



**HAL**  
open science

## A reversible jump MCMC algorithm for Bayesian curve fitting 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. A reversible jump MCMC algorithm for Bayesian curve fitting by using smooth transition regression models. ICASSP 2011 - IEEE International Conference on Acoustics, Speech and Signal Processing, May 2011, Prague, Czech Republic. pp.s/n. hal-00609133

**HAL Id: hal-00609133**

**<https://hal.science/hal-00609133>**

Submitted on 18 Jul 2011

**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.

# A REVERSIBLE JUMP MCMC ALGORITHM FOR BAYESIAN CURVE FITTING BY USING SMOOTH TRANSITION REGRESSION MODELS

Matthieu Sanquer<sup>\*</sup>      Florent Chatelain<sup>\*</sup>      Mabrouka El-Guedri<sup>†</sup>      Nadine Martin<sup>\*</sup>

<sup>\*</sup> University of Grenoble, GIPSA-lab, 961 rue de la Houille Blanche, BP 46, 38402, St Martin D'Hères, France

<sup>†</sup> Electricité de France – R&D, 4 Quai Watier, 78401 Chatou, France

## ABSTRACT

This paper proposes a Bayesian algorithm to estimate the parameters of a smooth transition regression model. Within this modelling, time series are divided into segments and a linear regression analysis is performed on each segment. Unlike piecewise regression model, smooth transition functions are introduced to model smooth transitions between the sub-models. Appropriate prior distributions are associated with each parameter to penalize a data-driven criterion, leading to a fully Bayesian model. Then, a reversible jump Markov Chain Monte Carlo algorithm is derived to sample the parameter posterior distributions. It allows one to compute standard Bayesian estimators, providing a sparse representation of the data. Results are obtained for real-world electrical transients with a view to non-intrusive load monitoring applications.

**Index Terms**— Bayesian segmentation, Smooth transition regression, Reversible Jump MCMC, Non Intrusive Load Monitoring

## 1. INTRODUCTION

Piecewise regression models are a flexible class of regression models that has been widely used for abrupt changes detection [1], or segmentation [2, 3] problems. Within this modelling, time series are divided into several segments. Each segment is associated with a specific regime that appears during a time interval. A classical regression analysis is independently performed on each segment. Thus, piecewise regression analysis often reduces to a changepoint detection problem.

However, in many applications, the change among two regimes is not abrupt but rather smooth and classical piecewise regression methods may produce oversegmented representation of the signal. To avoid this problem, some parametric transition functions are considered for modelling the smooth switch between two regimes. These transition functions are introduced to continuously mix the different regression models associated with neighboring segments. Some similar models have already been studied in econometrics [4, 5], or in railway monitoring applications [6] (note that in this case, the smooth switch process is hidden). However these approaches require the number of regimes to be fixed a priori.

When the number of ruptures is unknown, the estimation of the model parameters is a difficult problem. Indeed, it appears to be ill-posed and requires some regularization to be solved. To circumvent these difficulties, a fully Bayesian framework is introduced. Appropriate priors are associated with each parameter in order to penalize a likelihood-based criterion. As the resulting posterior density appears to be too complex to allow an analytic computation, a reversible jump Markov Chain Monte-Carlo (RJCMCMC) sampling method [7] is investigated. Classical Bayesian estimators are then derived from

the MCMC sample, leading to a quite sparse modelling of the original time series. The RJCMCMC algorithm is eventually used to provide an accurate and sparse representation of real-world electrical transient data.

This communication is organized as follows. In section 2, the smooth transition regression model is introduced. Section 3 presents the proposed hierarchical Bayesian framework. In section 4, the RJCMCMC algorithm is derived for parameters estimation. Results on real-world electrical transients are reported and discussed in 5. Conclusions are drawn in section 6.

## 2. SMOOTH TRANSITION REGRESSION MODEL

### 2.1. Model and notations

A sequence of  $n$  samples  $\mathbf{x} = (x_i)_{i=1,\dots,n}$  is considered. The time instant sequence associated with the observations  $\mathbf{x}$  is denoted as  $\mathbf{t} = (t_i)_{i=1,\dots,n}$ , while  $t_s$  is the sampling period. Let  $\tau_1 < \tau_2 < \dots < \tau_{K-1}$  be a configuration of  $K - 1$  changepoint locations in the signal,  $K$  being the number of segments. For notational convenience, the convention that  $\tau_0 = t_0$  and  $\tau_K = t_n$  is adopted. In this work, a polynomial regression analysis is performed on each segment. More precisely, the mean value at time  $t_i$  of the regime associated with the  $k$ th segment  $[\tau_{k-1} \tau_k]$  is

$$m_{i,k} = \sum_{0 \leq p \leq P} \beta_k^{(p)} t_i^p, \quad \forall 1 \leq i \leq n, \quad (1)$$

where  $P$  is the order of the polynomial, and  $\beta_k^{(0)}, \dots, \beta_k^{(P)}$  denotes the polynomial coefficients associated with the  $k$ th segment  $[\tau_{k-1} \tau_k]$ .

A transition function denoted as  $\pi_{\eta_k}(t)$  is associated with the  $k$ th rupture  $\tau_k$ , for all  $0 < k < K$ . This function is chosen to be monotonically increasing from 0 to 1 and is governed by the vector of parameters  $\eta_k$  i.e the rupture locations  $\tau_k$ , and possibly other kind of parameters that characterize the smoothness, or the shape (i.e. sigmoidal, exponential,...) of the rupture. By convention, the transitions associated with the first and last ruptures  $\tau_0$  and  $\tau_K$  are the step functions such that  $\pi_{\eta_0}(t) = 1$  and  $\pi_{\eta_K}(t) = 0$  for all  $t_0 \leq t \leq t_n$ . It yields the following expression of the observation at time  $t_i$

$$x_i = \sum_{1 \leq k \leq K} [\pi_{\eta_{k-1}}(t_i) - \pi_{\eta_k}(t_i)] m_{i,k} + \epsilon_i, \quad (2)$$

for all  $1 \leq i \leq n$ , where  $\pi_{\eta_{k-1}}(t_i)$  accounts for the transition from  $m_{i,k-1}$  to  $m_{i,k}$  and  $\pi_{\eta_k}(t_i)$  accounts for the transition from  $m_{i,k}$  to  $m_{i,k+1}$  (see figure 1(a)). The noise vector  $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$  is composed of i.i.d. centered Gaussian random variables with variance  $\sigma^2$ . Let  $\boldsymbol{\eta} = [\eta_1, \dots, \eta_{K-1}]^T$  be

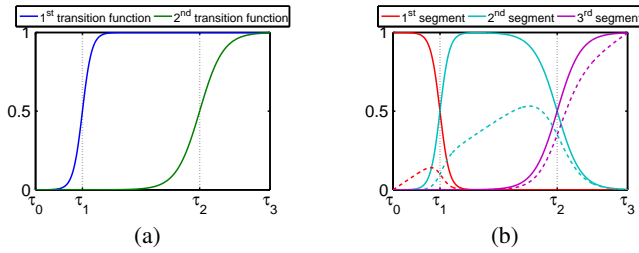
the parameter vector of the whole configuration of transitions, while  $\beta = [\beta_1^{(0)}, \dots, \beta_1^{(P)}, \dots, \beta_K^{(0)}, \dots, \beta_K^{(P)}]^T$  denotes the  $K(P+1)$ -length column vector of the regression coefficient parameters. An equivalent matrix formulation of the observation model (2) is given by

$$\mathbf{x} = Z(\boldsymbol{\eta})\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (3)$$

where  $Z(\boldsymbol{\eta})$  is a  $n \times K(P+1)$  matrix of regressors whose entries are the

$$Z_{i,(k-1)(P+1)+p+1} = [\pi_{\eta_{k-1}}(t_i) - \pi_{\eta_k}(t_i)] t_i^p, \quad (4)$$

for all  $k \in \{1 \dots K\}$  and  $p \in \{0 \dots P\}$ . Note that the columns of the matrix  $Z(\boldsymbol{\eta})$  represents the temporal evolution of the regressors. An example of a configuration of transition functions, and the corresponding regressors are depicted in Fig. 1.



**Fig. 1.** Configuration of transition functions and matrix of regressors  $Z(\boldsymbol{\eta})$  for  $K = 3$  segments. (a) transition functions associated with the two ruptures. (b) columns of the  $n \times K(P+1)$  matrix of regressors  $Z(\boldsymbol{\eta})$  for a first order polynomial ( $P = 1$ ) regression analysis on each segment. The same color is used for the regressors associated with the same segment  $k$ , while solid and dashed lines corresponds to the zero ( $p = 0$ ) and first ( $p = 1$ ) order monomial regressors respectively.

## 2.2. Constraint on the configuration of ruptures

Eq. (2) induces an additional assumption on the rupture configuration. Indeed, for each segment  $[\tau_{k-1} \tau_k]$ , only the transition functions associated with the  $(k-1)$ th rupture  $\tau_{k-1}$  and the  $k$ th rupture  $\tau_k$  are taken into account. Thus, all the previous transition functions must be close to 1 after  $\tau_{k-1}$ , and all the next ones must be close to 0 before  $\tau_k$ . More precisely, the following constraints are added

$$\begin{cases} 1 - \pi_{\eta_{k-2}}(\tau_{k-1}) < \varepsilon \\ \pi_{\eta_{k+1}}(\tau_k) < \varepsilon \end{cases} \quad \forall k \in [2, K-2], \quad (5)$$

where  $\varepsilon$  is some very low threshold (basically,  $\varepsilon = 10^{-10}$ ).

## 2.3. Transition function choice and parameterization

A large number of transition function  $\pi_\eta$  are possible, depending on the application. In this paper, the following class of Kohlrausch function [8] is considered

$$\pi_\eta(t) = \begin{cases} 1 - \exp[-(h(\alpha)\frac{t-\tau}{\lambda} + g(\alpha))^\alpha] & \text{for } t > \tau - \lambda\frac{g(\alpha)}{h(\alpha)} \\ 0 & \text{for } t < \tau - \lambda\frac{g(\alpha)}{h(\alpha)} \end{cases} \quad (6)$$

where

$$g(\alpha) = \begin{cases} 0 & \text{if } 0 < \alpha \leq 1 \\ (\frac{\alpha-1}{\alpha})^{\frac{1}{\alpha}} & \text{if } \alpha > 1, \end{cases} \quad (7)$$

$$h(\alpha) = \begin{cases} \frac{1}{\alpha}\Gamma(\frac{1}{\alpha}) & \text{if } 0 < \alpha \leq 1 \\ \left[ (\frac{\alpha-1}{\alpha})^{\frac{1}{\alpha}} + \frac{1}{\alpha}\Gamma(\frac{1}{\alpha}) (1 - 2P(\frac{1}{\alpha}, \frac{\alpha-1}{\alpha})) \right] & \text{if } \alpha > 1. \end{cases} \quad (8)$$

with  $P(a, x) = \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} dt$  being the incomplete gamma function. Such a family of functions is able to describe a wide range of shapes, from exponential to more symmetric functions such as sigmoids. This specific parameterization ensures that  $\tau$  and  $\lambda$  match the following definitions :

$$\tau = \operatorname{argmax}_t \frac{d\pi_\eta}{dt}(t), \quad (9)$$

$$\lambda = \int_{-\infty}^{\tau} \pi_\eta(t) dt + \int_{\tau}^{+\infty} [1 - \pi_\eta(t)] dt. \quad (10)$$

Finally, by introducing  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_{K-1})$ ,  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_{K-1})$  and  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{K-1})$ , the parameter vector associated with the configuration of transitions is  $\boldsymbol{\eta} = (\boldsymbol{\tau}, \boldsymbol{\lambda}, \boldsymbol{\alpha})$ , and the resulting vector of the model parameters is  $\boldsymbol{\theta} = (\boldsymbol{\eta}, \boldsymbol{\beta}, K, \sigma^2)$ .

## 2.4. Likelihood model

Based on eq. (3), the Gaussian assumption on the noise vector leads to the following expression of the observation likelihood

$$\mathcal{L}(\mathbf{x}|\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^n} e^{-\frac{1}{2\sigma^2}(\mathbf{x} - Z(\boldsymbol{\eta})\boldsymbol{\beta})^T(\mathbf{x} - Z(\boldsymbol{\eta})\boldsymbol{\beta})}. \quad (11)$$

When the number of segments  $K$  is given, the estimation of the parameters  $\boldsymbol{\theta}$  reduces to a classical non-linear least mean square problem. However, in the more general case where  $K$  is unknown, the problem is obviously ill-posed. In fact, a direct optimization of (11) w.r.t.  $\boldsymbol{\theta}$  leads to degenerated solutions with possibly as many parameters as observations. Following [2, 3], a hierarchical Bayesian approach is proposed to regularize the problem thanks to a fully data-driven criterion.

## 3. HIERARCHICAL BAYESIAN MODEL

In a Bayesian framework, the estimation problem is solved by assuming that  $\boldsymbol{\theta}$  is a random vector whose *a priori* distribution is known. The same framework as in [9] is followed.

### 3.1. Parameter priors

A binomial distribution is chosen for the number of segments parameter  $K$

$$f(K|\mu) = \mu^K (1 - \mu)^{n-K}, \quad (12)$$

where  $n$  is the number of samples, and  $\mu$  is an hyperparameter. For the parameters  $\boldsymbol{\beta}$  and  $\sigma^2$ , it is convenient to choose among the family of conjugate prior distributions with respect to the likelihood (11). The inverse gamma distribution for  $\sigma^2$  is

$$\sigma^2 | (\nu, \rho) \sim \mathcal{IG}\left(\frac{\nu}{2}, \frac{\rho}{2}\right), \quad (13)$$

where  $\rho$  is an hyperparameter, the inverse gamma shape parameter  $\nu$  being fixed to 2. The polynomial coefficients  $\boldsymbol{\beta}$  are characterized through the following conjugate multivariate Gaussian distribution

$$\boldsymbol{\beta} | (\boldsymbol{\eta}, \sigma^2, \delta^2) \sim \mathcal{N}\left(\mathbf{0}, \delta^2 \rho [Z(\boldsymbol{\eta})^T Z(\boldsymbol{\eta})]^{-1}\right). \quad (14)$$

The main advantage of this parameterization is that the resulting posterior distribution is now invariant to the choice of the polynomial time basis. Moreover, the hyperparameter  $\delta^2$  is then recognizable as a signal to noise ratio since

$$\delta^2 = \frac{E[(Z(\boldsymbol{\eta})\boldsymbol{\beta})^T (Z(\boldsymbol{\eta})\boldsymbol{\beta})]}{n\sigma^2}. \quad (15)$$

Unfortunately, it is not possible to choose some conjugate distributions for the parameter vectors  $\lambda$  and  $\alpha$ , as the dependence of the likelihood from the parameters is non-standard. Since  $\lambda$  characterizes the spreading times of the ruptures according to (10), a higher bound  $\lambda_{\max} = t_n - t_0$  is given. Then a truncated inverse-gamma prior is chosen

$$\lambda_k | (\nu', \omega) \sim \mathcal{IG}_{[0, \lambda_{\max}]} \left( \frac{\nu'}{2}, \frac{\omega}{2} \right) \quad \forall 0 < k < K, \quad (16)$$

where the shape parameter  $\nu'$  is fixed to 2. Moreover, the values of interest on the stretch exponential shape parameters  $\alpha_k$  can be bounded such that  $\alpha_k < \alpha_{\max}$ . Thus, the prior for the parameter  $\alpha_k$  is chosen to be uniform between 0 and  $\alpha_{\max}$ :  $\alpha_k \sim \mathcal{U}_{[0, \alpha_{\max}]}$  for all  $0 < k < K$ . In this work,  $\alpha_{\max}$  is set to 20. Eventually the hyperparameter vector is  $\phi = \{\mu, \rho, \delta^2, \omega\}$ .

### 3.2. Hyperparameter priors

A second level of hierarchy is introduced on the hyperparameter vector  $\phi$ . It consists of considering some hyperpriors on these hyperparameters. As no further information is available on these hyperparameters, a uniform distribution is chosen for  $\mu$ , while a vague gamma prior is chosen for  $\omega \sim \mathcal{G}(\chi, \psi)$  (with  $\chi = 1$  and scale parameter  $\psi = 100$ ). For the parameter  $\rho$ , an improper non-informative Jeffreys prior is introduced such that  $f(\rho) = \frac{1}{\rho} \mathbb{I}_{\mathbb{R}^+}(\rho)$ . The overall model is then a so called *hierarchical* Bayesian model.

Finally, a vague conjugate inverse-gamma distribution could be chosen for  $\delta^2$  such that  $\delta^2 \sim \mathcal{IG}(\Delta, \Xi)$ . However, a given value should be preferred in some cases. Indeed, according to equation (15),  $\delta^2$  is a signal-to-noise ratio and is then related to the reconstruction error. Setting  $\delta^2$  to a given value allows one to control the signal fitting, preventing for example the emergence of too many ruptures.

### 3.3. Joint posterior distribution

The posterior distribution of the parameters  $\theta$  expresses as

$$f(\theta | \mathbf{x}) = \int f(\theta, \phi | \mathbf{x}) d\phi \propto \int \mathcal{L}(\mathbf{x} | \theta) f(\theta | \phi) f(\phi) d\phi \quad (17)$$

In the expression of the joint posterior distribution, the parameters  $\mu$ ,  $\sigma^2$  and  $\beta$  can be integrated out, yielding the following expression:

$$f(\boldsymbol{\eta}, K, \rho, \delta^2 | \mathbf{x}) \propto \left( \frac{e^{-\frac{\Delta}{\delta^2}}}{(\delta^2)^{\Xi+1}} \right) \left( \frac{\prod_{k=1}^{K-1} \lambda_k^{-2}}{\left[ \frac{1}{\psi} + \sum_{k=1}^{K-1} \frac{1}{2\lambda_k} \right]^{(\chi+n-2)}} \right) \\ \times \left( \frac{\Gamma(K+1)\Gamma(n-1-K)(1+\delta^2)^{-\frac{K(P+1)}{2}}}{\left( \frac{\rho}{2} + \frac{1}{2} [\mathbf{x}^T \mathbf{x} - \frac{\delta^2}{1+\delta^2} (\mathbf{x}^T Z [Z^T Z]^{-1} Z^T \mathbf{x})] \right)^{\frac{P}{2}+1}} \right) \mathbb{I}_*(\boldsymbol{\eta}), \quad (18)$$

where the indicator function  $\mathbb{I}_*(\boldsymbol{\eta})$  is equal to 1 if  $\alpha_k \in [0, \alpha_{\max}]$  and  $\lambda_k \in [0, \lambda_{\max}]$  for all  $0 < k < K$  and if condition (5) is respected.

Closed-form expressions of the classical Bayesian estimators, namely the maximum a posteriori (MAP) or the minimum mean square error (MMSE), are not tractable from eq. (18). The approach driven here is to use a Markov Chain Monte-Carlo (MCMC) algorithm to generate samples from the posterior distribution (18). Bayesian estimates can be then deduced from the empirical distribution of the parameters. As the number of parameters of the model is not given, a RJMCMC sampler is used.

## 4. REVERSIBLE-JUMP MCMC ALGORITHM

The RJMCMC algorithm, introduced in [7], is an extension of the Metropolis-Hastings algorithm to the case where the number of variables is not given. In addition to the classical Metropolis-Hastings

moves for updating parameters values, changing-dimension moves are proposed and accepted with arbitrary probabilities. In this paper, the following moves are considered : 1) Birth or death of a rupture, 2) Update of the rupture parameters  $\eta_k$ , for all  $1 < k < K - 1$ , 3) Update of the hyperparameters  $\phi$ .

To assess the probability of acceptance for the changing-dimension moves, one must choose a probabilistic framework encompassing the whole model space. In this paper, a marked Poisson point process framework is chosen [10]. The rupture position  $\tau$  is the point process, whereas  $(\lambda, \alpha)$  are the associated marks. For each rupture parameter, a reference probability density must be chosen. A probability density  $l$  based on the discrete gradient of  $\mathbf{x}$  is chosen as a prior for the  $\tau_k$ ,  $0 < k < K - 1$ . This density  $l$  represents the intensity function of the marked Poisson point process. This specific choice of reference density reflects the *a priori* knowledge that the ruptures should coincide with high values of the gradient. This is in agreement with the definition (9) of the parameter  $\tau$ . In addition, non-informative reference probability densities  $d(\lambda) = \frac{1}{\lambda} \mathbb{I}_{[0, t_n]}$  and  $m(\alpha) = \mathbb{I}_{[0, \alpha_{\max}]}$  are chosen for the marks  $\lambda$  and  $\alpha$ . Finally, the overall reference density for the parameter vector  $\eta_k$  is  $Q(\eta_k) = l(\tau_k) d(\lambda_k) m(\alpha_k)$ .

### 4.1. Proposition distribution

To sample the  $\lambda_k$ , a method based on equation (10) is used. A subset  $\mathbf{x}_s$  of the observation sequence  $\mathbf{x}$  is extracted around  $\tau_k$  and is normalized to vary from 0 to 1. The following estimation of  $\lambda$  is obtained

$$\hat{\lambda} = \frac{1}{t_s} \sum_{(t < \tau_k)} (\mathbf{x}_s(t)) + \frac{1}{t_s} \sum_{(t > \tau_k)} (1 - \mathbf{x}_s(t)) \quad (19)$$

The proposition distribution  $d(\lambda | \boldsymbol{\eta})$  is then deduced as  $\lambda | \mathbf{x} \sim \mathcal{N}(\hat{\lambda}, s^2)$ , where  $s^2 = \text{var} [\mathbf{x}_s - \pi_{\tau_k, \hat{\lambda}, \alpha_k}(t)]$ .

The rupture location  $\tau$  and the rupture parameter  $\lambda$  are proposed according to their reference densities  $l$  and  $d$  respectively. Finally, the proposition distribution for the rupture parameters is  $P(\boldsymbol{\eta} | \boldsymbol{\eta}) = l(\tau) d(\lambda | \boldsymbol{\eta}) m(\alpha)$ .

### 4.2. Birth/death moves

Let  $\boldsymbol{\eta}^{(x)}$  be the parameter vector associated to the current configuration. In a birth move, a new marked point  $\eta_u = (\tau_u, \lambda_u, \alpha_u)$  is sampled according to the proposition distribution  $P(\eta_u | \boldsymbol{\eta}^{(x)})$ . The proposed new configuration is  $\boldsymbol{\eta}^{(y)} = \boldsymbol{\eta}^{(x)} \cup \eta_u$ , and the corresponding acceptance rate  $R(\boldsymbol{\eta}^{(x)}, \boldsymbol{\eta}^{(y)})$  expresses as

$$R(\boldsymbol{\eta}^{(x)}, \boldsymbol{\eta}^{(y)}) = \frac{f(\boldsymbol{\eta}^{(y)}, K^{(y)}, \rho, \delta^2 | \mathbf{x})}{f(\boldsymbol{\eta}^{(x)}, K^{(x)}, \rho, \delta^2 | \mathbf{x})} p_d(\eta_u | \boldsymbol{\eta}^{(y)}) \frac{Q(\eta_u)}{P(\eta_u | \boldsymbol{\eta}^{(x)})},$$

where  $p_d(\eta_u | \boldsymbol{\eta}^{(y)})$  is the probability for the rupture  $u$  to be selected in the death move, when the current configuration is  $\boldsymbol{\eta}^{(y)}$ :  $p_d(\eta_u | \boldsymbol{\eta}^{(y)}) = \frac{1}{K^{(y)} - 1}$ .

In a death move, a rupture  $\eta_u$  to be suppressed is selected uniformly from the current configuration  $\boldsymbol{\eta}^{(x)}$ . The proposed new configuration is  $\boldsymbol{\eta}^{(y)} = \boldsymbol{\eta}^{(x)} \setminus \eta_u$  and the acceptance rate becomes

$$R(\boldsymbol{\eta}^{(x)}, \boldsymbol{\eta}^{(y)}) = \frac{f(\boldsymbol{\eta}^{(y)}, K^{(y)}, \rho, \delta^2 | \mathbf{x})}{f(\boldsymbol{\eta}^{(x)}, K^{(x)}, \rho, \delta^2 | \mathbf{x})} \frac{1}{p_d(\eta_u | \boldsymbol{\eta}^{(x)})} \frac{P(\eta_u | \boldsymbol{\eta}^{(y)})}{Q(\eta_u)}$$

### 4.3. Update of the rupture parameters

A rupture  $\eta_u$  is chosen uniformly among the current rupture configuration  $\boldsymbol{\eta}^{(x)}$ . This rupture is updated to  $\eta_v$ . The new configuration becomes  $\boldsymbol{\eta}^{(y)} = \boldsymbol{\eta}^{(x)} \setminus \eta_u \cup \eta_v$ . If  $\eta_v$  is sampled according to  $P(\eta_v | \boldsymbol{\eta}^{(x)})$ , then the acceptance rate is:

$$R(\boldsymbol{\eta}^{(x)}, \boldsymbol{\eta}^{(y)}) = \frac{f(\boldsymbol{\eta}^{(y)}, K^{(y)}, \rho, \delta^2 | \mathbf{x}) Q(\eta_v) P(\eta_u | \boldsymbol{\eta}^{(x)} \setminus \eta_u)}{f(\boldsymbol{\eta}^{(x)}, K^{(x)}, \rho, \delta^2 | \mathbf{x}) Q(\eta_u) P(\eta_v | \boldsymbol{\eta}^{(y)} \setminus \eta_v)}$$

### 4.4. Update of the hyperparameters

One can deduce from the joint posterior distribution the marginal posterior distribution of each hyperparameter. Then the new values of these hyperparameters are sampled from their true conditional posterior, leading to an acceptance ratio equal to 1. These moves correspond to the classical Gibbs moves.

## 5. APPLICATION ON ELECTRICAL TRANSIENT DATA

The algorithm introduced in this paper has been tested on modelling electrical transient signals with a view to non-intrusive load monitoring. In this context, electrical transients are analyzed to detect the turned on appliances of an electrical network. The real-world transients used in this work were provided by Electricité de France (EDF), the French leading company for electricity distribution.

In this paper, an electrical transient due to a fridge turn on event is considered. The transient envelope contains the information useful for the classification task. Thus, the proposed MCMC algorithm is applied to fit this envelope. A polynomial regression model of order  $P = 0$  is applied on each segment. It represents a regime with a constant current consumption level. The following results are obtained in two steps. First the RJMCMC algorithm was run with a given value  $\delta^2 = 20$ , during  $0.5 \times 10^4$  iterations. This value reflects the desired reconstruction error on the observed signal. Figure 2(e) shows the resulting marginal posterior density of  $K$ . Then the same algorithm was run again, but with  $K = \hat{K}_{\text{map}} = 4$  (thus, without birth/death moves), while  $\delta^2$  was sampled according to his conditional posterior distribution. During the second step  $10^4$  iterations were performed. Fig. 2(b)-(c)-(d) show, for each rupture, the posterior distributions for the transition parameters  $(\alpha_k, \lambda_k)$ , while the posterior distribution of the rupture locations  $\tau$ , the observed time series and the MMSE estimate  $\hat{\mathbf{x}} = Z(\hat{\boldsymbol{\eta}}_{\text{mmse}}) \hat{\boldsymbol{\beta}}_{\text{mmse}}$  of the observed signal  $\mathbf{x}$  are displayed on Fig. 2(a).

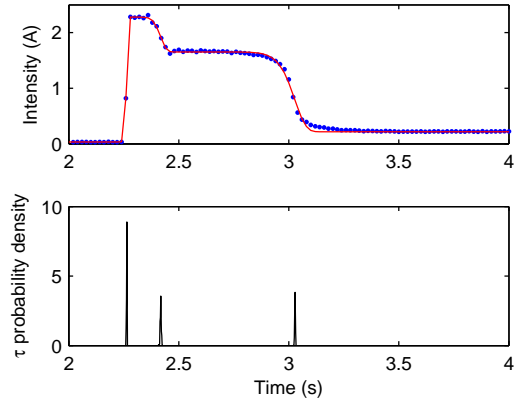
Although it slightly diverges from the observation around the third rupture, one can see that the signal estimate  $\hat{\mathbf{x}}$  provides an accurate representation of the observed signal  $\mathbf{x}$  and sparse as  $\hat{\mathbf{x}}$  depends on 13 parameters instead of initial 100 samples. The method proposed in this paper has been used for 50 real-world transients with similar results. The parameter estimates will be used in future works as characteristic features of the transient for classification.

## 6. CONCLUSION

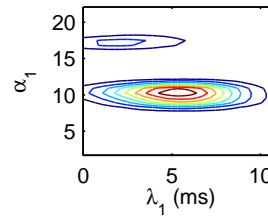
The development of a Bayesian framework for the joint estimation of the number of segments and the parameters of the flexible smooth transition regression model is the main contribution of this paper. A RJMCMC algorithm is derived to fit a given time serie in a full Bayesian framework. This method has been tested with promising results on the envelope of real-world electrical transients in a non-intrusive load monitoring context.

## 7. REFERENCES

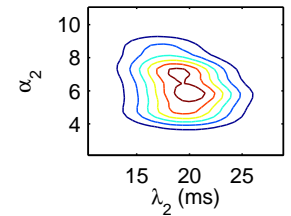
[1] Michèle Basseville and Igor V. Nikiforov, *Detection of abrupt changes : Theory and application*, 1993.



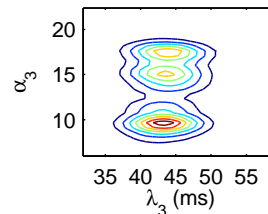
(a) top:(blue dots) observed signal, (red line) MMSE signal estimate  
bottom: $\tau$  posterior density



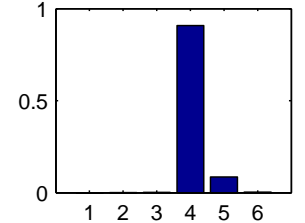
(b)  $(\alpha_1, \lambda_1)$  posterior distribution



(c)  $(\alpha_2, \lambda_2)$  posterior distribution



(d)  $(\alpha_3, \lambda_3)$  posterior distribution



(e)  $K$  posterior distribution

**Fig. 2. Results**

- [2] Elena Punskeya, Christophe Andrieu, Arnaud Doucet, and Fitzgerald William, "Bayesian Curve Fitting Using MCMC With Application to Signal Segmentation," *IEEE Transactions On Signal Processing*, vol. 50, no. 3, pp. 747–758, 2002.
- [3] Nicolas Dobigeon and Jean-Yves Tourneret, "Joint segmentation of piecewise autoregressive processes by using a hierarchical model and a bayesian sampling approach," *IEEE Transactions On Signal Processing*, vol. 55, no. 4, 2007.
- [4] Michel Lubrano, *Bayesian Analysis of Nonlinear Time Series Models with a Threshold*, pp. 79–118, Cambridge University Press, 2000.
- [5] Dick van Dijk, Terävirta Timo, and Philip Hans Franses, "Smooth transition autoregressive models - a survey on recent development," Tech. Rep., Econometric Institute, Erasmus University Rotterdam, 2000.
- [6] Faïcel Chamroukhi, Allou Samé, Gérard Govaert, and Patrice Aknin, "Time series modeling by a regression approach based on a latent process," *Neural Networks*, vol. 22, no. 5-6, pp. 593–602, 2009.
- [7] Peter J. Green, "Reversible Jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 52, pp. 711–732, 1995.
- [8] Kohlrusch, "Theorie des elektrischen rückstandes in der leidener flasche," *Pogg. Ann. Phys. Chem*, vol. 91, pp. 179–214, 1854.
- [9] Matthieu Sanguer, Florent Chatelain, Mabrouka El Guedri, and Nadine Martin, "Bayesian curve fitting for transient signals by using smooth transition regression models.," in *Conference on Condition Monitoring and Machinery Failure Prevention Technologies. CM 2010 and MFPT 2010*, Stratford-upon-Avon Royaume-Uni, 06 2010.
- [10] D.J Daley and D. Vera-Jones, *An Introduction To The Theory Of Point Process*, Springer Series in Statistics, 2nd edition, 2003.