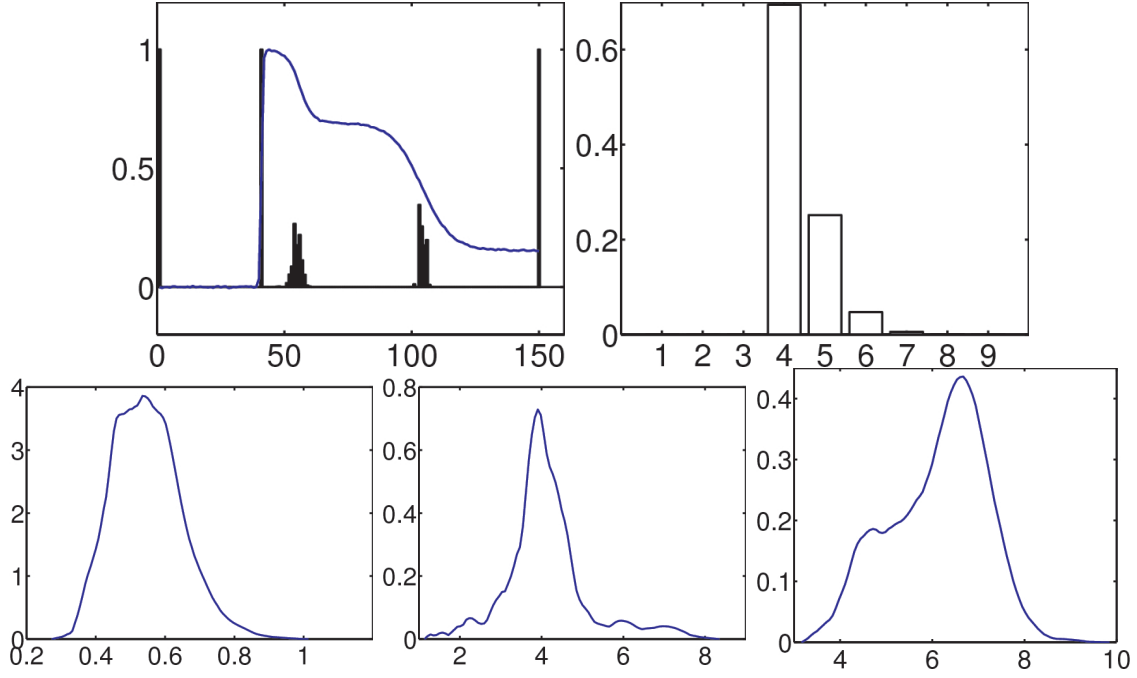
As in the synthetic case, the Bayesian estimates appears to be slightly biased with respect to the Maximum Likelihood estimates. However, one can see that these Bayesian estimates are in good agreement with the Maximum Likelihood estimates.

## 5 Conclusions

A smooth transition regression method, based on a hierarchical Bayesian model, has been studied. This method has been tested on both synthetic and real transients

$\tilde{\tau}$	(41 56 104)	$\tilde{\lambda}$	(0.3 2.5 5.8)
$\tau_{MAP}$	(41 56 103)	$\lambda_{MAP}$	(0.5 3.9 6.6)
$\tau_{MMSE}$	(41 54.9 104.2)	$\lambda_{MMSE}$	(0.5 4.1 6.6)

Table 1: Synthetic signal simulation results

Figure 3: Synthetic transient modelization : upper-left :  $(\mathbf{r}|K_{MAP})$  posterior distribution - upper-right :  $K$  posterior distribution - lower-left :  $\lambda_1$  posterior distribution - lower-middle :  $\lambda_2$  posterior distribution - lower-right :  $\lambda_3$  posterior distribution

$\tau_{ML}$	(41 56 104)	$\lambda_{ML}$	(0.3 2.5 5.8)
$\tau_{MAP}$	(41 55 105)	$\lambda_{MAP}$	(0.6 3.7 6.0)
$\tau_{MMSE}$	(41 51.9 104.2)	$\lambda_{MMSE}$	(0.6 3.8 5.5)

Table 2: Refrigerator signal simulation results

$\tau_{ML}$	(24 30)	$\lambda_{ML}$	(0.3 6)
$\tau_{MAP}$	(24 33)	$\lambda_{MAP}$	(0.4 5.7)
$\tau_{MMSE}$	(24 33)	$\lambda_{MMSE}$	(0.4 5.7)

Table 3: Hoover signal simulation results

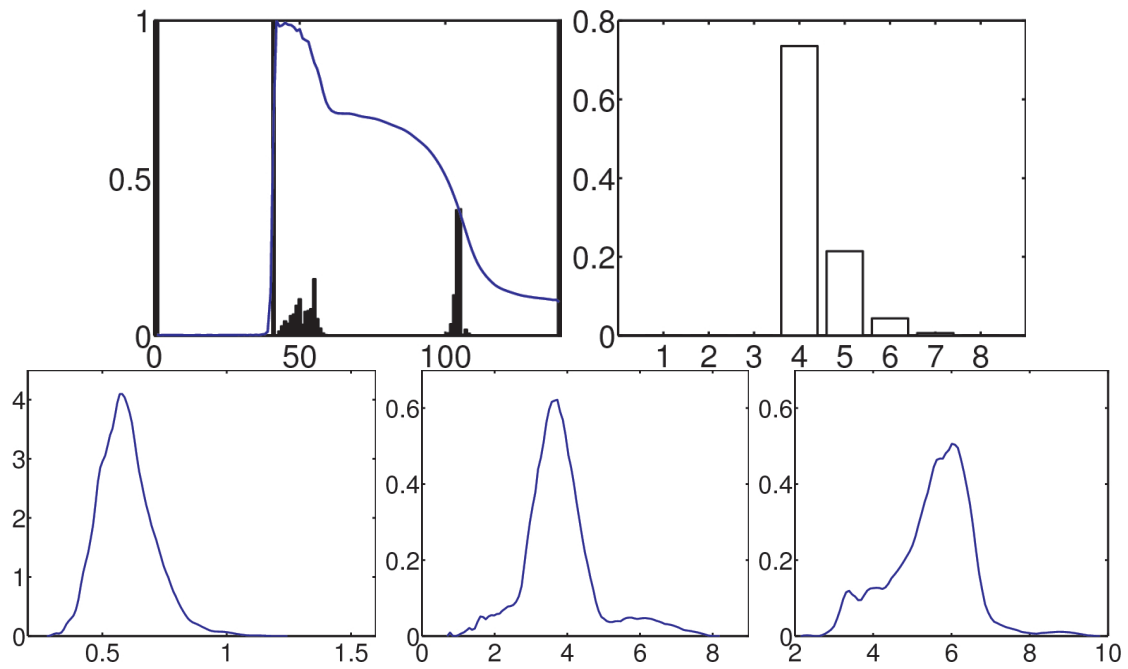


Figure 4: Refrigerator transient modelization : upper-left :  $(\mathbf{r}|K_{MAP})$  posterior distribution - upper-right :  $K$  posterior distribution - lower-left :  $\lambda_1$  posterior distribution - lower-middle :  $\lambda_2$  posterior distribution - lower-right :  $\lambda_3$  posterior distribution

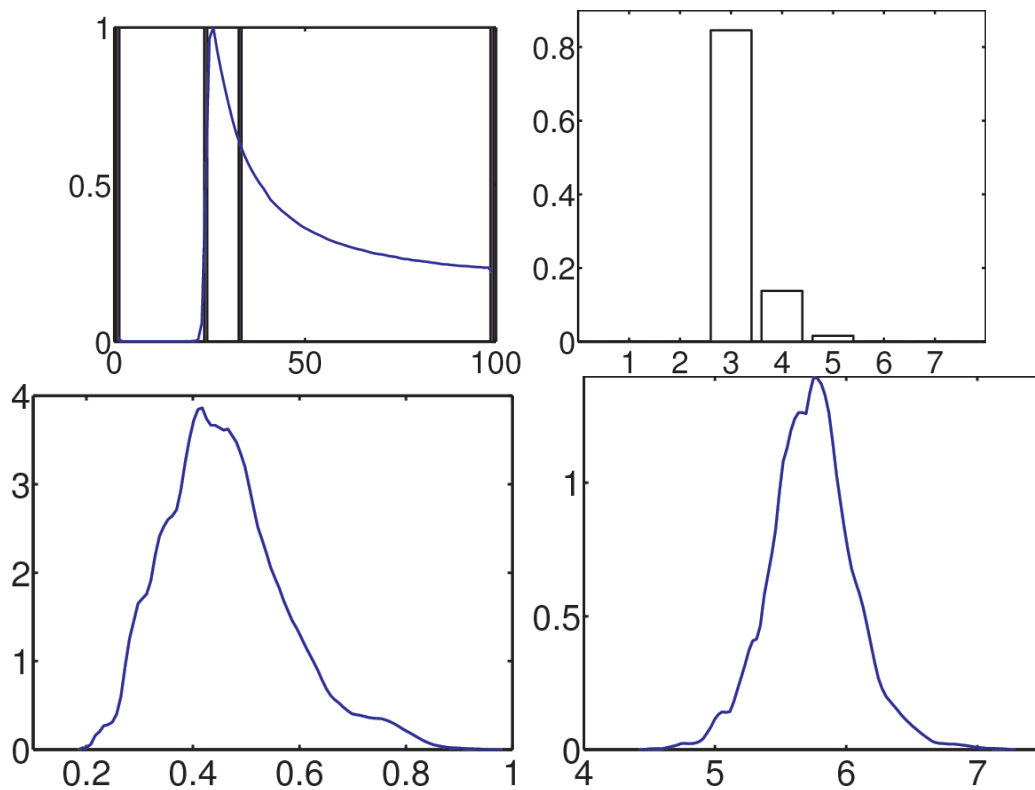


Figure 5: Hoover transient modelization : upper-left :  $(\mathbf{r}|K_{MAP})$  posterior distribution - upper-right :  $K$  posterior distribution - lower-left :  $\lambda_1$  posterior distribution - lower-right :  $\lambda_2$  posterior distribution



generated by different electrical appliances, in a context of residential non-intrusive load monitoring. It allows one to obtain a parsimonious representation of the transients that characterize the electrical appliances. The extension to other classes of electrical devices, and the learning of the parameters in order to detect and classify these different classes, are currently under investigation.

## References

- [1] F. Chamrouki, S. Allou, and P. Aknin. A probabilistic approach for the classification of railway switch operating states. In *Sixth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies. CM 2009 and MFPT 2009*, Dublin Ireland.
- [2] N. Dobigeon and J.-Y. Tourneret. Joint segmentation of piecewise autoregressive processes by using a hierarchical model and a bayesian sampling approach. *IEEE Transactions On Signal Processing*, 55(4), 2007.
- [3] J.-B. Durand, L. Bozzi, G. Celeux, and C. Derquenne. Analyse de courbes de consommation électrique par chaine de markov cachées. *Revue de Statistique Appliquée*, 12(4):71–91, 2004.
- [4] M. El Guedri, J. Bect, C. Lajaunie, G. Fleury, R. Seraoui, T. Garcia, and A. Girard. RJMCMC Point Process Sampler For Single Sensor Source Separation: An Application to Electric Load Monitoring. In *European Signal Processing Conference 2009*, Glasgow, Scotland.
- [5] A. E. Gelfand and A. F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409, 1990.
- [6] P. J. Green. Reversible Jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 52:711–732, 1995.
- [7] G. W. Hart. Nonintrusive Appliance Load Monitoring. *Proceedings of the IEEE*, pages 1870–1891, 1992.
- [8] S. B. Leeb. Transient Event Detection in Spectral Envelope Estimates for Nonintrusive Load Monitoring. *IEEE Transactions on Power Delivery*, 10(3), 1995.
- [9] M. Lubrano. *Bayesian Analysis of Nonlinear Time Series Models with a Threshold*, pages 79–118. Cambridge University Press, 2000.
- [10] E. Punskeya, C. Andrieu, A. Doucet, and F. William. Bayesian Curve Fitting Using MCMC With Application to Signal Segmentation. *IEEE Transactions On Signal Processing*, 50(3):747–758, 2002.