Approche bayésienne pour la segmentation de séries corrompues par un biais fonctionnel.

Meili Baragatti [1] & Karine Bertin [2] & Emilie Lebarbier [3] & Cristian Meza [2]

[1] *SupAgro-INRA UMR MISTEA, Montpellier,FRANCE. meili.baragatti@supagro.fr.*
[2] *CIMFAV-Facultad de Ingeniería, Universidad de Valparaíso, Valparaíso, Chili. karine.bertin@uv.cl, cristian.meza@uv.cl.*
[3] AgroParisTech/INRA UMR518, Paris 5e, France. emilie.lebarbier@agroparistech.fr.

Résumé. Nous proposons une approche bayésienne pour détecter des points de rupture multiples, dans un signal constant par morceaux corrompu par un biais fonctionnel. Ce biais peut correspondre à des perturbations environnementales ou expériementales. La partie constante par morceaux est exprimée comme le produit d'une matrice triangulaire inférieure avec un vecteur parcimonieux. la partie fonctionnelle est exprimée comme une combinaison linéaire de fonctions provenant d'un dictionnaire. Une approche de sélection de variables "Stochastic Search Variable Selection" est utilisée pour obtenir des estimations parcimonieuses des paramètres de segmentation (les points de rupture et les moyennes des segments) ainsi que de la partie fonctionnelle. Les performances de la méthode proposée seront illustrées sur des simulations, puis la méthode sera appliquée sur deux jeux de données réelles dans les domaines de la géodésie et de l'économie.

Mots-clés. Segmentation, série corrompue, dictionnaire de fonctions, sélection de variables, Stochastic search variable selection.

Abstract. We propose a Bayesian approach to detect multiple change-points in a piecewise-constant signal corrupted by a functional part corresponding to environmental or experimental disturbances. The piecewise constant part (also called segmentation part) is expressed as the product of a lower triangular matrix by a sparse vector. The functional part is a linear combination of functions from a large dictionary. A Stochastic Search Variable Selection approach is used to obtain sparse estimations of the segmentation parameters (the change-points and the means over the segments) and of the functional part. The performance of our proposed method is assessed using simulation experiments. Applications to two real datasets from geodesy and economy fields are also presented.

Keywords. Segmentation, Corrupted series, Dictionary approach, Stochastic search variable selection.

1 Introduction

The problem of detecting multiple change-points in signals arises in many fields such as biology (Boys and Henderson 2004), geodesy (Williams, 2003; Bertin et al., 2016), meteorology (Caussinus and Mestre, 2004; Fearnhead, 2066; Wyse et al., 2011; Ruggieri, 2013) or astronomy (Dobigeaon et al., 2007) among others. In addition to change-points, we may observe environmental or experimental disturbances (for instance geophysical signals or climatic effects) which need to be taken into account in the models. Since the form of these disturbances is in general unknown or partially unknown, it seems natural to model them as a functional part that has to be estimated. Our goal in this article is to develop a Bayesian approach that allows us to both estimate the segmentation part (the change-points and the means over the segments) and the functional part. The Bayesian approach has the advantage that expert knowledge can be introduced in the models through prior distributions. This can be useful in multiple change-points problems where change-points can be related to specific events such as instrumental changes, earthquakes, very hot years or months or economic crisis for example. Moreover, posterior distributions allow us for a quantification of the uncertainty, giving in particular posterior probabilities or credible intervals for the positions of change-points or the functional part. This is of particular interest for practitioners.

Several methods have been proposed in a Bayesian framework for the multiple change-points problem. These methods are based, mostly, on reversible jump Markov Chain Monte Carlo algorithms (Lavielle and Lebarbier, 2001; Boys and Henderson, 2004; Tai and Xing, 2010), Stochastic search Variable Selection (Dobigeon et al., 2007), dynamic programming recursions (Ruggieri, 2013) or non-parametric Bayesian approaches (Martinez and Mena, 2014 and references there in). All these Bayesian methods deal with the multiple change-points problem but they do not consider the presence of functional disturbances. However, as illustrated in Picard (2011) and Bertin et al. (2016) in simulation and real examples, taking into account the functional part in the segmentation model can be crucial for an accurate change-point detection and interesting information can be extracted from the form of the functional part.

We propose a novel Bayesian method to detect multiple change-points in a piecewise-constant signal corrupted by a functional part, where the functional part is estimated using a dictionary approach (Bickel et al., 2009) and the segmentation part is treated as a sparse problem. More precisely, concerning the segmentation, we follow Harchaoui and Lévy-Leduc (2010) by expressing the piecewise constant part of the model as a product of a lower triangular matrix by a sparse vector (which non-zero coordinates correspond to change-points positions). In addition, the functional part is represented as a linear combination of functions from a dictionary. Since a large variety of functions can be included in the dictionary, this leads generally to a sparse representation of the functional part in terms of functions from the dictionary. Hence, a Stochastic search Variable Selection approach can be used to estimate the sparse vectors, that is, both the location of the change-points and the functional part (Gearges and McCulloch, 1997).

2 Model

2.1 Segmentation model with functional part

We observe a series $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ that satisfies

$$Y_t = \mu_k + f(x_t) + \epsilon_t, \quad \forall t \in I_k = (\tau_{k-1}, \tau_k], k \in \{1, \ldots, K\}, \tag{1}$$

where K is the total number of segments of the series is unknown, the ϵ_t are i.i.d centered Gaussian variables with variance σ^2, x_t is a covariate (the simple one is the time t), f is an unknown function to be estimated, τ_k is the kth change-point, μ_k is the mean of the series on the segment I_k. We use the convention $\tau_0 = 0$ and $\tau_K = n$.

A classical approach in non-parametric framework is to expand the functional part f with respect to orthonormal basis, such as Fourier or wavelet ones (see Hardle et al., 1998 and references therein). Following Bickel et al. (2009) or Bertin et al. (2016), we choose here to adopt a dictionary approach, that consists in finding an over-complete representation of f. More precisely, we expand f with respect to a large family of functions $(\phi_j)_{j=1,\ldots,M}$, named dictionary, that can for example be the union of two orthonormal basis. Then f is assumed to be of the form

$$f(x) = \sum_{j=1}^{M} \lambda_j \phi_j(x),$$

where $\lambda = (\lambda_1, \ldots, \lambda_M)' \in \mathcal{R}^M$ is a vector of coordinates of f in the dictionary and

$$(f(x_1), \ldots, f(x_n))' = \mathbf{F}\lambda,$$

where $\mathbf{F}$ is the $n \times M$ matrix $\mathbf{F} = (\phi_j(x_i))_{i,j}$. Note that since large dictionaries are considered, this allows us to obtain a sparse representation of the function f, that is the vector λ is expected to be with few non-zero coordinates.

To estimate the change-points in the series, we follow the strategy proposed by Lévy-Leduc (2010), which consists in reframing this task in a variable selection context. We denote by $\mathbf{X}$ the $n \times n$ lower triangular matrix having only 1's on the diagonal and below it. We consider the $n \times 1$ vector β with only K non-zero coefficients at positions $(\tau_k + 1)_{k=0,\ldots,K-1}$ with $\beta_{\tau_k+1} = \mu_{k+1} - \mu_k$ and using the convention $\mu_0 = 0$. Note that the segmentation (the change-points τ_k and the means μ_k) will be recovered by the vector β.

The model (1) can then be rewritten as follows

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{F}\lambda + \epsilon,$$

where $\epsilon = (\epsilon_1, \ldots, \epsilon_n)'$. Our objective is now to estimate the parameters β, λ and σ^2. Since both β and λ vectors are expected to be sparse, we propose to use Bayesian methods of variable selection for their estimation.

2.2 Bayesian hiercharchical framework

Following George and McCulloch (1993), we first introduce latent variables $\boldsymbol{\gamma}$ and $\mathbf{r}$ to identify non-null components of the vectors β and λ. The vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)$ is such that $\gamma_i = I_{\{\beta_i \neq 0\}}$, where I denotes the indicator function and the vector $\mathbf{r} = (r_1, \ldots, r_M)$ satisfies $r_j = I_{\{\lambda_j \neq 0\}}$. The number of non-zero coordinates of $\boldsymbol{\gamma}$ and $\mathbf{r}$ are $d_{\boldsymbol{\gamma}} = K$ and $d_{\mathbf{r}}$ respectively. The product $\mathbf{X}\beta$ is equal to $\mathbf{X}_{\boldsymbol{\gamma}}\beta_{\boldsymbol{\gamma}}$ where $\mathbf{X}_{\boldsymbol{\gamma}}$ is the $n \times d_{\boldsymbol{\gamma}}$ matrix containing only the j columns of X such that γ_j is non-zero and $\beta_{\boldsymbol{\gamma}}$ is a $d_{\boldsymbol{\gamma}} \times 1$ vector containing only the non-zero coefficients of β. Similarly, we can express $\mathbf{F}\lambda$ as $\mathbf{F}_{\mathbf{r}}\lambda_{\mathbf{r}}$ where $\mathbf{F}_{\mathbf{r}}$ is a $n \times d_{\mathbf{r}}$ matrix and $\lambda_{\mathbf{r}}$ a $d_{\mathbf{r}} \times 1$ vector. The model (1) can be then rewritten as

$$\mathbf{Y} = \mathbf{X}_{\boldsymbol{\gamma}}\beta_{\boldsymbol{\gamma}} + \mathbf{F}_{\mathbf{r}}\lambda_{\mathbf{r}} + \epsilon$$

where the parameters to estimate are $\boldsymbol{\theta} = \{\beta_{\boldsymbol{\gamma}}, \boldsymbol{\gamma}, \lambda_{\mathbf{r}}, r, \sigma^2\}$.

Then, as usual in a Bayesian context, these parameters are treated as random variables, assumed here to be independent, and we consider the following prior distributions. The γ_i are independent Bernoulli variables with parameter $0 \leq \pi_i \leq 1$ for $i = 2, \ldots, n$ and with $\pi_1 = 1$ by convention. The r_j are also independent Bernoulli variables with parameter $0 \leq \eta_j \leq 1$ for $j = 1, \ldots, M$. Then the noise parameter follows a Jeffrey distribution, $\pi(\sigma^2) \propto \sigma^{-2}$. The conditional distribution of $\beta_{\boldsymbol{\gamma}}|\boldsymbol{\gamma}, \sigma^2$ is the classical g-prior of Zellner (1986) given by $\beta_{\boldsymbol{\gamma}}|\boldsymbol{\gamma}, \sigma^2 \sim \mathcal{N}_{d_{\boldsymbol{\gamma}}}\left(0, c_1\sigma^2\left(\mathbf{X}'_{\boldsymbol{\gamma}}\mathbf{X}_{\boldsymbol{\gamma}}\right)^{-1}\right)$. Finally the conditional distribution of $\lambda_{\mathbf{r}}|\mathbf{r}, \sigma^2$ is also a g-prior, with $\lambda_{\mathbf{r}}|\mathbf{r}, \sigma^2 \sim \mathcal{N}_{d_{\mathbf{r}}}\left(0, c_2\sigma^2\left(\mathbf{F}'_{\mathbf{r}}\mathbf{F}_{\mathbf{r}}\right)^{-1}\right)$.

The posterior distribution of $\boldsymbol{\theta}$ has the following expression

$$\pi(\boldsymbol{\theta}|\mathbf{Y}) = \frac{\pi(\mathbf{Y}|\boldsymbol{\theta})\pi(\beta_{\boldsymbol{\gamma}}|\boldsymbol{\gamma}, \sigma^2)\pi(\lambda_{\mathbf{r}}|\mathbf{r}, \sigma^2)\pi(\boldsymbol{\gamma})\pi(\mathbf{r})\pi(\sigma^2)}{\pi(\mathbf{Y})}, \tag{2}$$

where

$$\pi(\mathbf{Y}|\boldsymbol{\theta}) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\left(\mathbf{Y} - \mathbf{X}_{\boldsymbol{\gamma}}\beta_{\boldsymbol{\gamma}} - \mathbf{F}_{\mathbf{r}}\lambda_{\mathbf{r}}\right)'\left(\mathbf{Y} - \mathbf{X}_{\boldsymbol{\gamma}}\beta_{\boldsymbol{\gamma}} - \mathbf{F}_{\mathbf{r}}\lambda_{\mathbf{r}}\right)\right).$$

3 MCMC Schemes

A classical approach for the computational scheme would be to estimate the whole parameters at the same time $(\beta_{\boldsymbol{\gamma}}, \boldsymbol{\gamma}, \lambda_{\mathbf{r}}, \mathbf{r}, \sigma^2)$ using a Metropolis-within-Gibbs algorithm combined with the grouping (or blocking) technique of Liu (1994). However, we will see that two main drawbacks are associated with this algorithm. Therefore, we propose the following two-step strategy: the first step aims at detecting the positions of the change-points and at selecting the functions, that is, to estimate the latent vectors $\boldsymbol{\gamma}$ and $\mathbf{r}$. To this end, the parameters $\beta_{\boldsymbol{\gamma}}$, $\lambda_{\mathbf{r}}$ and σ^2 can be considered as nuisance parameters, and we use the joint posterior distribution integrated with respect to $\beta_{\boldsymbol{\gamma}}$, $\lambda_{\mathbf{r}}$ and σ^2. This can be viewed as a collapsing technique, see Liu (1994) and van Dyk and Park (2008). In the second part, we estimate $\beta_{\boldsymbol{\gamma}}$, $\lambda_{\mathbf{r}}$ and σ^2, conditionally to $\boldsymbol{\gamma}$ and $\mathbf{r}$. Some details of both steps will be given during the presentation.

4 Simulation study, applications and discussion

The performance of our proposed method is assessed using simulation experiments. During the presentation, main results will be given.In particular, good results for both the segmentation and the functional parts have been achieved.
Then we will show the results obtained on two real datasets: a GPS series from an Australian station and a series of daily records of Mexican Peso/US Dollar exchange rate, for which expected change-points are recovered. Moreover the benefits of the Bayesian approach will be illustrated.
Eventually, our approach will be discussed.

Bibliographie

[1] Bertin, K., Collilieux, X., Lebarbier, E., and Meza, C. (2016). Segmentation of multiple series using a Lasso strategy. Submitted - arXiv:1406.6627.
[2] Bickel, P. J., Ritov, Y., and Tsybakov, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. Ann. Statist., 37(4): 1705-1732.
[3] Boys, R. J. and Henderson, D. A. (2004). A Bayesian approach to DNA sequence segmentation. Biometrics, 60: 573-588. With discussion.
[4] Caussinus, H. and Mestre, O. (2004). Detection and correction of artificial shifts in climate series. Applied Statistics, 53: 405-425.34.
[5] Dobigeon, N., Tourneret, J.-Y., and Scargle, J. (2007). Joint segmentation of multivariate astronomical time series: bayesian sampling with a hierarchical model. IEEE Transactions on Signal Processing, 55: 414-423.
[6] Fearnhead, P. (2006). Exact and efficient Bayesian inference for multiple changepoint problems. Statistics and Computing, 16: 203-213.
[7] George, E.I. and McCulloch, R.E. (1993). Variable selection via Gibbs sampling. Journal of the American Statistical Association, 88(423): 881-889.
[8] Harchaoui, Z. and Lévy-Leduc, C. (2010). Multiple Change-Point Estimation With a Total Variation Penalty. Journal of the American Statistical Association, 105:1480-1493.
[9] Hardle, W., Kerkyacharian, G., Picard, D., and Tsybakov, A. (1998). Wavelets, approximation, and statistical applications, volume 129 of Lecture Notes in Statistics. Springer-Verlag, New York.

[10] Lavielle, M. and Lebarbier, E. (2001). An application of MCMC methods for the multiple change-points problem. Signal processing , 81: 39-53.
[11] Liu, J.S (1994). The collapsed Gibbs sampler in Bayesian computations with application to a gene regulation problem. J. Am. Stat. Ass., 89(427): 958-966.
[12] Martinez, A. and Mena, R.(2014). On a Nonparametric Change Point Detection Model in Markovian Regimes. Bayesian Analysis, 9: 823-858.
[13] Picard, F., Lebarbier, E., Budinska, E., and Robin (2011). Joint segmentation of multivariate Gaussian processes using mixed linear models. Comp. Stat. and Data Analysis, 55: 1160-1170.
[14] Ruggieri, E. (2013). A Bayesian approach to detecting change points in climatic records. International Journal of Climatology, 33: 520-528.
[15] Tai, T.L. and Xing, H. (2010). A fast Bayesian change point analysis for the segmentation of microarray data. Bioinformatics., 24: 2143-2148.
[16] van Dyk, D.A. and Park, T. (2008). Partially collapsed Gibbs samplers: theory and methods. J. Am. Stat. Ass., 103: 790-796.36.
[17] Williams, S. (2003). Offsets in Global Positioning System time series. Journal of Geophysical Research (Solid Earth), 108(B6, 2310).
[18] Wyse, J., Friel, N., and Rue, H. (2011). Approximate simulation-free Bayesian inference for multiple changepoint models with dependence within segments. Bayesian Analysis, 6: 501-528.
[19] Zellner, A. (1986). Bayesian inference and decision techniques - essays in honour of Bruno De Finetti, chapter On assessing prior distributions and Bayesian regression analysis with g-prior distributions. Goel, P.K. and Zellner, A, 233-243.