

# Bayesian mixture of spatial spline regressions

Faïcel Chamroukhi<sup>1,2</sup>

1- Université de Toulon, CNRS, LISIS, UMR 7296, 83957 La Garde, France

2- Laboratoire de Mathématiques Paul Painlevé, UMR CNRS 8524, Université Lille 1

**Abstract.** We introduce a Bayesian mixture of spatial spline regressions with mixed-effects (BMSSR) for density estimation and model-based clustering of spatial functional data. The model, through its Bayesian formulation, allows to integrate possible prior knowledge on the data structure and constitute a good alternative to a recent mixture of spatial spline regressions model estimated in a maximum likelihood framework via the expectation-maximization (EM) algorithm. The Bayesian model inference is performed by Markov Chain Monte Carlo (MCMC) sampling. We derive a Gibbs sampler to infer the model and apply it on simulated surfaces and a real problem of handwritten digit recognition using the MNIST data.

## 1 Introduction

Functional data analysis (FDA) [1] is the paradigm of data analysis in which the individuals are functions (e.g., curves or surfaces) rather than vectors of reduced dimension. Most of the classical analyses directly consider the data to be analyzed as vectors. However, in many areas of application, the analyzed data are often available in the form of (discretized) values of functions or curves (e.g., times series, waveforms, etc) and surfaces (2D-images, spatio-temporal data, etc) which makes them very structured. This “functional” aspect of the data adds additional difficulties in the the analysis compared to the case of a classical multivariate analysis. In this framework, several models have been introduced to model univariate and multivariate functional data for clustering or classification. Among these models, one distinguishes those based on the finite mixture model [2], on which we focus in this paper.

These models have however mainly focused on the study of univariate or multivariate functions. For the case of spatial functional data, [3, 4, 5, 6] proposed methods to deal with surfaces. In particular, the recent approach proposed by [6] for clustering and classification of surfaces is based on the regression spatial spline regression as in [5] in a mixture of linear mixed-effects model framework as in [7]. [6] indeed extended the functional data analysis framework for univariate functions to the analysis of spatial functions (i.e. surfaces) by introducing a spatial spline regression (SSR) model and a mixture of spatial spline regressions (MSSR) model, to respectively model homogeneous surfaces and heterogeneous surfaces with a clustering structure. The SSR model with mixed-effects is tailored to spatial regression data with both fixed-effects and random-effects. The mixture of spatial spline regression (MSSR) is dedicated to surface clustering, as in [8] for curve clustering, while the mixture of spatial spline regression discriminant analysis (MSSR-DA) is dedicated to curve discrimination. The usual used

tool for model estimation is maximum likelihood estimation (MLE) by using the expectation-maximization (EM) algorithm [9, 10].

In this paper, we present a probabilistic Bayesian formulation to model spatial functional data by extending the approaches of [6] and apply the proposal to surface approximation and clustering. The model is also related to the random-effects mixture model of [11] in which we explicitly add mixed-effects and derive it for spatial functional data by using the Nodal basis functions (NBFs). The NBFs [3] used in [4, 5, 6] represent an extension of the univariate B-spline bases to bivariate surfaces. We thus introduce the Bayesian mixtures of SSR (BMSSR) for fitting populations of heterogeneous surfaces organized in groups. The model is applied for fitting a population of homogeneous surfaces and the model-based surface clustering by considering handwritten digits from the MNIST data [12].

This paper is organized as follows. Section 2 presents the Bayesian mixture of spatial spline regressions (BMSSR) model and its inference technique using Gibbs sampling. Then, in section 3, we apply the proposed model on simulated surfaces and on a real handwritten digit recognition problem. Finally, in Section 4, we draw some conclusions and mention some future work.

## 2 Bayesian mixture spatial spline regressions with mixed-effects (BMSSR)

We introduce a Bayesian probabilistic formulation to the mixture of spatial spline regressions with mixed-effects presented in [6] in a maximum likelihood context. The proposed model is thus the Bayesian mixture of spatial spline regressions with mixed-effects (BMSSR).

### 2.1 The model

Consider that there are  $K$  sub-populations in the set of  $n$  surfaces  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$ . The proposed BMSSR model has the following stochastic representation. Conditional on component  $k$ , the individual  $\mathbf{y}_i$  is modeled by a BSSR model as:

$$\mathbf{y}_i = \mathbf{S}_i(\boldsymbol{\beta}_k + \mathbf{b}_{ik}) + \mathbf{e}_{ik} \quad (1)$$

where the spatial regression matrix  $\mathbf{S}_i$  is computed from the Nodal basis functions. Introduced by [3], the idea of Nodal basis functions (NBFs) extends the use of B-splines for univariate function approximation [1], to the approximation of surfaces. For a fixed number of basis functions  $d$ , defined on a regular grid with regularly spaced points  $c(l)$  ( $l = 1, \dots, d$ ) of the domain we are working on, with  $d$  defined as  $d = d_1 d_2$  where  $d_1$  and  $d_2$  are respectively the columns and rows number of nodes, the  $i$ th surface can be approximated using piecewise linear Lagrangian triangular finite element NBFs constructed as in [5, 6]  $s(\mathbf{x}, \mathbf{c}, \delta_1, \delta_2)$  (the shape parameters  $\delta_1$  and  $\delta_2$  being constant). An example of a NBF function defined on the rectangular domain  $(x_1, x_2) \in [-1, 1] \times [-1, 1]$  with a single node  $\mathbf{c} = (0, 0)$  and  $\delta_1 = \delta_2 = 1$  is presented in the Figure 1.

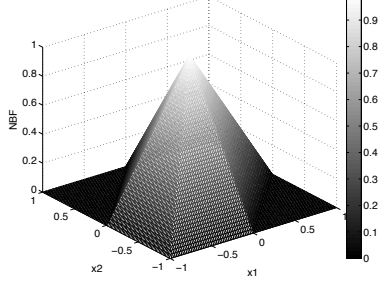


Figure 1: Nodal basis function  $s(\mathbf{x}, \mathbf{c}, \delta_1, \delta_2)$ , where  $\mathbf{c} = (0, 0)$  and  $\delta_1 = \delta_2 = 1$ .

Thus, a  $K$  component Bayesian mixture of spatial spline regression models with mixed-effects (BMSSR) has the following density:

$$f(\mathbf{y}_i | \mathbf{S}_i; \Psi) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{y}_i; \mathbf{S}_i(\boldsymbol{\beta}_k + \mathbf{b}_{ik}), \sigma_k^2 \mathbf{I}_{m_i}) \quad (2)$$

where the parameter vector of the model is given by

$$\Psi = (\pi_1, \dots, \pi_{K-1}, \boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_K^T, \mathbf{B}_1^T, \dots, \mathbf{B}_K^T, \sigma_1^2, \dots, \sigma_K^2, \xi_1^2, \dots, \xi_K^2)^T,$$

$\mathbf{B}_k = (\mathbf{b}_{1k}^T, \dots, \mathbf{b}_{nk}^T)^T$  being the vector of the random-effect coefficients of the  $k$ th BSSR component. The BMSSR model is indeed composed of Bayesian Spatial Spline Regression components, each of them has parameters  $\Psi_k = (\boldsymbol{\beta}_k^T, \mathbf{B}_k^T, \sigma_k^2, \xi_k^2)^T$  and a mixing proportion parameter  $\pi_k$ . In this Bayesian setting, we therefore just need to specify the prior distribution on the mixing proportions  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_K)$  which follow the Multinomial distribution in the generative model of the non-Bayesian mixture. We use a conjugate prior as for the other parameters, that is, a Dirichlet prior with hyper-parameters  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)$ . The hierarchical prior from for the BMSSR model parameters is therefore given by:

$$\begin{aligned} \boldsymbol{\pi} &\sim \text{Dir}(\alpha_1, \dots, \alpha_K) \\ \boldsymbol{\beta}_k &\sim \mathcal{N}(\boldsymbol{\beta}_k | \boldsymbol{\mu}_0, \Sigma_0) \\ \mathbf{b}_{ik} | \xi_k^2 &\sim \mathcal{N}(\mathbf{b}_{ik} | \mathbf{0}_d, \xi_k^2 \mathbf{I}_d) \\ \xi_k^2 &\sim \text{IG}(\xi_k^2 | a_0, b_0) \\ \sigma_k^2 &\sim \text{IG}(\sigma_k^2 | g_0, h_0). \end{aligned} \quad (3)$$

## 2.2 Bayesian inference using Gibbs sampling

In this section we derive the Gibbs sampler to infer the model parameters. The pseudo-code 1 summarizes the Gibbs of the proposed BMSSR model and specifies the full conditional distributions.

---

**Algorithm 1** Gibbs sampler for the BMSSR model
 

---

**Inputs:** The observations  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$  and the spatial spline regression matrices  $(\mathbf{S}_1, \dots, \mathbf{S}_n)$  and the number of mixture components  $K$

**Initialize:** the hyper-parameters  $(\boldsymbol{\alpha}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, g_0, h_0, a_0, b_0)$  and the parameters  $(\boldsymbol{\pi}, \boldsymbol{\beta}, \mathbf{B}, \boldsymbol{\xi}^2, \boldsymbol{\sigma}^2)$

**for**  $t = 1$  **to** #Gibbs samples **do**

**for**  $i = 1$  **to**  $n$  **do**

1. The allocation variables:  $z_i^{(t)} \sim \text{Mult}(1; \tau_{i1}^{(t)}, \dots, \tau_{iK}^{(t)})$  with the posterior probabilities  $\tau_{ik}^{(t)}$  calculated according to  $\tau_{ik}^{(t)} = \frac{\pi_k^{(t)} \mathcal{N}(\mathbf{y}_i | \mathbf{S}_i(\boldsymbol{\beta}_k^{(t)} + \mathbf{b}_{ik}^{(t)}), \sigma_k^2 \mathbf{I}_{m_i})}{\sum_{l=1}^K \pi_l^{(t)} \mathcal{N}(\mathbf{y}_i | \mathbf{S}_i(\boldsymbol{\beta}_l^{(t)} + \mathbf{b}_{il}^{(t)}), \sigma_l^2 \mathbf{I}_{m_i})}$

**end for**

2. The mixing proportions:  $\boldsymbol{\pi}^{(t)} \sim \text{Dir}(\alpha_1 + n_1^{(t)}, \dots, \alpha_K + n_K^{(t)})$  with  $n_k^{(t)} = \sum_{i=1}^n z_{ik}^{(t)}$

**for**  $k = 1$  **to**  $K$  **do**

3. The random-effects variance:  $\xi_k^2 \sim IG\left(a_0 + \frac{n}{2}, b_0 + \frac{\sum_{i=1}^n \mathbf{b}_{ik}^{(t-1)T} \mathbf{b}_{ik}^{(t-1)}}{2}\right)$

4. Sample the noise variance:  
 $\sigma_k^2 \sim IG\left(g_0 + \frac{n_k}{2}, h_0 + \frac{\sum_{i=1}^n z_{ik} (\mathbf{y}_i - \mathbf{S}_i \boldsymbol{\beta}_k^{(t-1)} - \mathbf{S}_i \mathbf{b}_{ik}^{(t-1)})^T (\mathbf{y}_i - \mathbf{S}_i \boldsymbol{\beta}_k^{(t-1)} - \mathbf{S}_i \mathbf{b}_{ik}^{(t-1)})}{2}\right)$

5. The fixed-effects coefficient vector:  $\boldsymbol{\beta}_k^{(t)} \sim \mathcal{N}(\boldsymbol{\nu}_0^{(t)}, \mathbf{V}_0^{(t)})$  with  
 $\mathbf{V}_0^{-1(t)} = \boldsymbol{\Sigma}_0^{-1} + \frac{1}{\sigma_k^2} \sum_{i=1}^n z_{ik} \mathbf{S}_i^T \mathbf{S}_i,$   
 $\boldsymbol{\nu}_0^{(t)} = \mathbf{V}_0 \left( \frac{1}{\sigma_k^2} \sum_{i=1}^n z_{ik} \mathbf{S}_i^T (\mathbf{y}_i - \mathbf{S}_i \mathbf{b}_{ik}^{(t-1)}) - \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0 \right)$

**for**  $i = 1$  **to**  $n$  **do**

6. The random-effects coefficient vector:  $\mathbf{b}_{ik}^{(t)} \sim \mathcal{N}(\boldsymbol{\nu}_1^{(t)}, \mathbf{V}_1^{(t)})$  with  
 $\mathbf{V}_1^{-1(t)} = \frac{1}{\sigma_k^2} \mathbf{S}_i^T \mathbf{S}_i + \frac{1}{\xi_k^2}; \boldsymbol{\nu}_1^{(t)} = \mathbf{V}_1 \left( \frac{1}{\sigma_k^2} \mathbf{S}_i^T (\mathbf{y}_i - \mathbf{S}_i \boldsymbol{\beta}_k^{(t)}) \right)$

**end for**

**end for**

**end for**

---

### 3 Application to simulated data and real data

We first consider simulated surfaces to test the model in terms of surface approximation. Then, we apply it on a handwritten character recognition problem by considering real images from the MNIST data set [12] to test it in terms of surface approximation and clustering.

#### 3.1 Simulated surface approximation

We consider the bi-dimensional arbitrary function  $\mu(\mathbf{x}) = \frac{\sin(\sqrt{1+x_1^2+x_2^2})}{\sqrt{1+x_1^2+x_2^2}}$

and we attempt to approximate it from a sample of simulated noisy surfaces. We simulate a sample of 100 random surfaces  $\mathbf{y}_i (i = 1, \dots, 100)$  as follows. Each surface  $\mathbf{y}_i$  is composed of  $m_i = 21 \times 21$  observations generated on a square domain  $(x_1, x_2) \in [-10, 10] \times [-10, 10]$ . To generate the surface  $\mathbf{y}_i$ , we first add random effects to the mean surface by computing  $\boldsymbol{\mu}_i(\mathbf{x}) + \mathbf{b}_i$  and then  $\mathbf{y}_i$  is simulated by adding a random error term, that is,  $\mathbf{y}_i = \boldsymbol{\mu}_i(\mathbf{x}) + \mathbf{b}_i + \mathbf{e}_i$  with  $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, 0.1^2 \mathbf{I}_{m_i})$  and  $\mathbf{e}_i \sim \mathcal{N}(\mathbf{0}, 0.1^2 \mathbf{I}_{m_i})$ . Then, the sample of simulated

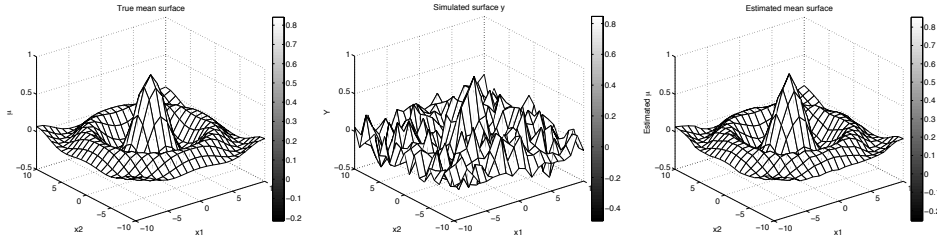


Figure 2: True mean surface (left), an example of noisy surface (middle), A BSSR fit from 100 surfaces using  $15 \times 15$  NBFs (right).

surfaces  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_{100})$  is approximated by applying the BMSSR model with one component. Figure 2 shows an example of actual arbitrary mean function before the noise and the random effects are added, an example of simulated surface the fitted mean surface  $\hat{\mu}(\mathbf{x}) = \mathbf{S}_i \hat{\beta}$  from a set of 100 surfaces with  $d = 15 \times 15$  NBFs. It can be seen that for the two cases, the approximated surface resembles the actual one. In particular, the second approximation, using a reasonable number of basis functions, is very close to the true surface. This is confirmed by the value of the empirical sum of squared error between the true surface and the fitted one  $SSE = \sum_{j=1}^m (\mu_j(\mathbf{x}) - \hat{\mu}_j(\mathbf{x}))^2$  ( $m = 441$  here), which equal 0.0865 in this case and which corresponds to a very reasonable fit.

### 3.2 Handwritten digit clustering using the BMMSSR model

In this section we apply the BMSSR model on a subset of the ZIPcode data set [13], which is issued from the MNIST data set [12]. The data set contains 9298 16 by 16 pixel gray scale images of Hindu-Arabic handwritten digits. Each individual  $\mathbf{y}_i$  contains  $m_i = 256$  observations  $\mathbf{y}_i = (y_{i1}, \dots, y_{i256})^T$  values in the range  $[-1, 1]$ . We used a subset of 1000 digits randomly chosen from the Zipcode testing set with  $d = 8 \times 8$  NBFs. The best partition is obtained for  $K = 12$  clusters and the corresponding mean Adjusted Rand Index (ARI) value equals 0.5238. Figure 3 shows the cluster means obtained by the proposed Bayesian model (BMSSR). It clearly shows that the model is able to recover the ten digits as well as subgroups of the digit 0 and the digit 5.

## 4 Conclusion and future work

We introduced the Bayesian mixture of spatial spline regressions with mixed-effects (BMSSR) for spatial functional data. The model is able to accommodate individuals with both fixed and random effect variability. We derived a Gibbs sampler to infer the model parameters. Application on simulated surfaces and real data in a handwritten digit recognition framework shows the potential benefit of the proposed model for practical applications. The BMSSR can be directly extended to be used for supervised surface classification. A future work will consist in conducting additional experiments on real data clustering and discrimination as well as model selection.



Figure 3: Cluster means obtained by the proposed BMSSR model

## References

- [1] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer Series in Statistics. Springer, June 2005.
- [2] G. J. McLachlan and D. Peel. *Finite mixture models*. New York: Wiley, 2000.
- [3] N. Malfait and J. O. Ramsay. The historical functional linear model. *The Canadian Journal of Statistics*, 31(2), 2003.
- [4] J.O. Ramsay, T.O. Ramsay, and L.M. Sangalli. *Recent Advances in Functional Data Analysis and Related Topics*, chapter Spatial functional data analysis, pages 269–275. Springer, 2011.
- [5] L.M. Sangalli, J.O. Ramsay, and T.O. Ramsay. Spatial spline regression models. *Journal of the Royal Statistical Society (Series B)*, 75:681–703, 2013.
- [6] Hien D. Nguyen, Geoffrey J. McLachlan, and Ian A. Wood. Mixtures of spatial spline regressions for clustering and classification. *Computational Statistics & Data Analysis*, 93:76 – 85, 2016.
- [7] G. Celeux, O. Martin, and C. Lavergne. Mixture of linear mixed models for clustering gene expression profiles from repeated microarray experiments. *Statistical Modelling*, 5:1–25, 2005.
- [8] G. M. James and C. Sugar. Clustering for sparsely sampled functional data. *Journal of the American Statistical Association*, 98(462), 2003.
- [9] G. J. McLachlan and T. Krishnan. *The EM algorithm and extensions*. New York: Wiley, second edition, 2008.
- [10] 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*, 39(1):1–38, 1977.
- [11] P. Lenk and W. DeSarbo. Bayesian inference for finite mixtures of generalized linear models with random effects. *Psychometrika*, 65(1):93–119, 2000.
- [12] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, Nov 1998.
- [13] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning, Second Edition: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, second edition edition, January 2010.