Statistica Sinica **33** (2023), 2715-2737
doi:https://doi.org/10.5705/ss.202021.0288

LÉVY ADAPTIVE B-SPLINE REGRESSION VIA OVERCOMPLETE SYSTEMS

Sewon Park[1], Hee-Seok Oh[2] and Jaeyong Lee[2]

[1]*Samsung SDS and* [2]*Seoul National University*

Abstract: Estimating functions with varying degrees of smoothness is a challenging problem in nonparametric function estimation. In this paper, we propose the Lévy adaptive B-Spline regression (LABS) model, an extension of the Lévy adaptive regression kernels (LARK) models, for estimating functions with varying degrees of smoothness. The LABS model is a LARK model with B-spline basis functions as generating kernels. The B-spline basis functions consist of piecewise k-degree polynomials with $k-1$ continuous derivatives, and can systematically express functions with varying degrees of smoothness. By changing the order of the B-spline basis, the LABS model systematically adapts to the smoothness of the functions, for example, jump discontinuities, sharp peaks, and so on. The results of simulation studies and real-data examples show that the proposed model captures smooth areas, jumps, and sharp peaks of functions. The proposed model also perform best in almost all examples. Finally, we provide theoretical results that the mean function for the LABS model belongs to certain Besov spaces, based on the order of the B-spline basis, and that the prior of the model has full support on the Besov spaces.

Key words and phrases: Besov space, Lévy random measure, Nonparametric function estimation, reversible jump Markov chain Monte Carlo.

1. Introduction

Suppose we observe n pairs of observations, $(x_1, y_1), \ldots, (x_n, y_n)$, where

$$y_i = \eta(x_i) + \epsilon_i, \quad \epsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2), \quad i = 1, \ldots, n, \tag{1.1}$$

and η is an unknown real-valued function that maps $\mathbb{R}$ to $\mathbb{R}$. We wish to estimate the mean function η, which may have varying degrees of smoothness, including discontinuities. In nonparametric function estimation, particularly in climate and economic data sets, we often encounter smooth curves, except for a finite number of jump discontinuities and sharp peaks. For example, heavy rainfall can cause a sudden rise in the water level of a river, epidemics such as the COVID-19 can increase unemployment rates, and policymakers' decisions can cause abrupt

Corresponding author: Sewon Park, Security Algorithm Lab, Samsung SDS, Seoul 05510, Korea. E-mail: swpark0413@gmail.com.

changes. For instance, the United States Congress passed the National Minimum Drinking Age Act in 1984, establishing 21 as the minimum legal age for purchasing alcohol. This Act caused a sudden rise in mortality for young Americans around the age of 21. Abrupt changes can provide us with meaningful information about such issues.

There has been much research on estimating the local smoothness of functions. The first approach is to minimize the penalized sum of squares based on a locally varying smoothing parameter or penalty function across the whole domain. Pintore, Speckman and Holmes (2006), Liu and Guo (2010), and Wang, Du and Shen (2013) modeled the smoothing parameter of a smoothing spline to vary over the domain. Ruppert and Carroll (2000), Crainiceanu et al. (2007), and Yang and Hong (2017) suggested penalized splines based on a local penalty that adapts to spatial heterogeneity in the regression function. The second approach is to use adaptive free-knot splines to choose the number and locations of the knots from the data. Friedman (1991) and Luo and Wahba (1997) determined a set of knots using stepwise forward/backward knot selection procedures. Zhou and Shen (2001) avoided the problems of stepwise schemes by proposing optimal knot selection schemes that introduced a knot relocation step. Smith and Kohn (1996), Denison, Mallick and Smith (1998a), Denison, Mallick and Smith (1998b), and DiMatteo, Genovese and Kass (2001) studied Bayesian estimations of free knot splines using Markov chain Monte Carlo (MCMC) techniques. The third approach is to use wavelet shrinkage estimators, including VisuShrink based on the universal threshold (Donoho and Johnstone (1994)), SureShrink based on Stein's unbiased risk estimator (SURE) function (Donoho and Johnstone (1995)), Bayesian thresholding rules that use a mixture prior (Abramovich, Sapatinas and Silverman (1998)), and empirical Bayes methods for level-dependent threshold selection (Johnstone and Silverman (2005)).

We consider a function estimation method using overcomplete systems. A subset of the vectors $\{\phi\}_{j\in J}$ of a Banach space $\mathcal{F}$ is called a *complete system* if

$$\left\| \eta - \sum_{j\in J} \beta_j \phi_j \right\| < \epsilon, \quad \forall \eta \in \mathcal{F}, \forall \epsilon > 0,$$

where $\beta_j \in \mathbb{R}$ and $J \in \mathbb{N} \cup \{\infty\}$. Such a complete system is *overcomplete* if removing a vector ϕ_j from the system does not alter the completeness. In other words, an overcomplete system is constructed by adding basis functions to a complete basis (Lewicki and Sejnowski (2000)). The coefficients β_j in the expansion of η with an overcomplete system are not unique, owing to the redundancy intrinsic

in the system. This non-uniqueness property can provide representations that are more parsimonious than those offered by a complete system (Simoncelli et al. (1992)).

The Lévy adaptive regression kernels (LARK) model, first proposed by Tu (2006), is a Bayesian regression model that uses overcomplete systems with Lévy process priors. Tu (2006) showed that the LARK model has sparse representations for η from an overcomplete system and improvements in terms of nonparametric function estimation. Pillai et al. (2007) determined the relationship between the LARK model and a reproducing kernel Hilbert space (RKHS), and Pillai (2008) proved the posterior consistency of the LARK model. Chu, Clyde and Liang (2009) used continuous wavelets as the elements of an overcomplete system. Wolpert, Clyde and Tu (2011) obtained sufficient conditions for LARK models to lie in some Besov space or Sobolev space. Lee, Mano and Lee (2020) devised an extended LARK model with multiple kernels instead of only one type of kernel.

In this study, we develop a fully Bayesian approach with B-spline basis functions as the elements of an overcomplete system , called Lévy adaptive B-Spline regression (LABS). Although the approach of Chu, Clyde and Liang (2009) and the LARK methods are useful tools, the LABS model demonstrates clear advantages over them. First, the approach of Chu, Clyde and Liang (2009) uses wavelet functions to generate an overcomplete system. Note that the wavelets of the Daubechies' family other than the Haar wavelet do not have closed-form expressions. Thus, we need to use a numerical algorithm, namely, the Daubechies-Lagarias pyramidial algorithm (Vidakovic (2009)), to evaluate wavelets at arbitrary points. The Daubechies-Lagarias algorithm requres multiplying a number of $(2N-1)\times(2N-1)$ matrices, where N is the number of vanishing moments of the mother wavelet. In theory, the number of matrices in the product needs to be taken to ∞, and it increases as more precision in the computation is required. In addition, if N is large, the matrix multiplication can be burdensome. In a typical wavelet application of equally spaced data, the Daubechies-Lagarias algorithm can be avoided by using a discrete wavelet transform. However, in the LARK model, the kernels are not equally spaced, and so we need to use a computationally expensive algorithm. The B-spline basis, defined as the convolution of the unit box function, has a simple explicit format in both the time and the frequency domains, which may be useful for statistical analysis. Second, splines are piecewise polynomials, and it is easier to obtain their derivatives and integrals, which may be required in the posterior analysis, than it is for other generating functions, including wavelet and kernel functions. Third, splines provide a flexi-

ble framework that can switch between two extreme cases, namely, the piecewise constant model (degree zero) and the band-limited model (degree infinite), which is not feasible with a wavelet basis and a kernel function. Finally, splines smooth the signal by imposing smoothness, whereas wavelets impose sparsity. Therefore, splines are more natural for smoothing than wavelets are.

The main contributions of this work are as follows.

1. The LABS model can systematically represent the smoothness of functions that vary locally by changing the order of the B-spline basis. The varying degree of the B-spline basis enables the LABS model to adapt to the smoothness of the functions. The LABS model can also be used to construct overcomplete systems with B-spline bases with different types of differentiability. In contrast, Tu (2006), Chu, Clyde and Liang (2009), and Wolpert, Clyde and Tu (2011) use only one type of generating element in overcomplete systems. Using two or more types of generating functions as elements of an overcomplete system is more effective for estimating the mean function η with varying degrees of smoothness.

2. We investigate two theoretical properties of the LABS model. First, the mean function of the LABS model exists in certain Besov spaces, based on the degree of the B-spline basis. Second, the prior of the LABS model has full support on some Besov spaces. Thus, the proposed model extends the range of the smoothness classes of the mean function.

3. We provide empirical results demonstrating that our model performs well in spatially inhomogeneous functions, such as functions with jump discontinuities, sharp peaks, and smooth parts. The LABS model achieves the best results in almost all experiments in comparison with other popular nonparametric function estimation methods. In particular, the LABS model shows remarkable performance when estimating functions with jump discontinuities, outperforming other competing models.

The rest of the paper is organized as follows. In Section 2, we introduce the LARK model and discuss its properties. In Section 3, we propose the LABS model and present an equivalent model with latent variables that make the posterior computation tractable. We present three theorems that show that the function spaces for the proposed model depend on the degree of the B-spline basis, and that the prior has large support in some function spaces. In Section 4, we compare the LABS model with other methods in two simulation studies, and in Section 5, we analyze a real-world data set using the LABS model. The final section

concludes the paper.

2. The LARK Model

In this section, we give a brief introduction to the LARK model. Let Ω be a complete separable metric space, and ν be a positive measure on $\mathbb{R} \times \Omega$, with $\nu(\{0\}, \Omega) = 0$ satisfying the L_1 integrability condition

$$\int \int_{\mathbb{R} \times A} (1 \wedge |\beta|) \nu(d\beta, d\omega) < \infty, \tag{2.1}$$

for each compact set $A \subset \Omega$. The Lévy random measure L with the Lévy measure ν is defined as

$$L(d\omega) = \int_{\mathbb{R}} \beta N(d\beta, d\omega),$$

where N is a Poisson random measure with intensity measure ν. We denote this as $L \sim \text{Lévy}(\nu)$. For any $t \in \mathbb{R}$, the characteristic function of $L(A)$ is

$$\mathbb{E}\left[e^{itL(A)}\right] = \exp\left\{\int \int_{\mathbb{R} \times A} (e^{it\beta} - 1)\nu(d\beta, d\omega)\right\}, \quad \text{for all } A \subset \Omega. \tag{2.2}$$

Let $g(x, \omega)$ be a real-valued function defined on $\mathcal{X} \times \Omega$, where $\mathcal{X}$ is another set. By integrating g with respect to a Lévy random measure L, we define the following real-valued function on $\mathcal{X}$:

$$\eta(x) \equiv L[g(x)] = \int_{\Omega} g(x, \omega) L(d\omega) = \int_{\Omega} \int_{\mathbb{R}} g(x, \omega) \beta N(d\beta, d\omega), \forall x \in \mathcal{X}. \tag{2.3}$$

We call g a *generating function* of η.

When $\nu(\mathbb{R} \times \Omega) = M$ is finite, a Lévy random measure can be represented as $L(d\omega) = \sum_{j \leq J} \beta_j \delta_{\omega_j}$, where J has a Poisson distribution with mean $M > 0$, and $\{(\beta_j, \omega_j)\} \overset{i.i.d.}{\sim} \pi(d\beta_j, d\omega_j) := \nu/M$, for $j = 1, \ldots, J$. In this case, equation (2.3) can be expressed as

$$\eta(x) = \sum_{j=1}^{J} g(x, \omega_j) \beta_j, \tag{2.4}$$

where $\{(\beta_j, \omega_j)\}$ is a random set of the finite support points of a Poisson random measure. If g is bounded, the L_1 integrability condition (2.1) implies the existence of (2.3) for all x. See Lee, Mano and Lee (2020).

If a Lévy measure satisfying (2.1) is infinite, the number of support points of $N(\mathbb{R}, \Omega)$ is infinite, almost surely. Tu (2006) proved that the truncated Lévy

random field $L_\epsilon[g]$ converges in distribution to $L[g]$ as $\epsilon \to 0$, where

$$L_\epsilon[g] = \int\int_{[-\epsilon,\epsilon]^c \times \Omega} g(x,\omega)\beta N(d\beta, d\omega) = \int\int_{\mathbb{R}\times\Omega} g(x,\omega)\beta N_\epsilon(d\beta, d\omega),$$

and N_ϵ is a Poisson measure on $\mathbb{R}$ with mean measure

$$\nu^{(\epsilon)}(d\beta, d\omega) := \nu(d\beta, d\omega) I_{|\beta|>\epsilon}.$$

This truncation is often used as an approximation of the posterior. For posterior computation methods for the Poisson random measure without truncation, see Lee (2007) and Lee and Kim (2004).

Together with the data-generating mechanism (1.1), the LARK model is defined as follows:

$$\begin{aligned} \mathbb{E}[Y|L,\theta] = \eta(x) &\equiv \int_\Omega g(x,\omega)L(d\omega) \\ L|\theta &\sim \text{Lèvy}(\nu) \\ \theta &\sim \pi_\theta(d\theta), \end{aligned}$$

where Lèvy(ν) denotes a Lèvy process that has the characteristic function, and ν is a Lèvy measure satisfying (2.1). Tu (2006) used gamma, symmetric gamma, and symmetric α-stable (SαS) ($0 < \alpha < 2$) Lèvy random fields. The conditional distribution for Y has a hyperparameter θ, and π_θ denotes the prior distribution of θ. The function $g(x,\omega)$ is used as elements of an overcomplete system. Tu (2006) and Lee, Mano and Lee (2020) used the following Gaussian kernel, Laplace kernel, and Haar wavelet as generating functions:

- Haar kernel: $g(x,\omega) := I\left(|(x-\chi)/\lambda| \le 1\right)$
- Gaussian kernel: $g(x,\omega) = \exp\left\{-(x-\chi)^2/2\lambda^2\right\}$
- Laplacian Kernel: $g(x,\omega) = \exp\left\{-|x-\chi|/\lambda\right\}$,

with $\omega := (\chi,\lambda) \in \mathbb{R}\times\mathbb{R}^+ := \Omega$. All of the above generating functions are bounded.

The LARK model can be represented in a hierarchical structure as follows:

$$\begin{aligned} Y_i|\,\eta(\mathbf{x}_i) &\overset{ind.}{\sim} \mathcal{N}(\eta(\mathbf{x}_i), \sigma^2), \\ \eta(\mathbf{x}_i) &= \sum_{j=1}^{J} g(\mathbf{x}_i, \boldsymbol{\omega}_j)\beta_j, \quad J\,|\,\epsilon \sim \text{Pois}(\nu^{(\epsilon)}(\mathbb{R},\Omega)) \end{aligned}$$

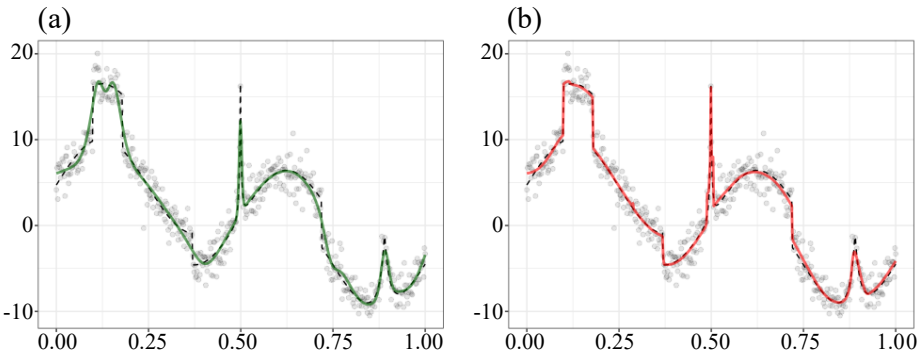

Figure 1. Comparison of curve fitting functions using (a) the LARK model, and (b) the LABS model for the modified HeaviSine data set. The solid lines are estimated functions and the dashed line represents the true function.

$$(\beta_j, \boldsymbol{\omega}_j)|J, \epsilon \overset{i.i.d.}{\sim} \pi(d\beta_j, d\boldsymbol{\omega}_j) := \frac{\nu^{(\epsilon)}(d\beta_j, d\boldsymbol{\omega}_j)}{\nu^{(\epsilon)}(\mathbb{R}, \Omega)}$$

for $j = 1, \ldots, J$. Here J is a random number determined stochastically using a Lèvy random measure, $(\beta_1, \ldots, \beta_J)$ are the unknown coefficients of a mean function, and $(\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_J)$ are the parameters of the generating functions. Because some parameters have varying dimensions, we use the reversible jump Markov chain Monte Carlo (RJMCMC) proposed by Green (1995) to obtain samples from the posterior distribution under the LARK model.

The LARK model stochastically extracts features and finds a compact representation for $\eta(\cdot)$ based on an overcomplete system. That is, it enables functions to be represented by a small number of elements from an overcomplete system. However, the LARK model and most methods for function estimation use only one type of kernel or basis, and can only be used to determine the restricted smoothness of a target function. These models cannot capture all parts of a function that has varying degrees of smoothness. For example, we consider a noisy modified HeaviSine function sampled at $n = 512$ equally spaced points on $[0, 1]$ in Figure 1. The data contain both smooth regions and nonsmooth regions, such as peaks and jumps. As shown in panel (a) of Figure 1, it is difficult for the LARK model with a finite Lèvy measure using a Gaussian kernel to estimate the jumps in the data. Therefore, we propose a new model that adapts the smoothness of a function systematically by using a variety of B-spline basis functions as the generating elements of an overcomplete system.

3. Lévy Adaptive B-Spline Regression

We consider a general type of basis function to generate elements of an overcomplete system, rather than using specific kernel functions, such as the Haar, Laplacian, or Gaussian functions. The LABS model uses B-spline basis functions, which can all systematically express jumps, sharp peaks, and smooth parts of a function.

3.1. B-spline basis

The B-spline basis function consists of piecewise $k \in \mathbb{N} \cup \{0\}$-degree polynomials, where $\mathbb{N}$ is the set of natural numbers. The B-spline of degree $k(\geq 1)$ has $k-1$ continuous derivatives at the knots. In general, the B-spline basis of degree k can be derived using the Cox-de Boor recursion formula:

$$\begin{aligned} B^*_{0,i}(x) &:= \begin{cases} 1 & \text{if} \quad t_i \leq x < t_{i+1} \\ 0 & \text{otherwise} \end{cases} \\ B^*_{k,i}(x) &:= \frac{x - t_i}{t_{i+k} - t_i} B^*_{k-1,i}(x) + \frac{t_{i+k+1} - x}{t_{i+k+1} - t_{i+1}} B^*_{k-1,i+1}(x), \end{aligned} \tag{3.1}$$

where t_i are called knots, which must be in nondescending order $t_i \leq t_{i+1}$ (de Boor (1972), Cox (1972)). The B-spline basis of degree k, $B^*_{k,i}(x)$, then needs $(k+2)$ knots, $(t_i, \ldots, t_{i+k+1})$. For convenience of notation, we redefine the B-spline basis of degree k with a knot sequence $\boldsymbol{\xi}_k := (\xi_{k,1}, \ldots, \xi_{k,k+2})$ as follows:

$$\begin{aligned} B_0(x; \boldsymbol{\xi}_0) &:= \begin{cases} 1 & \text{if} \quad \xi_{0,1} \leq x < \xi_{0,2} \\ 0 & \text{otherwise} \end{cases} \\ B_k(x; \boldsymbol{\xi}_k) &:= \frac{x - \xi_{k,1}}{\xi_{k,(k+1)} - \xi_{k,1}} B_{k-1}(x; \boldsymbol{\xi}^\star_k) + \frac{\xi_{k,(k+2)} - x}{\xi_{k,(k+2)} - \xi_{k,2}} B_{k-1}(x; \boldsymbol{\xi}^{\star\star}), \end{aligned} \tag{3.2}$$

where $\boldsymbol{\xi}^\star_k := (\xi_{k,1}, \xi_{k,2}, \ldots, \xi_{k,(k+1)})$ and $\boldsymbol{\xi}^{\star\star}_k := (\xi_{k,2}, \xi_{k,3}, \ldots, \xi_{k,(k+2)})$.

The B-spline basis functions can vary in terms of shape and smoothness depending on the knot locations and degrees. For example, a B-spline basis function can be a piecewise constant ($k = 0$), linear ($k = 1$), quadratic ($k = 2$) , or cubic ($k = 3$) function. Furthermore, B-spline basis functions with equally spaced knots have a symmetric form on the interval on which they exist. These bases are called uniform B-splines. Examples of B-spline basis functions of different degrees with equally spaced knots are shown in Figure 2.

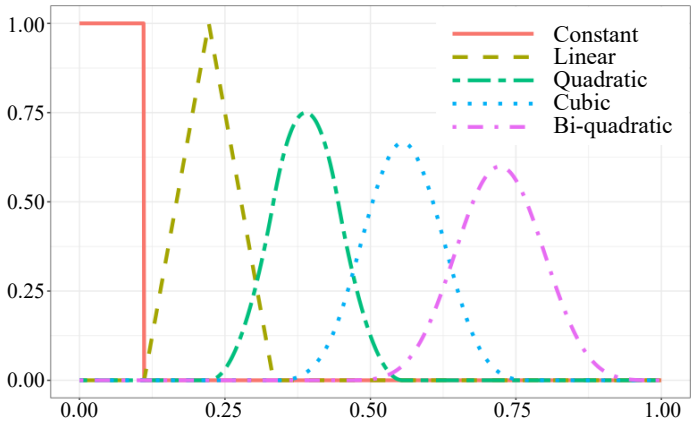

Figure 2. Different shapes of the B-spline basis function by increasing the degree k.

3.2. Model specification

The LARK model with one type of kernel does not well estimate functions with both continuous and discontinuous parts. To improve this, we consider various B-spline basis functions simultaneously to estimate all parts of the unknown function. The new model uses a B-spline basis to systematically generate an overcomplete system with varying degrees of smoothness. For example, the B-spline basis functions of degree zero, one, and two or more are for jumps, sharp peaks, and the smooth parts of the function, respectively.

We express the mean function as a random finite sum:

$$\eta(x) = \sum_{k\in S} \sum_{1\le l\le J_k} B_k(x;\boldsymbol{\xi}_{k,l})\beta_{k,l}, \tag{3.3}$$

where S denotes a subset of degree numbers of B-spline bases and $B_k(x;\boldsymbol{\xi}_k)$ is a B-spline basis of degree k with knots $\boldsymbol{\xi}_k \in \mathcal{X}^{(k+2)} := \Omega$. The generating functions of the LARK model are replaced with the B-spline basis functions. J_k has a Poisson distribution with rate $M_k > 0$, and $\{(\beta_{k,l},\boldsymbol{\xi}_{k,l})\} \overset{i.i.d.}{\sim} \pi_k(d\beta_k, d\boldsymbol{\xi}_k) := \nu_k(d\beta_k, d\boldsymbol{\xi}_k)/\nu_k(\mathbb{R}\times\Omega)$. We assume

$$\pi_k(d\beta_k, d\boldsymbol{\xi}_k) = \mathcal{N}(\beta_k; 0, \phi_k^2)\, d\beta_k \cdot \mathcal{U}(\boldsymbol{\xi}_k; \mathcal{X}^{(k+2)}) d\boldsymbol{\xi}_k.$$

The mean function can be also defined as

$$\eta(x) \equiv \sum_{k\in S} \int_\Omega B_k(x;\boldsymbol{\xi}_k) L_k(d\boldsymbol{\xi}_k). \tag{3.4}$$

The stochastic integral representation of the mean function is determined by

$$L_k \sim \text{Lévy}(\nu_k(d\beta_k, d\boldsymbol{\xi}_k)), \quad \forall k \in S,$$

where $\nu_k(d\beta_k, d\boldsymbol{\xi}_k)$ is a finite Lévy measure satisfying $M_k \equiv \nu_k(\mathbb{R} \times \Omega) < \infty$. Although the Lévy measure ν_k satisfying (2.1) may be infinite, the aforemetioned Poisson integrals and sums are well defined for all bounded measurable compactly supported $B_k(\cdot, \cdot)$ for which for all $k \in S$,

$$\int \int_{\mathbb{R}\times\Omega} (1 \wedge |\beta_k B_k(\cdot; \boldsymbol{\xi}_k)|)\nu_k(d\beta_k, d\boldsymbol{\xi}_k) < \infty. \tag{3.5}$$

We consider only finite Lévy measures in the proposed model. In other words, we restrict our attention to the Lévy measure of a compound Poisson process. The proposed model is more complex than the LARK model with one kernel, and is expected to give a more accurate estimate of the regression function. It can estimate a mean function that has both smooth and peak shapes. The proposed model can be written in hierarchical form as

$$\begin{aligned}
Y_i|x_i &\overset{ind.}{\sim} \mathcal{N}(\eta(x_i), \sigma^2), \\
\eta(x_i) &= \beta_0 + \sum_{k\in S} \sum_{1\le l\le J_k} B_k(x_i; \boldsymbol{\xi}_{k,l})\beta_{k,l}, \quad i = 1, \ldots, n, \\
\sigma^2 &\sim \text{IG}\left(\frac{r}{2}, \frac{rR}{2}\right), \\
J_k &\sim \text{Poi}(M_k), \quad M_k \sim \text{Ga}(a_{\gamma_k}, b_{\gamma_k}), \\
\beta_{k,l} &\overset{i.i.d.}{\sim} \mathcal{N}(0, \phi_k^2), \quad \boldsymbol{\xi}_{k,l} \overset{i.i.d.}{\sim} \mathcal{U}(\mathcal{X}^{(k+2)}), \quad l = 1, \ldots, J_k,
\end{aligned} \tag{3.6}$$

for $k \in S$. We set $\beta_0 = \overline{Y}$ and $\phi_k = 0.5 \times (\max_i\{Y_i\} - \min_i\{Y_i\})$ or $\sqrt{\text{Var}(Y)}$.

The LABS model intrinsically tries to lead to sparse representations by using a Levy process prior. Specifically, the $\log(J_k!)$ terms in the log posterior for the Levy process regularize the number of coefficients of the model. This directly prevents the LABS model from causing over-parametrization issues. Furthermore, the prior distribution on the B-spline coefficients indirectly penalizes the model complexity, such as the Bayesian ridge regression and Bayesian LASSO. Refer to Clyde and Wolpert (2007) and Jang et al. (2017) for further details.

As in the LARK models, becuase some parameters have varying dimensions in the LABS model (3.11), the posterior sampling algorithm for the LABS model has the RJMCMC and Metropolis-Hastings within Gibbs sampling part. Specifically, the RJMCMC procedure for the LABS model is repeated $|S|$ times by

simultaneously considering various types of generating functions. A detailed summary and the pseudocode of the MCMC algorithm for the posterior sampling are described in the Supplementary Material.

3.3. Support of LABS model

In this section, we present three theorems on the support of the LABS model. We first define the modulus of the smoothness and Besov spaces.

Definition 1. Let $0 < p \leq \infty$ and $h > 0$. For $f \in L^p(\mathcal{X})$, the rth-order modulus of the smoothness of f is defined by

$$\omega_r(f,t)_p := \sup_{h \leq t} \|\Delta_h^r f\|_p,$$

where $\Delta_h^r f(x) = \sum_{k=0}^r [r!/\{k!(r-k)!\}](-1)^{r-k} f(x+kh)$, for $x \in \mathcal{X}$ and $x+kh \in \mathcal{X}$.

If $r = 1$, $\omega_1(f,t)_p$ is the modulus of continuity. There exist equivalent definitions in defining Besov spaces. Here, we follow DeVore and Lorentz (1993, Chap. 2.10).

Definition 2. Let $\alpha > 0$ be given, and let r be the smallest integer such that $r > \alpha$. For $0 < p, q < \infty$, the Besov space $\mathbb{B}_{p,q}^{\alpha}$ is the collection of all functions $f \in L_p(\mathcal{X})$ such that

$$|f|_{\mathbb{B}_{p,q}^{\alpha}} = \left(\int_0^{\infty} \{t^{-\alpha} \omega_r(f,t)_p\}^q \frac{dt}{t} \right)^{1/q}$$

is finite. The norm on $\mathbb{B}_{p,q}^{\alpha}$ is defined as

$$\|f\|_{\mathbb{B}_{p,q}^{\alpha}} = \|f\|_p + |f|_{\mathbb{B}_{p,q}^{\alpha}}.$$

The Besov space is a general function space that depends on the smoothness of the functions in $L_p(\mathcal{X})$ and, in particular, can allow smoothness of spatially inhomogeneous functions, including spikes and jumps. The Besov space has three parameters, α, p, and q, where α is the degree of smoothness, p represents that $L_p(\Omega)$ is the function space where the smoothness is measured, and q is a parameter for finer tuning on the degree of smoothness.

Theorem 1. *For fixed $k \in S$ and $\boldsymbol{\xi}_k \in \mathcal{X}^{(k+2)}$, the B-spline basis $B_k(x; \boldsymbol{\xi}_k)$ falls in $\mathbb{B}_{p,q}^{\alpha}(\mathcal{X})$, for all $1 \leq p, q < \infty$ and $\alpha < k + 1/p$.*

The proof of Theorem 1 is provided in the Supplementary Material. For instance, the B-spline basis with degree zero satisfies $B_k(\cdot, \boldsymbol{\xi}_k) \in \mathbb{B}_{p,q}^{\alpha}$ for $\alpha < 1/p$,

the B-spline basis with degree one is in $\mathbb{B}^{\alpha}_{p,q}$, for $\alpha < 1 + 1/p$ and the B-spline basis with degree two falls in $\mathbb{B}^{s}_{p,q}$ for $\alpha < 2 + 1/p$.

Theorem 2 states that the mean function of the LABS model, η, is in a Besov space with smoothness corresponding to the degrees of the B-spline bases used in the LABS model. The proof of the theorem closely follows that of Wolpert, Clyde and Tu (2011). The proof for Theorem 2 is provided in the Supplementary Material.

Theorem 2. *Suppose $\mathcal{X}$ is a compact subset of $\mathbb{R}$. Let ν_k be a Lévy measure on $\mathbb{R} \times \mathcal{X}^{(k+2)}$ that satisfies the following integrability condition:*

$$\int \int_{\mathbb{R}\times\mathcal{X}^{(k+2)}} (1 \wedge |\beta_k|)\nu_k(d\beta_k, d\boldsymbol{\xi}_k) < \infty. \tag{3.7}$$

and let $L_k \sim$ Lévy(ν_k) for all $k \in S$. Define the mean function of the LABS model, $\eta(\cdot) = \sum_{k\in S} \int_{\mathcal{X}^{(k+2)}} B_k(x;\boldsymbol{\xi}_k)L_k(d\boldsymbol{\xi}_k)$, on $\mathcal{X}$, where $B_k(x;\boldsymbol{\xi}_k)$ satisfies (3.7) *for each fixed $x \in \mathcal{X}$. Then, η has the convergent series*

$$\eta(x) = \sum_{k\in S}\sum_{l} B_k(x;\boldsymbol{\xi}_{k,l})\beta_{k,l}, \tag{3.8}$$

where S is a finite set, including the degree numbers of the B-spline bases. Furthermore, η lies in the Besov space $\mathbb{B}^{\alpha}_{p,q}(\mathcal{X})$, with $\alpha < \min(S) + 1/p$, almost surely.

For example, if a zero element is included in S, then the mean function of the LABS model, η, falls in $\mathbb{B}^{\alpha}_{p,q}$, with $\alpha < 1/p$, almost surely, and consists of functions that are no longer continuous. If $S = \{3, 5, 8\}$, then η belongs to $\mathbb{B}^{\alpha}_{p,q}$, with $\alpha < 3 + 1/p$, almost surely. Moreover, it is highly possible that the function spaces for the LABS model are larger than those of the LARK model with one type of kernel function. Specifically, the mean function for the LABS model, with $S = \{0, 1\}$ falls in $\mathbb{B}^{\alpha}_{p,p}$, with $\alpha < 1/p$, almost surely. If the mean function of the LARK model with only one Laplacian kernel falls in $\mathbb{B}^{\alpha}_{p,p}$, with $\alpha < 1+1/p$, then the function spaces of the LABS model with given $\alpha < 1/p$ are larger than those of the LARK model for the range of the smoothness parameter, $1/p < \alpha < 1 + 1/p$, by the properties of the Besov space.

The next theorem shows that the prior distribution of our LABS model has sufficiently large support on the Besov space $\mathbb{B}^{\alpha}_{p,q}$, with $1 \le p, q < \infty$ and $\alpha > 0$. For $\eta_0 \in \mathbb{B}^{\alpha}_{p,q}(\mathcal{X})$, denote the ball around η_0 of radius δ,

$$\bar{b}_{\delta}(\eta_0) = \{\eta : \|\eta - \eta_0\|_p < \delta\},$$

where $\|\cdot\|_p$ is a L_p norm. The proof of Theorem 3 is given in the Supplementary Material.

Theorem 3. *Let $\mathcal{X}$ be a bounded domain in $\mathbb{R}$. Let ν_k be a finite measure on $\mathbb{R}\times\mathcal{X}^{(k+2)}$, and let $L_k \sim Levy(\nu_k)$, for all $k \in S$. Suppose η has a prior Π for the LABS model (3.6). Then, $\Pi\{\bar{b}_\delta(\eta_0)\} > 0$, for every $\eta_0 \in \mathbb{B}^{\alpha}_{p,q}(\mathcal{X})$ and all $\delta > 0$.*

4. Simulation Studies

In this section, we evaluate and compare the performance of the LABS model (3.6) and that of competing methods on simulated data sets. First, we apply the proposed method to four standard examples: the Bumps, Blocks, Doppler and HeaviSine test functions introduced by Donoho and Johnstone (1994). Second, we consider three functions that we created ourselves with jumps and peaks to assess the practical performance of the proposed model.

The simulated data sets are generated from equally spaced x on $\mathcal{X} = [0, 1]$, with sample sizes $n = 128$ and 512. Independent normally distributed noise $\mathcal{N}(0, \sigma^2)$ is added to the true function $\eta(\cdot)$. The root signal-to-noise ratio (RSNR) is defined as

$$\text{RSNR} := \sqrt{\frac{\int_{\mathcal{X}}\{f(x) - \bar{f}\}^2\,dx}{\sigma^2}},$$

where $\bar{f} := (1/|\mathcal{X}|)\int_{\mathcal{X}} f(x)\,dx$ and is set to 3, 5, and 10. We run the LABS model for 200,000 iterations, with the first 100,000 iterations discarded as burn-in, and retain every 10th sample. For comparison between the methods, we compute the mean squared error (MSE) of all methods using 100 replicate data sets for each test function. The average of the posterior curves is used for the estimate of the test function:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\{\eta(x_i) - \hat{\eta}(x_i)\}^2.$$

4.1. Simulation 1: DJ test functions

We carry out a simulation study using the benchmark test functions suggested by Donoho and Johnstone (1994) that are often used in the field of wavelet and nonparametric function estimation. The four Donoho and Johnstone test functions include sharp peaks (Bumps), jump discontinuities (Blocks), oscillating behavior (Doppler), and jumps/peaks in smooth functions (HeaviSine) (see Figure 3).

The hyperparameters and types of basis functions shown in Table S1 of the Supplementary Material were used in (3.6). For Bumps and Doppler, we set the

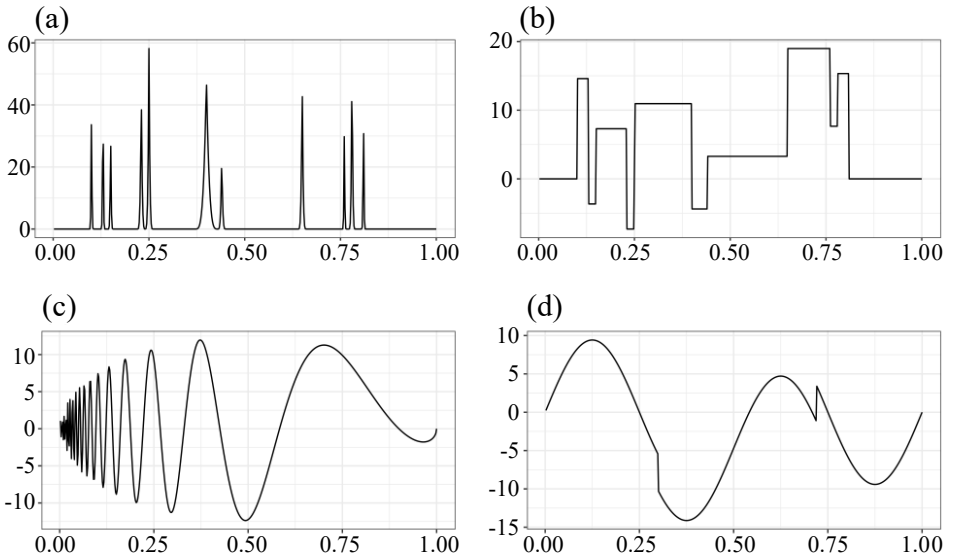

Figure 3. The Donoho and Johnstone test functions: (a) Bumps, (b) Blocks, (c) Doppler, and (d) HeaviSine.

parameter r of the prior distribution for σ^2 to 100 to speed up the convergence. We also considered combinations of B-spline bases, based on the shapes of the test functions. We provide suggestions on how to choose the appropriate degrees in S in the Supplementary Material.

We compared our model with various methods: a B-spline curve of degree two with 50 knots (denoted as BSP-2); a local polynomial regression with automatic smoothing parameter selection (denoted by LOESS); a smoothing spline with the smoothing parameter selected using cross-validation (denoted by SS); a Nadaraya–Watson kernel regression using the Gaussian kernel with bandwidth h that minimizes the CV error (denoted by NWK); an empirical Bayes approach for wavelet shrinkage using a Laplace prior with Daubechies "least asymmetric" (la8) wavelets, except for the Blocks example, where it uses the Haar wavelet (Johnstone and Silverman (2005)) (denoted by EBW); trend filtering with order # based on an optimal regularization parameter (Tibshirani (2014)) (denoted by TF-#); a Gaussian process regression with a Radial basis or Laplacian kernel (denoted by GP-R and GP-L, respectively); a Bayesian curve fitting using piecewise polynomials with $l = \#1$ and $l_0 = \#2$ (Denison, Mallick and Smith (1998a)) (denoted by BPP-#1-#2); Bayesian adaptive spline surfaces with degree # (Francom et al. (2018)) (denoted by BASS-#); and a Lévy adaptive regression with multiple kernels (Lee, Mano and Lee (2020)) (denoted by LARMuK). These com-

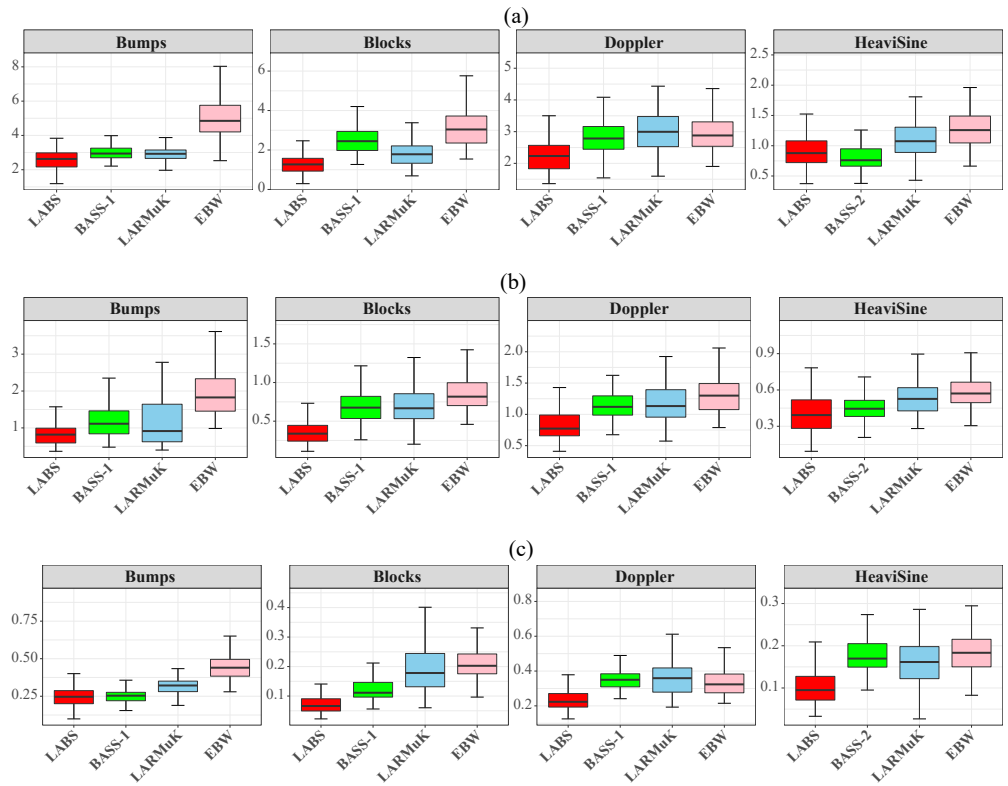

Figure 4. Box plots of the MSEs from the simulation study with $n = 128$ and RSNR = (a) 3, (b) 5, and (c) 10.

petitive models are implemented in R (R Core Team (2020)) using the `splines2`, `fANCOVA`, `EbayesThresh`, `genlasso`, `kernlab`, `miscF`, and `BASS` packages.

Figure 4 and Figure 5 both show that our model outperforms the other methods, in general. The models in the two figures are selected by better outcomes from simulations. More detailed simulation results can be seen in the Supplementary Material. Figure 4 shows that the LABS model is superior to the other models regardless of the noise levels, with $n = 128$. The proposed model also has the smallest average MSE for all test functions, except for the HeaviSine example with RSNR = 3. Similarly, for the sample size $n = 512$, the LABS model performs best in Figure 5, except for the Doppler function, where it is competitive. Our model removes high frequencies in the interval $[0, 0.1]$ and produces a smooth curve within the corresponding domain. In contrast, because there are few data points in the Doppler example with $n = 128$, most models yield similar smooth curves in $[0, 0.1]$. As a result, the LABS model demonstrates excellent numerical

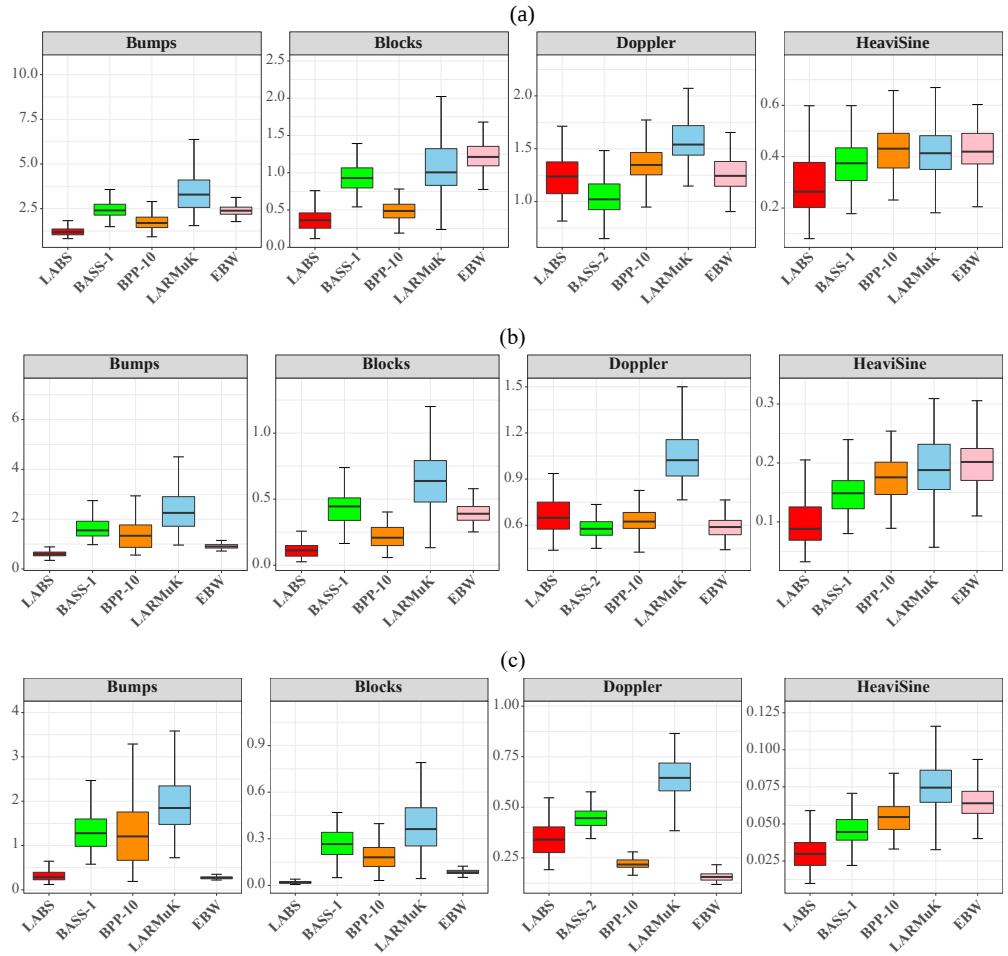

Figure 5. Box plots of the MSEs from the simulation study with $n = 512$ and RSNR = (a) 3, (b) 5, and (c) 10.

performance. For the Blocks example, the LABS model yields the lowest average and standard deviation of MSEs in all scenarios. This suggests that our model has an excellent ability to find jump points. Furthermore, the LABS model consistently outperforms the B-spline regression using only one basis function for four simulated data sets, because its overcomplete systems can be constructed using various combinations of B-spline basis functions; see the Supplementary Material.

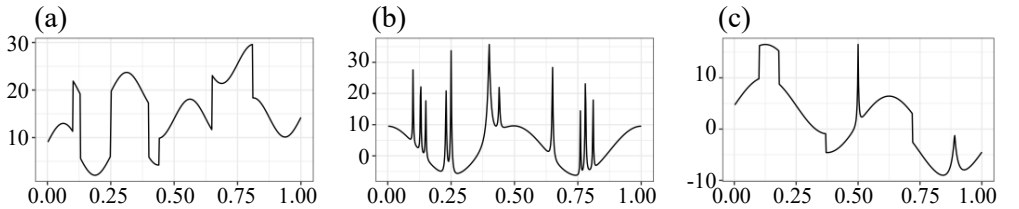

Figure 6. The three test functions used in the second simulation: modified Blocks (left), modified Bumps (center), and modified HeaviSine (right).

4.2. Simulation 2: Smooth functions with jumps and peaks

Our main interest lies in estimating smooth functions with discontinuities such as jumps or sharp peaks, or both. We design three test functions to assess the practical performance of the proposed method. The first and second examples are modified by adding some smooth parts, unlike the original versions of the Bumps and Blocks test functions. Each test function is given by

$$\begin{aligned}\eta_1(x) =& \frac{0.6}{0.92}\{4\mathrm{ssgn}(x-0.1) - 5\mathrm{ssgn}(x-0.13) + 5\mathrm{ssgn}(x-0.25)\\ &- 4.2\mathrm{ssgn}(x-0.4) + 2.1\mathrm{ssgn}(x-0.44) + 4.3\mathrm{ssgn}(x-0.65)\\ &- 4.2\mathrm{ssgn}(x-0.81) + 2\} + 0.2 + \sin(8\pi x),\\ \eta_2(x) =& \{7K_{0.005}(x-0.1) + 5K_{0.07}(x-0.25) + 4.2K_{0.03}(x-0.4)\\ &+ 4.3K_{0.01}(x-0.65) + 5.1K_{0.008}(x-0.78) + 3.1K_{0.1}(x-0.9)\}\\ &+ \cos(4\pi x),\end{aligned}$$

where $\mathrm{sgn}(x) = \mathrm{I}_{(0,\infty)}(x) - \mathrm{I}_{(-\infty,0)}(x)$, $\mathrm{ssgn}(x) = 1 + \mathrm{sgn}(x)/2$, and $K_w(x) := (1 + |x/w|)^{-4}$. Finally, we create a sum of jumps, peaks, and some smoothness. The formula for the last test function is

$$\begin{aligned}\eta_3(x) = 6\sin(4\pi x) &+ 7\left\{1 + \frac{\mathrm{sgn}(x-0.1)}{2}\right\} - 7\left\{1 + \frac{\mathrm{sgn}(x-0.18)}{2}\right\}\\ &- 2\mathrm{sgn}(x-0.37) + 17K_{0.01}(x-0.5) - 3\mathrm{sgn}(x-0.72)\\ &+ 10K_{0.05}(x-0.89).\end{aligned}$$

The functions are displayed in Figure 6.

In these experiments, we use two or more types of B-spline bases as elements of overcomplete systems, because the three functions have different shapes, unlike in previous simulation studies. Hyperparameters are similar to the previous ones. All hyperparameters for the prior distributions are summarized in Table S1 of

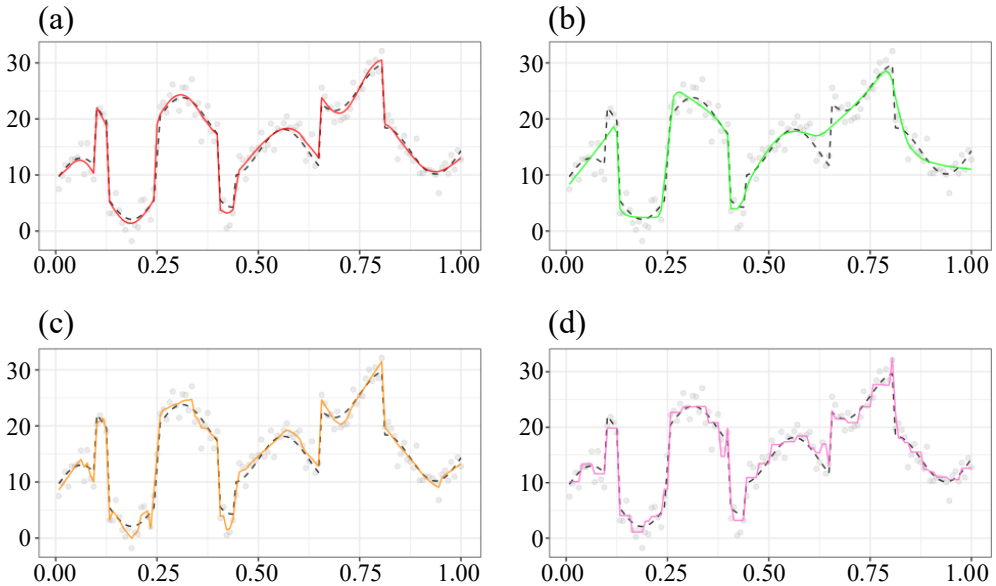

Figure 7. Comparisons of the estimates of a data set generated from the modified Blocks with $n = 128$ and RSNR $= 3$ using (a) LABS, (b) BASS-1, (c) BPP-10, and (d) EBW. Dashed lines represent true curves, and solid lines represent estimates of the curves.

the Supplementary Material. This time, we only compare our model with the BPP, BASS, EBW, TF, and LARMuK models, which exhibit relatively good performance in some test functions of Simulation 1.

Table S12 in the Supplementary Material shows that the LABS model has the best outcomes when the sample size is 128, which is difficult to estimate. Furthermore, when $n = 512$, Table S13 in the Supplementary Material shows that the LABS model performs well in most cases, with either the lowest or the second lowest average MSE values across 100 replicates. In particular, the LABS model outperforms its competitors in the modified Blocks example, irrespective of the sample size and the noise level as expected. The BASS-2 model performs worst, because it does not estimate many of the jumps and peak points for the given test functions. Figure 7 shows that the LABS model can overcome the noise and adapt to smooth functions with discontinuities, such as jumps or sharp peaks, or both.

5. Real-Data Analysis: Fine Particulate Matter in Seoul

We now apply the LABS model (3.6) to an air pollution data set containing the daily maximum value of concentrations of fine particulate matter (PM2.5)

in Seoul. This data set exhibits wildly varying patterns that may have jumps or peaks. These fluctuating patterns are expected to further illustrate the features of the LABS model.

We set the hyperparameter values of the proposed model as follows: $a_J = 5$, $b_J = 1$, $r = 0.01$, and $R = 0.01$. In this analysis, we practically choose $S = \{0, 1, 2\}$, because the true curve of the real-data is unknown, and may have varying smoothness. We run it 200,000 times with a burn-in of 100,000 and thin by 10 to achieve convergence of the MCMC algorithm. We also compare the performance of our model with that of other good methods in the simulated studies.

Fine dust has become a national issue, with the result that much research has been conducted on fine particulate matter (PM2.5). According to these studies, Korea's fine dust particles originate both from within the country and from external sources, such as China. Many factors cause PM2.5 concentration to rapidly rise or fall, and make it difficult to accurately predict its behavior.

We estimate the unknown function of daily maximum concentrations of PM2.5 in Seoul. The PM2.5 data set collected from the AIRKOREA (`https://www.airkorea.or.kr`) includes 1261 daily maximum values of PM2.5 concentration from January 1, 2015, to June 30, 2018. We removed all observations that have missing values.

Figure 8 displays the daily fluctuations and seasonality. PM2.5 concentrations are higher in winter and spring than they are in summer and fall. We take advantage of combinations of basis functions, $S = \{0, 1, 2\}$ to grasp the characteristics of the PM2.5 data with multiple jumps and peak points. As shown in Figure 8, the four methods represent different estimated lines of the unknown mean function and pick features of the data in their own way. Interestingly, LABS, BASS-1, and BPP-10 react differently in terms of detecting peaks, jumps, and smooth parts of the PM2.5 data. The GP-R reflects seasonality, but does not capture these features.

We also compute the average and standard deviation of the cross-validated errors of LABS, BPP-10, BASS-1, LARMuK, and GP-R, which are given in Table 1. The LABS model has the lowest cross-validation error among all methods. Moreover, a comparably low standard deviation of the LABS model supports that it has more stable performance when estimating any shape of function, because it uses all three types of B-spline bases.

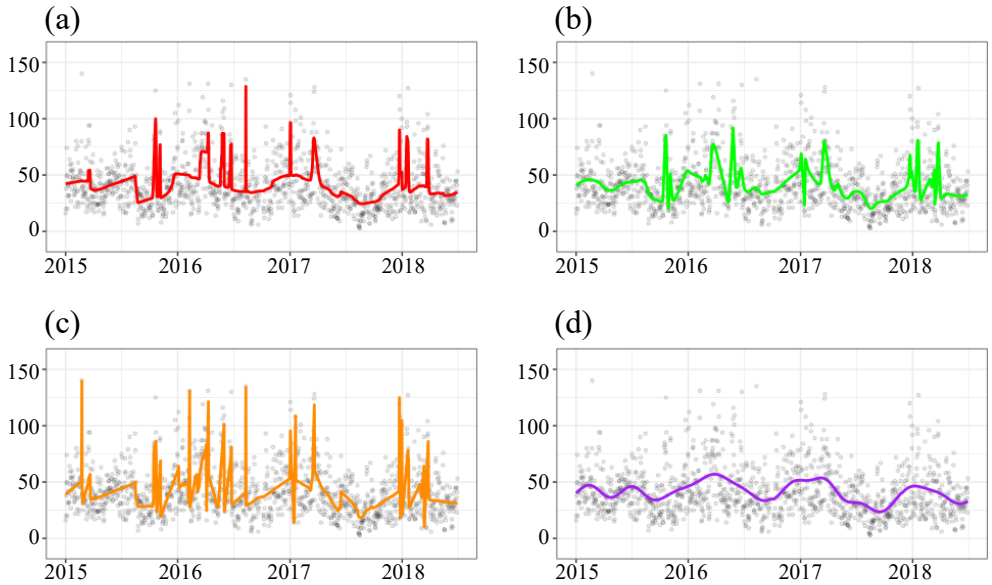

Figure 8. Posterior mean of the mean function on the PM2.5 data set using four models: (a) LABS, (b) BASS-1, (c) BPP-10, and (d) GP-R.

	LABS	BASS-1	BPP-10	LARMuK	GP-R
Mean	**384.8863**	393.6049	398.17	399.6718	436.2286
Standard Deviation	56.88069	60.38016	58.63784	53.02499	67.98722

Table 1. Mean and standard deviation for the error rate of 10-fold cross-validation on the Seoul PM2.5 data set.

6. Conclusion

We have proposed general function estimation methodologies using B-spline basis functions as the elements of an overcomplete system. A B-spline basis can systematically represent functions with varying smoothness, because it has nice properties, such as local support and differentiability. The overcomplete system and Lévy random measure enable a function with both continuous and discontinuous parts to capture all features of the unknown regression function. Simulation studies and a real-data analysis show that the proposed model outperforms competing models. We also showed that the prior has full support in certain Besov spaces. The prominent limitation of the LABS model is the slow mixing of the MCMC algorithm. In future work, we will develop an efficient algorithm for the LABS model and extend it to include multivariate analyses.

Supplementary Material

The online Supplementary Material contains proofs of the theorems in Section 3, the model hyperparameters used for all experiments, additional simulation results, and details about the steps of the MCMC algorithm and the derivation of the full conditionals.

Acknowledgments

The authors are grateful to the editor-in-chief, associate editor, and two referees for their helpful suggestions and comments that led to significant improvements to this work.

References

Abramovich, F., Sapatinas, T. and Silverman, B. W. (1998). Wavelet thresholding via a Bayesian approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **60**, 725–749.

Chu, J.-H., Clyde, M. A. and Liang, F. (2009). Bayesian function estimation using continuous wavelet dictionaries. *Statistica Sinica* **19**, 1419–1438.

Clyde, M. A. and Wolpert, R. L. (2007). Nonparametric function estimation using overcomplete dictionaries. *Bayesian Statistics* **8**, 91–114.

Cox, M. G. (1972). The numerical evaluation of B-splines. *IMA Journal of Applied Mathematics* **10**, 134–149.

Crainiceanu, C. M., Ruppert, D., Carroll, R. J., Joshi, A. and Goodner, B. (2007). Spatially adaptive Bayesian penalized splines with heteroscedastic errors. *Journal of Computational and Graphical Statistics* **16**, 265–288.

de Boor, C. (1972). On calculating with B-splines. *Journal of Approximation Theory* **6**, 50–62.

Denison, D., Mallick, B. and Smith, A. (1998a). Automatic Bayesian curve fitting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **60**, 333–350.

Denison, D. G., Mallick, B. K. and Smith, A. F. (1998b). Bayesian MARS. *Statistics and Computing* **8**, 337–346.

DeVore, R. A. and Lorentz, G. G. (1993). *Constructive Approximation*. Springer-Verlag

DiMatteo, I., Genovese, C. R. and Kass, R. E. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika* **88**, 1055–1071.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association* **90**, 1200–1224.

Donoho, D. L. and Johnstone, J. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**, 425–455.

Francom, D., Sansó, B., Kupresanin, A. and Johannesson, G. (2018). Sensitivity analysis and emulation for functional data using Bayesian adaptive splines. *Statistica Sinica* **28**, 791–816.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics* **19**, 1–67.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732.

Jang, P. A., Loeb, A. E., Davidow, M. B. and Wilson, A. G. (2017). Scalable Lévy process priors for spectral kernel learning. In *Proceedings of the 31st Conference on Neural Information Processing Systems* (Edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett). Curran Associates, Inc.

Johnstone, I. M. and Silverman, B. W. (2005). Empirical Bayes selection of wavelet thresholds. *The Annals of Statistics* **33**, 1700–1752.

Lee, J. (2007). Sampling methods of neutral to the right processes. *Journal of Computational and Graphical Statistics* **16**, 656–671.

Lee, J. and Kim, Y. (2004). A new algorithm to generate beta processes. *Computational Statistics & Data Analysis* **47**, 441–453.

Lee, Y., Mano, S. and Lee, J. (2020). Bayesian curve fitting for discontinuous functions using an overcomplete system with multiple kernels. *Journal of the Korean Statistical Society* **49**, 1–21.

Lewicki, M. S. and Sejnowski, T. J. (2000). Learning overcomplete representations. *Neural Computation* **12**, 337–365.

Liu, Z. and Guo, W. (2010). Data driven adaptive spline smoothing. *Statistica Sinica* **20**, 1143–1163.

Luo, Z. and Wahba, G. (1997). Hybrid adaptive splines. *Journal of the American Statistical Association* **92**, 107–116.

Pillai, N. S. (2008). *Lévy Random Measures: Posterior Consistency and Applications*. PhD dissertation. Duke University.

Pillai, N. S., Wu, Q., Liang, F., Mukherjee, S. and Wolpert, R. L. (2007). Characterizing the function space for Bayesian kernel models. *Journal of Machine Learning Research* **8**, 1769–1797.

Pintore, A., Speckman, P. and Holmes, C. C. (2006). Spatially adaptive smoothing splines. *Biometrika* **93**, 113–125.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Ruppert, D. and Carroll, R. J. (2000). Theory & methods: Spatially-adaptive penalties for spline fitting. *Australian & New Zealand Journal of Statistics* **42**, 205–223.

Simoncelli, E. P., Freeman, W. T., Adelson, E. H. and Heeger, D. J. (1992). Shiftable multiscale transforms. *IEEE Transactions on Information Theory* **38**, 587–607.

Smith, M. and Kohn, R. (1996). Nonparametric regression using Bayesian variable selection. *Journal of Econometrics* **75**, 317–343.

Tibshirani, R. J. (2014). Adaptive piecewise polynomial estimation via trend filtering. *The Annals of Statistics* **42**, 285–323.

Tu, C. (2006). *Bayesian Nonparametric Modeling Using Levy Process Priors with Applications for Function Estimation, Time Series Modeling and Spatio-Temporal Modeling*. PhD dissertation, Duke University.

Vidakovic, B. (2009). *Statistical Modeling by Wavelets*. John Wiley & Sons.

Wang, X., Du, P. and Shen, J. (2013). Smoothing splines with varying smoothing parameter. *Biometrika* **100**, 955–970.

Wolpert, R. L., Clyde, M. A. and Tu, C. (2011). Stochastic expansions using continuous dictionaries: Lévy adaptive regression kernels. *The Annals of Statistics* **39**, 1916–1962.

Yang, L. and Hong, Y. (2017). Adaptive penalized splines for data smoothing. *Computational Statistics & Data Analysis* **108**, 70–83.

Zhou, S. and Shen, X. (2001). Spatially adaptive regression splines and accurate knot selection schemes. *Journal of the American Statistical Association* **96**, 247–259.

Sewon Park

Security Algorithm Lab, Samsung SDS, Seoul 05510, Korea.

E-mail: swpark0413@gmail.com

Hee-Seok Oh

Department of Statistics, Seoul National University, Seoul 08826, Korea.

E-mail: heeseok@stats.snu.ac.kr

Jaeyong Lee

Department of Statistics, Seoul National University, Seoul 08826, Korea.

E-mail: jylc@snu.ac.kr

(Received August 2021; accepted April 2022)