Statistica Sinica Preprint No: SS- 2021-0288	
Title	Lévy Adaptive B-spline Regression via Overcomplete Systems
Manuscript ID	SS- 2021-0288
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202021.0288
Complete List of Authors	Sewon Park, Hee-Seok Oh and Jaeyong Lee
Corresponding Authors	Sewon Park
E-mails	swpark0413@gmail.com
Notice: Accepted version subject to English editing.	

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Statistica Sinica

Lévy Adaptive B-spline Regression via Overcomplete Systems

Sewon Park, Hee-Seok Oh and Jaeyong Lee

SAMSUNG SDS and Seoul National University

Abstract: The estimation of functions with varying degrees of smoothness is a challenging problem in the nonparametric function estimation. In this paper, we propose the LABS (Lévy Adaptive B-Spline regression) model, an extension of the LARK (Lévy Adaptive Regression Kernels) models, for the estimation of functions with varying degrees of smoothness. LABS model is a LARK 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 express systematically functions with varying degrees of smoothness. By changing the orders of the B-spline basis, LABS can systematically adapt the smoothness of functions, i.e., jump discontinuities, sharp peaks, etc. Results of simulation studies and real data examples support that this model catches not only smooth areas but also jumps and sharp peaks of functions. The proposed model also has the best performance in almost all examples. Finally, we provide theoretical results that the mean function for the LABS model belongs to the certain Besov spaces based on the orders of the B-spline basis and that the prior of the model has the full support on the Besov spaces.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Key words and phrases: Nonparametric Function Estimation, Lévy Random Measure, Besov Space, Reversible Jump Markov Chain Monte Carlo.

1. Introductionn

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

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

and η is an unknown real-valued function which maps $\mathbb{R}$ to $\mathbb{R}$. We wish to estimate the mean function η that may have varying degrees of smoothness including discontinuities. In nonparametric function estimation, we often face smooth curves except for a finite number of jump discontinuities and sharp peaks, which are common in many climate and economic datasets. Heavy rainfalls cause a sudden rise in the water level of a river. The COVID-19 epidemic brought about a sharp drop in unemployment rates. Policy-makers' decisions can give rise to abrupt changes. For instance, the United States Congress passed the National Minimum Drinking Age Act in 1984, which has been debated over several decades in the United States, establishing 21 as the minimum legal alcohol purchase age. This act caused a sudden rise in mortality for young Americans around 21. The abrupt changes can provide us with meaningful information on these issues, and it is important to grasp the changes.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

There has been much research into the estimation of local smoothness of the 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 et al. (2006), Liu and Guo (2010), and Wang et al. (2013) modeled the smoothing parameter of smoothing spline to vary over the domain. Ruppert and Carroll (2000), Crainiceanu et al. (2007), and Yang and Hong (2017) suggested the penalized splines based on the local penalty that adapts to spatial heterogeneity in the regression function. The second approach is the adaptive free-knot splines that choose the number and location 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 and proposed optimal knot selection schemes introducing the knot relocation step. Smith and Kohn (1996), Denison et al. (1998a), Denison et al. (1998b), and DiMatteo et al. (2001) studied Bayesian estimation of free knot splines using 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 by utilizing

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

a mixture prior (Abramovich et al., 1998), and empirical Bayes methods for level-dependent threshold selection (Johnstone and Silverman, 2005).

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

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

where $\beta_j \in \mathbb{R}$ and $J \in \mathbb{N} \cup \infty$. Such a complete system is *overcomplete* if removal of 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). Coefficients β_j in the expansion of η with an overcomplete system are not unique owing to the redundancy intrinsic in the overcomplete system. The non-uniqueness property can provide more parsimonious representations than those with a complete system (Simoncelli et al., 1992).

The Lévy Adaptive Regression Kernels (LARK) model, first proposed by Tu (2006), is a Bayesian regression model utilizing overcomplete systems with Lévy process priors. Tu (2006) showed the LARK model had sparse representations for η from an overcomplete system and improvements in nonparametric function estimation. Pillai et al. (2007) found out the relationship between the LARK model and a reproducing kernel Hilbert space

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

(RKHS), and Pillai (2008) proved the posterior consistency of the LARK model. Chu et al. (2009) used continuous wavelets as the elements of an overcomplete system. Wolpert et al. (2011) obtained sufficient conditions for LARK models to lie in some Besov space or Sobolev space. Lee et al. (2020) devised an extended LARK model with multiple kernels instead of only one type kernel.

In this paper, we develop a fully Bayesian approach with B-spline basis functions as the elements of an overcomplete system and call it the Lévy Adaptive B-Spline regression (LABS). For function estimation, while Chu et al. (2009)'s approach and LARK methods have still been nice tools, the LABS has clear advantages over them. First, Chu et al. (2009)'s approach is based on the wavelet functions to generate an overcomplete system. Note that the wavelets of Daubechies' family except the Haar wavelet do not have closed form expressions and one needs to resort to a numerical algorithm, Daubechies-Lagarias pyramidial algorithm (Vidakovic, 2009) to evaluate wavelets at arbitrary points. For the Daubechies-Lagarias algorithm, a number of $(2N-1)\times(2N-1)$ matrices need to be multiplied, 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 gets larger as more precision in the computation is required. Also, if N is large, the

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

matrix multiplication can be burdensome. In a typical wavelet application of equally-spaced data, the Daubechies-Lagarias algorithm can be avoided by discrete wavelet transform, but in the LARK model the kernels are not equally-spaced and the computationally expensive algorithm needs to be employed. However, the B-spline basis, defined as the convolution of the unit box function, has a simple explicit format in both the time and frequency domain, which may be useful for further 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 other generating functions including wavelet and kernel functions. Third, splines provide a flexible framework that can switch two extreme cases, the piecewise constant model (degree 0) and the bandlimited model (degree infinite), which is not feasible with wavelet basis and kernel function. Finally, splines achieve smoothing of the signal by imposing smoothness, whereas wavelets impose sparsity. Therefore, splines are more natural for smoothing than wavelets.

Our main contributions of this work can be outlined as follows.

1. The LABS model can systematically represent the smoothness of functions varying locally by changing the orders of the B-spline basis. The varying degree of B-spline basis enables the LABS model to adapt to

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

the smoothness of functions. The LABS model can also construct overcomplete systems with B-spline bases with different types of differentiability, while Tu (2006), Chu et al. (2009) and Wolpert et al. (2011) can use only one type of generating elements in the overcomplete system. 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 types of degrees of B-spline basis. Second, the prior of the LABS model has full support on some Besov spaces. Thus, the proposed LABS model extends the range of smoothness classes of the mean function.

3. We provide empirical results demonstrating that our model performs well in the spatially inhomogeneous functions such as the functions with both jump discontinuities, sharp peaks, and smooth parts. The LABS model achieved the best results in almost every experiments compared to the popular nonparametric function estimation methods. In particular, the LABS model showed remarkable performance

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

in estimating functions with jump discontinuities and outperformed other competing models.

The rest of the paper is organized as follows. In section 2, we introduce the Lévy Adaptive Regression Kernels 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 the function spaces for the proposed model depend upon the degree of B-spline basis and that the prior has large support in some function spaces. In section 4, the LABS model is compared with other methods in two simulation studies and in section 5 an real-world dataset is analysed using the LABS model. In the last section, we discuss some improvements and possible extensions of the proposed model.

2. Lévy adaptive regression kernels

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 L_1 integrability condition,

$$\int \int_{\mathbb{R}\times A} (1 \wedge |\beta|)\nu(d\beta, d\omega) < \infty, \tag{2.2}$$

for each compact set $A \subset \Omega$. The Lévy random measure L with 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 $L \sim \text{Lévy}(\nu)$. For any $t \in \mathbb{R}$, the characteristic function of $L(A)$ is

$$\mathbb{E}\left[e^{itL(A)}\right] = \exp\left\{\int\int_{\mathbb{R}\times A}(e^{it\beta} - 1)\nu(d\beta, d\omega)\right\}, \quad \text{for all } A \subset \Omega. \tag{2.3}$$

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 a real-valued function on $\mathcal{X}$:

$$\eta(x) \equiv L[g(x)] = \int_{\Omega} g(x, \omega)L(d\omega) = \int_{\Omega}\int_{\mathbb{R}} g(x, \omega)\beta N(d\beta, d\omega), \forall x \in \mathcal{X}. \tag{2.4}$$

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 $L(d\omega) = \sum_{j\le J} \beta_j \delta_{\omega_j}$, where J has a Poisson distribution with mean $M > 0$ and $\{(\beta_j, \omega_j)\} \overset{iid}{\sim} \pi(d\beta_j, d\omega_j) := \nu/M$, $j = 1, \ldots, J$. In this case, equation (2.4) can be expressed as

$$\eta(x) = \sum_{j=1}^{J} g(x, \omega_j)\beta_j, \tag{2.5}$$

where $\{(\beta_j, \omega_j)\}$ is the random set of finite support points of a Poisson

random measure. If g is bounded, L_1 integrability condition (2.2) implies the existence of (2.4) for all x. See Lee et al. (2020).

If a Lévy measure satisfying (2.2) is infinite, the number of the 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 data generating mechanism (1.1), the LARK model is defined as follows:

$$\mathbb{E}[Y|L,\theta] = \eta(x) \equiv \int_\Omega g(x,\omega)L(d\omega)$$

$$L|\theta \sim \text{Lèvy}(\nu)$$

$$\theta \sim \pi_\theta(d\theta),$$

where Lèvy(ν) denotes the Lèvy process which has the characteristic function and ν is a Lèvy measure satisfying (2.2). Tu (2006) used gamma, sym-

metric 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 generating function $g(x, \omega)$ is used as elements of an overcomplete system.Tu (2006) and Lee et al. (2020) used the Gaussian kernel, the Laplace kernel, and Haar wavelet as generating functions:

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

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

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

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

for $j = 1, \ldots, J$. J is the random number that is stochastically determined by 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. To obtain samples from the posterior distribution under the LARK model, the reversible jump Markov chain Monte Carlo (RJMCMC) proposed by Green (1995) is used because some parameters have varying dimensions.

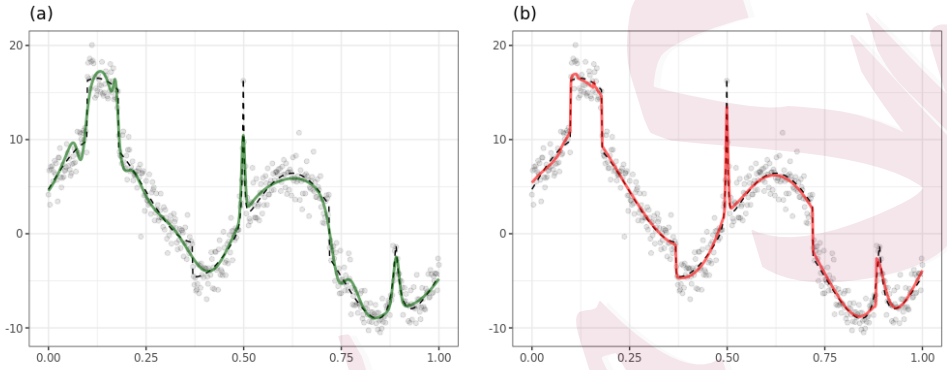

Figure 1: Comparison of curve fitting functions with (a) LARK, and (b) LABS model for the modified Heavisine dataset. The solid lines are estimated functions and the dashed line represents the true function.

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 the small number of elements from an overcomplete system. However, both the LARK model and most methods for function estimation use only one type of kernel or basis and can find out the restricted smoothness of the target function. These models

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

cannot afford to capture all parts of the function with various 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 contains both smooth and non-smooth 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 Gaussian kernel to estimate jump parts of the data. We, therefore, propose a new model which can adapt the smoothness of 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 as the generating elements of an overcomplete system instead of specific kernel functions such as Haar, Laplacian, and Gaussian. The LABS model uses B-spline basis functions which can all systematically express jumps, sharp peaks, and smooth parts of the 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 $k(\geq 1)$

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

has $k-1$ continuous derivatives at the knots. In general, the B-spline basis of degree k can be derived utilizing the Cox-de Boor recursion formula:

$$B^*_{0,i}(x) := \begin{cases} 1 & \text{if} \quad t_i \le x < t_{i+1} \\ 0 & \text{otherwise} \end{cases} \tag{3.6}$$

$$B^*_{k,i}(x) := \frac{x - t_i}{t_{i+k} - t_i} B^*_{k-1,i}(x) + \frac{t_{i+k+1} - x}{t_{i+k+1} - t_{i+1}} B^*_{k-1,i+1}(x),$$

where t_i are called knots which must be in non-descending order $t_i \le t_{i+1}$ (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.

$$B_0(x; \boldsymbol{\xi}_0) := \begin{cases} 1 & \text{if} \quad \xi_{0,1} \le x < \xi_{0,2} \\ 0 & \text{otherwise} \end{cases} \tag{3.7}$$

$$B_k(x; \boldsymbol{\xi}_k) := \frac{x - \xi_{k,1}}{\xi_{k,(k+1)} - \xi_{k,1}} B_{k-1}(x; \boldsymbol{\xi}^\star_k) + \frac{\xi_{k,(k+2)} - x}{\xi_{k,(k+2)} - \xi_{k,2}} B_{k-1}(x; \boldsymbol{\xi}^{\star\star}),$$

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 have a variety of shapes and smoothness determined by knot locations and degrees. For example, a B-spline basis function can be a piecewise constant ($k = 0$), linear ($k = 1$), quadratic ($k = 2$) , and cubic ($k = 3$) functions. Furthermore, the B-spline basis functions with equally spaced knots have the symmetric form on the interval where they exist. These bases are called uniform B-splines. Examples of

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

the B-spline basis functions of different degrees with equally spaced knots are shown in Figure 2.

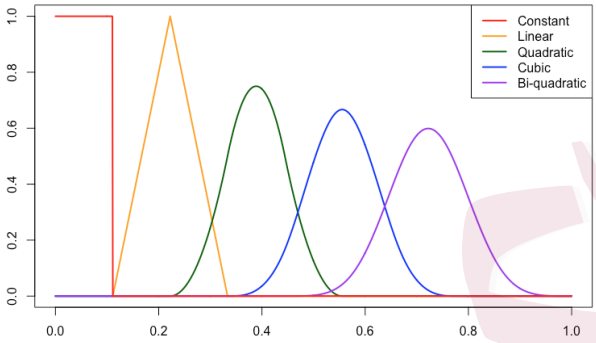

Figure 2: Different shapes of the B-spline basis function by increasing the degree k.

3.2 Model Specification

The LARK model with one type of kernel can not estimate well functions with both continuous and discontinuous parts. To improve this, we consider various a B-spline basis functions simultaneously for estimating all parts of the unknown function. The new model uses B-spline basis to systematically generate an overcomplete system with varying degrees of smoothness. For example, the B-spline basis functions of degrees 0, 1 and 2 or more are for jumps, sharp peaks and smooth parts of the function, respectively.

We consider the mean function can be expressed as a random finite

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

sum:

$$\eta(x) = \sum_{k \in S} \sum_{1 \le l \le J_k} B_k(x; \boldsymbol{\xi}_{k,l}) \beta_{k,l}, \tag{3.8}$$

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

$$\eta(x) \equiv \sum_{k \in S} \int_{\Omega} B_k(x; \boldsymbol{\xi}_k) L_k(d\boldsymbol{\xi}_k). \tag{3.9}$$

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 (2.2) may be infinite, the Poisson integrals and sums aboves are well defined for all bounded measurable compactly-supported $B_k(\cdot, \cdot)$ for which for all $k \in S$,

$$\int \int_{\mathbb{R} \times \Omega} (1 \wedge |\beta_k B_k(\cdot; \boldsymbol{\xi}_k)|) \nu_k(d\beta_k, d\boldsymbol{\xi}_k) < \infty. \tag{3.10}$$

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

In this paper, 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 expected to give a more accurate estimate of the regression function. It can estimate a mean function having both smooth and peak shapes. The proposed model can be written in hierarchical form as

$$\begin{aligned} Y_i|x_i &\overset{ind}{\sim} \mathcal{N}(\eta(x_i), \sigma^2), \\ \eta(x_i) = \beta_0 + \sum_{k\in S}\sum_{1\le l\le J_k} B_k(x_i;\boldsymbol{\xi}_{k,l})\beta_{k,l}, &\quad i = 1,\cdots,n, \\ \sigma^2 &\sim \mathrm{IG}\left(\frac{r}{2}, \frac{rR}{2}\right), \\ J_k \sim \mathrm{Poi}(M_k), &\quad M_k \sim \mathrm{Ga}(a_{\gamma_k}, b_{\gamma_k}), \\ \beta_{k,l} \overset{iid}{\sim} \mathcal{N}(0,\phi_k^2), &\quad \boldsymbol{\xi}_{k,l} \overset{iid}{\sim} \mathcal{U}(\mathcal{X}^{(k+2)}), \quad l = 1,\cdots,J_k, \end{aligned} \tag{3.11}$$

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 $\sqrt{\mathrm{Var}(Y)}$.

The LABS model intrinsically tries to lead sparse representations via a Levy process prior. Specifically, $\log(J_k!)$ 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 model complexity such as Bayesian ridge regression and Bayesian

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

LASSO. Refer to Clyde and Wolpert (2007) and Jang et al. (2017) for further details.

As in the LARK models, since 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. Especially, the RJMCMC procedure for the LABS is repeated $|S|$ times by considering simultaneously various types of generating functions. A detailed summary and pseudo-code 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 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 smoothness of f is defined by

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

where $\Delta_h^r f(x) = \sum_{k=0}^{r} \frac{r!}{k!(r-k)!}(-1)^{r-k} f(x+kh)$ 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. We follow DeVore and Lorentz (1993, Chap. 2.10).

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

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 $\mathbb{B}^{\alpha}_{p,q}$ is the collection of all functions $f \in L_p(\mathcal{X})$ such that

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

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

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

The Besov space is a general function space depending on the smoothness of functions in $L_p(\mathcal{X})$ and especially 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 smoothness is measured, and q is a parameter for a 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 $\mathbb{B}^{\alpha}_{p,q}(\mathcal{X})$ 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 0 satisfies $B_k(\cdot, \boldsymbol{\xi}_k) \in \mathbb{B}^{\alpha}_{p,q}$ for $\alpha < 1/p$, the B-spline basis with degree 1 is in $\mathbb{B}^{\alpha}_{p,q}$ for $\alpha < 1 + 1/p$ and the B-spline basis with degree 2 falls in $\mathbb{B}^{s}_{p,q}$ for $\alpha < 2 + 1/p$.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

The following theorem describes the mean function of the LABS model, η, is in a Besov space with smoothness corresponding to degrees of B-spline bases used in the LABS model. The proof of the theorem closely follows that of Wolpert et al. (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,*

$$\int\int_{\mathbb{R}\times\mathcal{X}^{(k+2)}} (1 \wedge |\beta_k|)\nu_k(d\beta_k, d\boldsymbol{\xi}_k) < \infty. \tag{3.12}$$

and $L_k \sim$ 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 (3.12) for each fixed $x \in \mathcal{X}$. Then, η has the convergent series

$$\eta(x) = \sum_{k\in S}\sum_{l} B_k(x;\boldsymbol{\xi}_{k,l})\beta_{k,l} \tag{3.13}$$

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

For example, if a zero element is included in S then the mean function of the LABS, η falls in $\mathbb{B}^{\alpha}_{p,q}$ with $\alpha < \frac{1}{p}$ almost surely, which consists of functions no longer continuous. If $S = \{3, 5, 8\}$, then, η belongs to $\mathbb{B}^{\alpha}_{p,q}$

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

with $\alpha < 3 + \frac{1}{p}$ almost surely. Moreover, it is highly possible that the function spaces for the LABS model may be larger than those of the LARK model using 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 $\alpha < \frac{1}{p}$ almost surely. If that of the LARK model using only one Laplacian kernel falls in $\mathbb{B}^{\alpha}_{p,p}$ with $\alpha < 1 + \frac{1}{p}$, then the function spaces of the LABS model with given $\alpha < \frac{1}{p}$ are larger than those of the LARK model for the range of smoothness parameter, $\frac{1}{p} < \alpha < 1 + \frac{1}{p}$ 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 $1 \leq p, q < \infty$ 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 $L_k \sim Levy(\nu_k)$ for all $k \in S$. Suppose η has a prior Π for the LABS model* (3.11). *Then, $\Pi(\bar{b}_{\delta}(\eta_0)) > 0$ for every $\eta_0 \in \mathbb{B}^{\alpha}_{p,q}(\mathcal{X})$ and all $\delta > 0$.*

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

4. Simulation Studies

In this section, we evaluate the performance of the LABS model (3.11) and competing methods on simulated data sets. First, we apply the proposed method to four standard examples: 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's on $\mathcal{X} = [0, 1]$ with sample sizes $n = 128$ and 512. Independent normally distributed noises $\mathcal{N}(0, \sigma^2)$ are 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 $\bar{f} := \frac{1}{|\mathcal{X}|} \int_{\mathcal{X}} f(x) \, dx$ and set at 3, 5 and 10. We also 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.

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

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

4.1 Simulation 1 : DJ test functions

We carry out a simulation study using the benchmark test functions suggested by Donoho and Johnstone (1994) often used in the field of wavelet and nonparametric function estimation. The Donoho and Johnstone test functions consist of four functions called Bumps, Blocks, Doppler and Heavisine. These test functions are composed of various shapes such as sharp peaks (Bumps), jump discontinuities (Blocks), oscillating behavior (Doppler) and jumps/peaks in smooth functions (Heavisine) (See Figure 3).

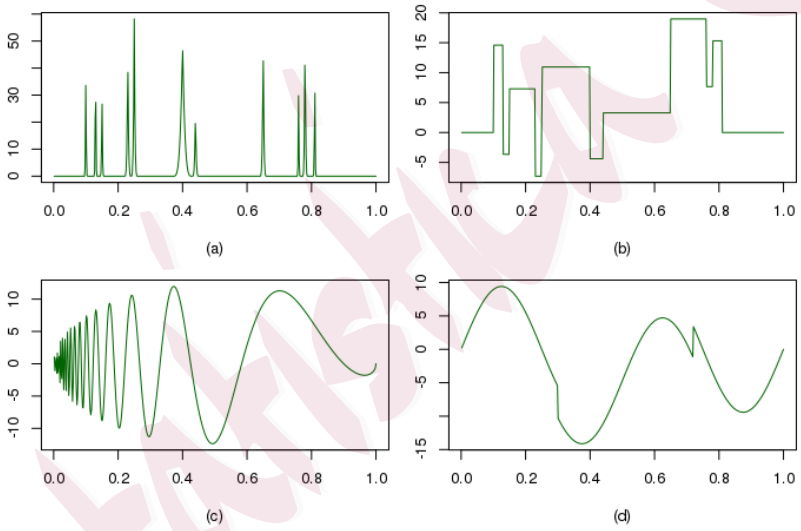

Figure 3: The Donoho and Johnstone test functions: (a) Bumps, (b) Blocks, (c) Doppler and (d) Heavisine.

The hyperparameters and types of basis functions displayed in Table S1

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

of the Supplementary Material were used in (3.11). For Bumps and Doppler, the parameter r of prior distribution for σ^2 was set to 100 to speed up convergence. We also took account of the combinations of a B-spline basis based on the shapes of test functions. We give tips on how to choose the appropriate degrees to be included in S in the Supplementary Material.

We compared our model with a variety of methods such as B-spline curve of degree 2 with 50 knots (denoted as BSP-2), Local polynomial regression with automatic smoothing parameter selection (denoted by LOESS), Smoothing spline with smoothing parameter selected by cross-validation (denoted by SS), Nadaraya–Watson kernel regression using the Gaussian kernel with bandwidth h which minimizes CV error (denoted by NWK), 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-#), Gaussian process regression with the Radial basis or Laplacian kernel (denoted by GP-R or GP-L), Bayesian curve fitting using piecewise polynomials with $l = \#1, l_0 = \#2$; Denison et al. (1998a) (denoted by BPP-#1-#2), Bayesian adaptive spline surfaces with degree #; Francom et al. (2018) (denoted by

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

BASS-#), and Lévy adaptive regression with multiple kernels; Lee et al. (2020) (denoted by LARMuK). These competitive models are implemented in R (R Core Team, 2020) with various packages: `splines2`, `fANCOVA`, `EbayesThresh` , `genlasso`, `kernlab`, `miscF`, and `BASS`.

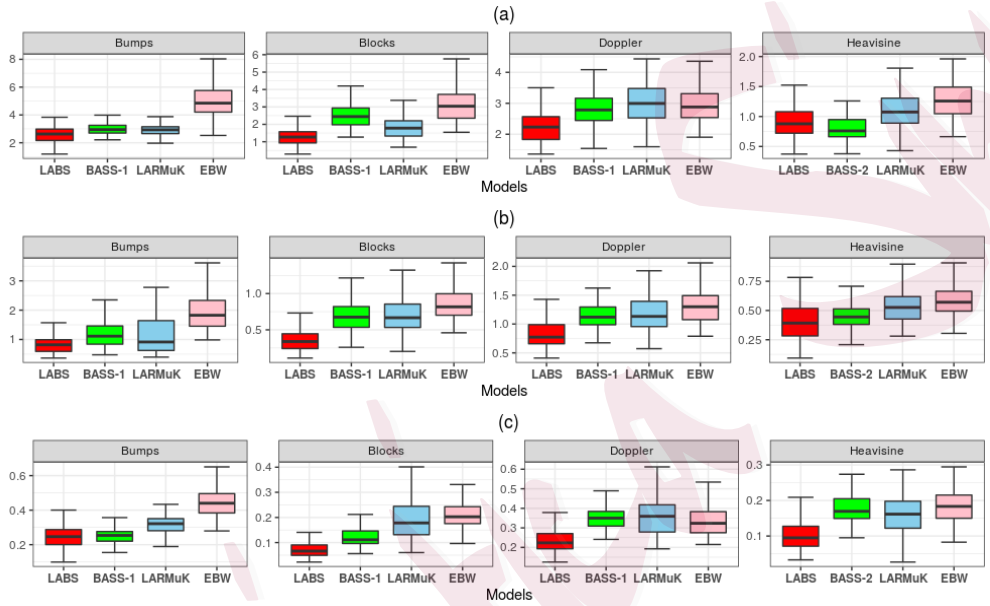

Figure 4: Boxplots of MSEs from the simulation study with $n = 128$ and RSNR = (a) 3, (b) 5 and (c) 10.

Both Figure 4 and Figure 5 show that the performance of our model is generally more accurate than other methods. 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 others regardless of noise levels with $n = 128$. It

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

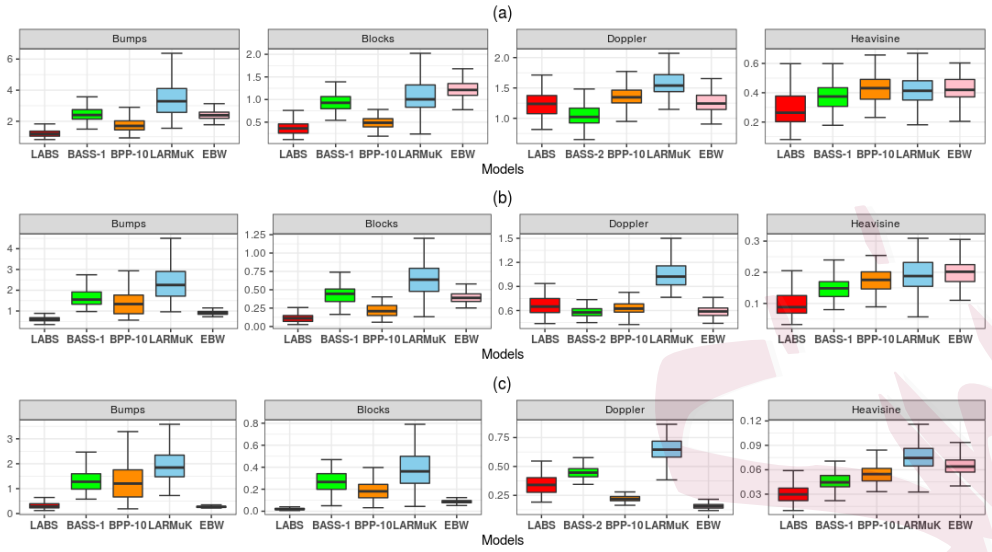

Figure 5: Boxplots of MSEs from the simulation study with $n = 512$ and RSNR = (a) 3, (b) 5 and (c) 10.

also has the smallest average mean square errors for all test functions except the Heavisine example with RSNR = 3. Similarly, for sample size $n = 512$, the LABS model comes up with the best performance 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. On the contrary, due to a small number of 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 has an excellent numerical performance. For Blocks example, LABS, in particular, yields

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

the lowest average and standard deviation of mean square errors in all scenarios. This suggests that our model has an excellent ability to find jump points. Furthermore, LABS has consistently better performance than B-spline regression using only one basis function for four simulated data sets since its overcomplete systems can be constructed by various combinations of B-spline basis functions. See the Supplementary Material.

4.2 Simulation 2 : Smooth functions with jumps and peaks

Our main interest lies in estimating smooth functions with either discontinuity such as jumps or sharp peaks or both. We design three test functions to assess the practical performance of the proposed method for our concerns. The first and second example is modified by adding some smooth parts, unlike the original version of the Bumps and Blocks of DJ test functions. Each test function provided is given by

$$\begin{aligned}\eta_1(x) =& \frac{0.6}{0.92}[4\text{ssgn}(x-0.1) - 5\text{ssgn}(x-0.13) + 5\text{ssgn}(x-0.25) \\ & - 4.2\text{ssgn}(x-0.4) + 2.1\text{ssgn}(x-0.44) + 4.3\text{ssgn}(x-0.65) \\ & - 4.2\text{ssgn}(x-0.81) + 2] + 0.2 + \sin(8\pi x),\end{aligned}$$

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

$$\begin{aligned}\eta_2(x) =&[7K_{0.005}(x-0.1)+5K_{0.07}(x-0.25)+4.2K_{0.03}(x-0.4)\\&+4.3K_{0.01}(x-0.65)+5.1K_{0.008}(x-0.78)+3.1K_{0.1}(x-0.9)]\\&+\cos(4\pi x),\end{aligned}$$

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. A formula for the last test function is

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

They are displayed in Figure 6. We call in turn them modified Blocks, modified Bumps, and modified Heavisine, respectively.

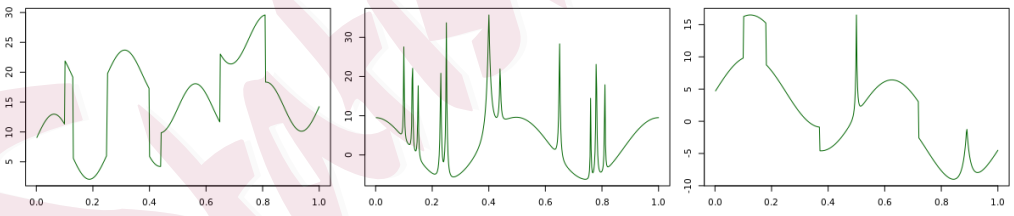

Figure 6: The three test functions used in the second simulation: modified Blocks (left), modified Bumps (center) and modified Heavisine (right).

In these experiments, we use two or more types of B-spline basis as ele-

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

ments of overcomplete systems since three functions have different shapes, unlike previous simulation studies. Hyperparameters are similar to the previous ones. All hyperparameters for the prior distributions are summarized in Table S1 of the Supplementary Material. This time again, we only compare our model with BPP, BASS, EBW, TF, and LARMuK models which have relatively good performance in some test functions of Simulation 1.

Table S12 in the Supplementary Material furnishes that the LABS model has the best outcomes when the sample size is 128, which is difficult to estimate. Furthermore, when $n = 512$, we find out from Table S13 in the Supplementary Material 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 outperforms competitors in modified Blocks, irrespective of the sample size and noise levels as expected. Among all models, the worst performing method is the BASS-2 since it cannot estimate well many jumps or peak points for given test functions. Figure 7 supports that the LABS model has the abilities to overcome the noise and adapt to smooth functions with either discontinuity such as jumps or sharp peaks or both.

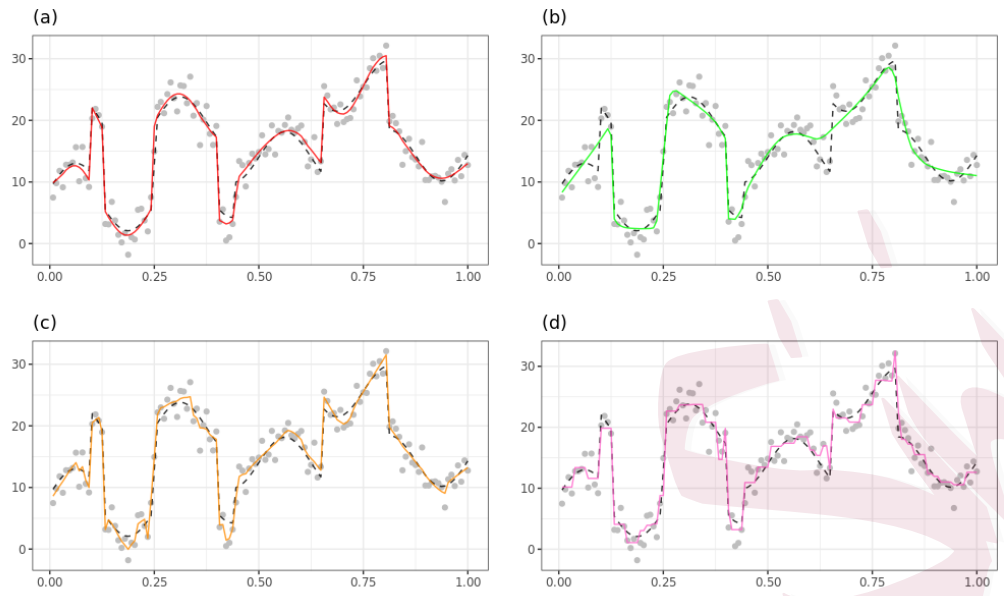

Figure 7: Comparisons of the estimates of a data set generated from the modified Blocks with $n = 128$ and RSNR $= 3$ using (a) LABS, (b) BASS-1, (c) BPP-10, and (d) EBW. Dashed lines represent true curves, solid lines represent estimates of curve.

5. Real data anlysis: Fine particulate matter in Seoul

We now apply LABS model (3.11) to an air pollution dataset, the daily maximum value of concentrations of fine particulate matter (PM2.5) in Seoul. This dataset exhibits wildly varying patterns that may have jumps or peaks. These fluctuating patterns are expected to further illustrate the features of the LABS model.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

We set the hyperparameter values of the proposed model, LABS: $a_J = 5$, $b_J = 1$, $r = 0.01$, and $R = 0.01$. In this analysis, we practically choose $S = \{0, 1, 2\}$ because the true curve of real data is unknown and it 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. Performance comparisons of our model and some rather good methods in the simulated studies are also conducted.

The fine dust has become a national issue and its forecast received great attention from the media. A lot of research on fine particulate matter (PM2.5) have been carried out as it gained social attention. According to the studies, Korea's fine dust particles originated from within the country and external sources from 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 dataset 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 daily fluctuations and seasonality. PM2.5 concentrations are higher in winter and spring than in summer and fall. We take

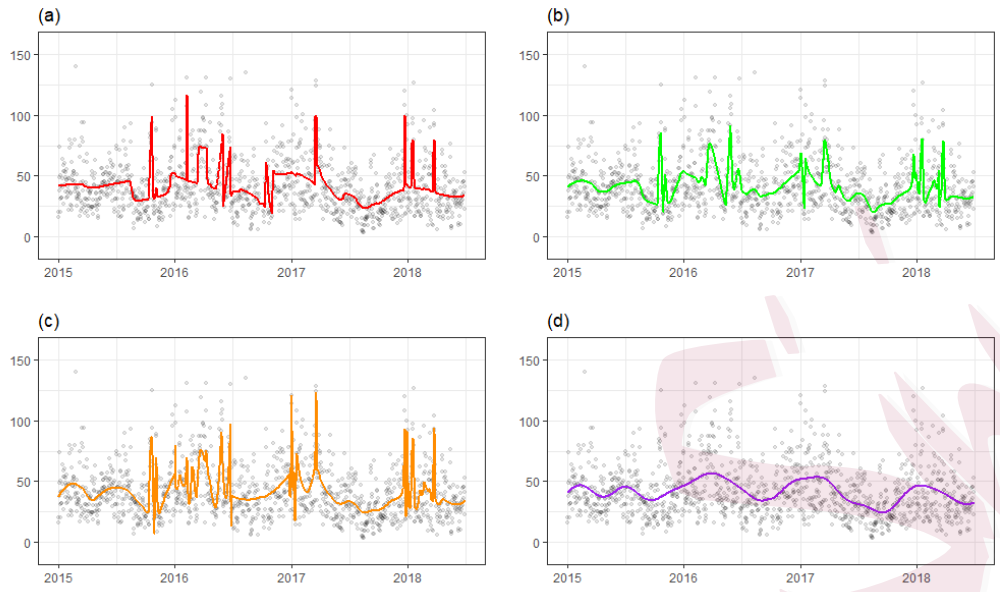

Figure 8: Posterior mean of the mean function on PM2.5 dataset using four models: (a) LABS, (b) BASS-1, (c) BPP-10 and (d) GP-R.

advantage of combinations of basis functions, $S = \{0, 1, 2\}$ to grasp such characteristics of PM2.5 data with multiple jumps and peak points. As shown in Figure 8, all four methods represent different estimated lines of the unknown mean function and pick features of the data up in their way. Interestingly, LABS, BASS-1, and BPP-10 react in different ways while they detect peaks, jumps, and smooth parts of PM2.5 data. The GP-R only 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

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

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 supports that it has a more stable performance for estimating any shape of functions due to using all three types of B-spline basis.

	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 Seoul PM2.5 dataset.

6. Conclusions

We suggested general function estimation methodologies using the B-spline basis function as the elements of an overcomplete system. The B-spline basis can systematically represent functions with varying smoothness since it has nice properties such as local support and differentiability. The overcomplete system and a Lévy random measure enable a function that has both continuous and discontinuous parts to capture all features of the unknown regression function. Simulation studies and real data analysis also present that the proposed models show better performance than other competing

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

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. Future work will develop an efficient algorithm for the LABS model and extend the LABS model for multivariate analysis.

Supplementary Materials

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

References

Abramovich, F., T. Sapatinas, and B. W. Silverman (1998). Wavelet thresholding via a bayesian approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60*(4), 725–749.

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

Clyde, M. A. and R. L. Wolpert (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 Mathemat-*

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

ics 10(2), 134–149.

Crainiceanu, C. M., D. Ruppert, R. J. Carroll, A. Joshi, and B. Goodner (2007). Spatially adaptive bayesian penalized splines with heteroscedastic errors. *Journal of Computational and Graphical Statistics 16*(2), 265–288.

De Boor, C. (1972). On calculating with b-splines. *Journal of Approximation Theory 6*(1), 50–62.

Denison, D., B. Mallick, and A. Smith (1998a). Automatic bayesian curve fitting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60*(2), 333–350.

Denison, D. G., B. K. Mallick, and A. F. Smith (1998b). Bayesian mars. *Statistics and Computing 8*(4), 337–346.

DeVore, R. A. and G. G. Lorentz (1993). *Constructive Approximation*, Volume 303. Springer Science & Business Media.

DiMatteo, I., C. R. Genovese, and R. E. Kass (2001). Bayesian curve-fitting with free-knot splines. *Biometrika 88*(4), 1055–1071.

Donoho, D. L. and I. M. Johnstone (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the american statistical association 90*(432), 1200–1224.

Donoho, D. L. and J. M. Johnstone (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika 81*(3), 425–455.

Francom, D., B. Sansó, A. Kupresanin, and G. Johannesson (2018). Sensitivity analysis and

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

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*, 1–67.

Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika 82*(4), 711–732.

Jang, P. A., A. Loeb, M. Davidow, and A. G. Wilson (2017). Scalable levy process priors for spectral kernel learning. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, Volume 30. Curran Associates, Inc.

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

Lee, J. (2007). Sampling methods of neutral to the right processes. *Journal of Computational and Graphical Statistics 16*(3), 656–671.

Lee, J. and Y. Kim (2004). A new algorithm to generate beta processes. *Computational Statistics & Data Analysis 47*(3), 441–453.

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

Lewicki, M. S. and T. J. Sejnowski (2000). Learning overcomplete representations. *Neural computation 12*(2), 337–365.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

REFERENCES

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

Luo, Z. and G. Wahba (1997). Hybrid adaptive splines. *Journal of the American Statistical Association 92*(437), 107–116.

Pillai, N. S. (2008). *Lévy random measures: Posterior consistency and applications*. PhD dissertation, Duke University.

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

Pintore, A., P. Speckman, and C. C. Holmes (2006). Spatially adaptive smoothing splines. *Biometrika 93*(1), 113–125.

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

Ruppert, D. and R. J. Carroll (2000). Theory & methods: Spatially-adaptive penalties for spline fitting. *Australian & New Zealand Journal of Statistics 42*(2), 205–223.

Simoncelli, E. P., W. T. Freeman, E. H. Adelson, and D. J. Heeger (1992). Shiftable multiscale transforms. *IEEE Transactions on Information Theory 38*(2), 587–607.

Smith, M. and R. Kohn (1996). Nonparametric regression using bayesian variable selection. *Journal of Econometrics 75*(2), 317–343.

Tibshirani, R. J. (2014). Adaptive piecewise polynomial estimation via trend filtering. *The*

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Annals of Statistics 42(1), 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*, Volume 503. John Wiley & Sons.

Wang, X., P. Du, and J. Shen (2013). Smoothing splines with varying smoothing parameter. *Biometrika 100*(4), 955–970.

Wolpert, R. L., M. A. Clyde, and C. Tu (2011). Stochastic expansions using continuous dictionaries: Lévy adaptive regression kernels. *The Annals of Statistics 39*(4), 1916–1962.

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

Zhou, S. and X. Shen (2001). Spatially adaptive regression splines and accurate knot selection schemes. *Journal of the American Statistical Association 96*(453), 247–259.

Security Algorithm Lab, SAMSUNG SDS

E-mail: swpark0413@gmail.com

Department of Statistics, Seoul National University

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

Department of Statistics, Seoul National University

E-mail: jylc@snu.ac.kr