The Annals of Statistics
2011, Vol. 39, No. 4, 1916–1962
DOI: 10.1214/11-AOS889

STOCHASTIC EXPANSIONS USING CONTINUOUS DICTIONARIES: LÉVY ADAPTIVE REGRESSION KERNELS

BY ROBERT L. WOLPERT[1], MERLISE A. CLYDE[2] AND CHONG TU

Duke University, Duke University and PIMCO

This article describes a new class of prior distributions for nonparametric function estimation. The unknown function is modeled as a limit of weighted sums of kernels or generator functions indexed by continuous parameters that control local and global features such as their translation, dilation, modulation and shape. Lévy random fields and their stochastic integrals are employed to induce prior distributions for the unknown functions or, equivalently, for the number of kernels and for the parameters governing their features. Scaling, shape, and other features of the generating functions are location-specific to allow quite different function properties in different parts of the space, as with wavelet bases and other methods employing overcomplete dictionaries. We provide conditions under which the stochastic expansions converge in specified Besov or Sobolev norms. Under a Gaussian error model, this may be viewed as a sparse regression problem, with regularization induced via the Lévy random field prior distribution. Posterior inference for the unknown functions is based on a reversible jump Markov chain Monte Carlo algorithm. We compare the Lévy Adaptive Regression Kernel (LARK) method to wavelet-based methods using some of the standard test functions, and illustrate its flexibility and adaptability in nonstationary applications.

1. Introduction. Popular approaches for nonparametric Bayesian estimation of unobserved functions generally employ as prior distributions either Gaussian processes (or random fields, in two or more dimensions) or mixtures of Dirichlet processes. In this article, we focus attention on a wider class of processes, Lévy random fields and their stochastic integrals. These include Gaussian random fields as a limiting case, while Dirichlet processes may be represented as "normalized" variants of the Gamma Lévy random field; Lévy random fields thus provide an important link between two of the random processes that form the foundation of Bayesian nonparametric methods (see Section 6). In this article, we construct prior distributions for the mean function in nonparametric regression as stochastic integrals of Lévy random fields. Under suitable regularity, these can be expressed as

Received May 2010; revised January 2011.

[1]Supported by NSF Grants DMS-07-57549, PHY-09-41373 and NASA AISR Grant NNX09AK60G.

[2]Supported by NSF Grants DMS-0342172 and DMS-04-06115.

MSC2010 subject classifications. Primary 62G08; secondary 60E07.

Key words and phrases. Bayes, Besov, kernel regression, LARK, Lévy random field, nonparametric regression, relevance vector machine, reversible jump Markov chain Monte Carlo, splines, support vector machine, wavelets.

stochastic expansions using continuous dictionaries, permitting tractable Bayesian inference. While our focus is on nonparametric regression, we hope that the reader will see the possibilities of using Lévy random fields in other contexts.

To begin, suppose we have noisy measurements $\{Y_i\}_{i\in I}$ of an unknown real-valued function $f:\mathcal{X}\to\mathbb{R}$ observed at points $\{x_i\}_{i\in I}$ in some complete separable metric space $\mathcal{X}$, with $\mathsf{E}[Y_i]=f(x_i)$. In nonparametric regression models, the mean function $f(\cdot)$ is often regarded as an element of some Hilbert space $\mathcal{H}$ of real-valued functions on $\mathcal{X}$, and is expressed as a linear combination of basis functions $\{g_j\}\subset\mathcal{H}$:

$$f(x_i)=\sum_{0\le j<J} g_j(x_i)\beta_j \tag{1}$$

with some (finite or infinite) number J of unknown coefficients $\{\beta_j\}_{0\le j<J}$. There is a vast literature on classical and Bayesian approaches for estimating f from noisy data using such methods as regression splines, Fourier expansions, wavelet expansions, and kernel methods, including kernel regression and support (or relevance) vector machines [see Chu and Marron (1991), Cristianini and Shawe-Taylor (2000), Denison et al. (2002), Vidakovic (1999), Wahba (1992), for background and references]. Many approaches, including smoothing splines and support vector machines, use as many basis elements, J, as there are data points, $n=|I|$, but employ regularization to avoid over-fitting. Sparser solutions (using fewer basis elements, $J\ll n$) may be obtained through more stringent regularization penalties, as in the Lasso [Tibshirani (1996)] and Dantzig Selector [Candès and Tao (2007)] approaches, or (often equivalently) in Bayesian methods through choice of prior distributions, as in relevance vector machines [Tipping (2001)]. Sparse solutions may also be achieved by using variable selection techniques to choose a few well-placed basis functions, perhaps in conjunction with regularization [Chen, Donoho and Saunders (1998), Denison, Mallick and Smith (1998), DiMatteo, Genovese and Kass (2001), Mallat and Zhang (1993), Johnstone and Silverman (2005b), Smith and Kohn (1996), Wolfe, Godsill and Ng (2004)].

In most signal processing and other nonstationary applications, no single (especially orthonormal) basis will lead to a sparse representation [Donoho and Elad (2003), Wolfe, Godsill and Ng (2004)]. Overcomplete dictionaries and frames [Daubechies (1992), Mallat and Zhang (1993)] provide larger collections of generating elements $\{g_\omega\}_{\omega\in\Omega}$ than would a single basis for $\mathcal{H}$, potentially allowing for more effective signal extraction and data compression. Examples of overcomplete dictionaries include unions of bases, Gabor frames, nondecimated or translational invariant wavelets, wavelet packets, or more general kernel functions or generating functions $g(x,\omega)$ where $\omega\in\Omega$ controls features (local or global) of the generating function, such as translations, dilations, modulations and shapes. Because of the redundancy inherent in overcomplete representations, coefficients for expansions using overcomplete dictionaries are not uniquely determined. This lack of uniqueness is advantageous, permitting more parsimonious representations from the dictionary than those obtained using any single basis.

In this article, we develop a fully Bayesian method for the sparse regression problem using stochastic expansions [Abramovich, Sapatinas and Silverman (2000)] of continuous dictionaries. We begin in Section 2 by introducing Lévy random fields, which are used to induce prior distributions for $f \in \mathcal{H}$ through stochastic integration of a kernel function with respect to a signed infinitely divisible random measure. We call the new model class *Lévy Adaptive Regression Kernel* or "LARK" models. The LARK framework allows both the number of kernels and kernel-specific parameters to adapt to any nonstationary features of f. Both finite and infinite expansions are considered. Exploiting the construction of Lévy random fields through Poisson random fields, we develop finite approximations to infinite expansions in Section 3 that permit tractable inference. In Section 4, we provide conditions under which the functions are almost surely in the same function space as the generating kernel. We describe the hierarchical representations of LARK models in Section 5 that enable posterior inference for the LARK model using reversible jump Markov chain Monte Carlo (RJ-MCMC) methods. In Section 6, we discuss relationships among LARK and other popular parametric and nonparametric methods. We then compare our LARK method to other procedures using simulated data in Section 7 and real data in Section 8. In Section 9, we discuss possible extensions of the LARK model.

2. Stochastic expansions and prior distributions. To make inference about the unknown mean function $f \in \mathcal{H}$ given noisy observations Y_i of $f(x_i)$ for $\{x_i\} \subset \mathcal{X}$, we must first propose a prior distribution on $\mathcal{H}$ for f. Let Ω be a complete separable metric space and $\phi : \mathcal{X} \times \Omega \to \mathbb{R}$ a Borel measurable function, and set $\phi_j(x_i) \equiv \phi(x_i, \omega_j)$ for some collection $\{\omega_j\} \subset \Omega$. As a slight extension of the basis expansion of (1), set

$$f(x) \equiv \sum_{0 \le j < J} \phi(x, \omega_j)\beta_j \tag{2}$$

for a *random* number $J \le \infty$ of randomly drawn pairs $(\beta_j, \omega_j) \in \mathbb{R} \times \Omega$. This is equivalent to specifying a random signed Borel measure $\mathcal{L}(d\omega) = \sum \beta_j \delta_{\omega_j}(d\omega)$ on Ω, giving the equivalent representation:

$$f(x) = \int_\Omega \phi(x, \omega)\mathcal{L}(d\omega). \tag{3}$$

The task of assigning prior distributions to functions $f(\cdot)$ of the form (2) is equivalent to that of specifying prior distributions for the random measure $\mathcal{L}(d\omega)$ in (3), that is, to specifying consistent joint probability distributions for all random vectors of the form $(\mathcal{L}(A_1), \dots, \mathcal{L}(A_k))$ for disjoint Borel sets $A_i \subset \Omega$. Lévy random measures, those for which $\{\mathcal{L}(A_i)\}$ are independent for disjoint $\{A_i\}$, are ideal for this purpose, since (as we will see in Section 5.3) they are simple to construct and amenable to posterior simulation. To make ideas more concrete, we first describe possible choices for the generating functions $\phi(x, \omega)$ used in our stochastic expansions and then proceed with the presentation of Lévy random measures in Section 2.2.

2.1. *Generating functions.* Possible choices for $\phi(x,\omega)$ for $\mathcal{X}=\mathbb{R}$ include translation-invariant kernel functions, such as the Gaussian

$$\phi_G(x,\omega) \equiv \exp\{-\tfrac{1}{2}\lambda(x-\chi)^2\} \tag{4a}$$

or the Laplace

$$\phi_L(x,\omega) \equiv \exp\{-\lambda|x-\chi|\} \tag{4b}$$

kernels with $\omega \equiv (\chi,\lambda) \in \mathcal{X} \times \mathbb{R}^+ \equiv \Omega$. There is no need to restrict attention to symmetric (e.g., Mercer) kernels, as required in the conventional Support Vector Machine (SVM) approach [Law and Kwok (2001), Sollich (2002)]. Asymmetric kernels, such as the one-sided exponential

$$\phi_E(x,\omega) \equiv \exp\{-\lambda(x-\chi)\}\mathbf{1}_{\{x>\chi\}} \tag{4c}$$

are useful, for example, in modeling pollutant dissipation over time. Other possibilities include piecewise-constant Haar wavelets on $\mathcal{X}=(0,1]$,

$$\phi_H(x,\omega) \equiv \mathbf{1}_{\{0<\lambda(x-\chi)\le 1\}} \tag{4d}$$

or continuous rescaling and shifting of other wavelet functions

$$\phi_\psi(x,\omega) \equiv \lambda^{1/2}\psi(\lambda(x-\chi)). \tag{4e}$$

In each of these examples, Ω is a location-scale space with location parameter χ and parameter λ determining the scale. Higher-dimensional spaces $\mathcal{X}$ may be accommodated in a similar way; for example, in Section 8.2 we use space–time kernel

$$\phi_{\mathrm{ST}}(x,\omega) \equiv \exp\{-\tfrac{1}{2}(s-\sigma)'\Lambda(s-\sigma)-\lambda|t-\tau|\} \tag{4f}$$

for space–time point $x=(s,t)\in\mathbb{R}^2\times\mathbb{R}_+$; here $\omega=(\sigma,\tau,\Lambda,\lambda)$ includes a space–time point $(\sigma,\tau)\in\mathbb{R}^2\times\mathbb{R}_+$, a positive-definite spatial dispersion matrix $\Lambda\in\mathcal{S}_2^+$, and a temporal decay rate $\lambda\in\mathbb{R}_+$.

2.2. *Lévy random measures.* For any $\nu^+\ge 0$ and any probability distribution $\pi(d\beta\, d\omega)$ on $\mathbb{R}\times\Omega$, let $J\sim\mathsf{Po}(\nu^+)$ be Poisson-distributed with mean ν^+, and let $\{(\beta_j,\omega_j)\}_{0\le j<J} \overset{\text{i.i.d.}}{\sim} \pi(d\beta\, d\omega)$; then the random measure given by

$$\mathcal{L}(A) \equiv \sum_{0\le j<J} \mathbf{1}_A(\omega_j)\beta_j \tag{5}$$

assigns independent infinitely-divisible (henceforth "ID") random variables $\mathcal{L}(A_i)$ to disjoint Borel sets $A_i\subset\Omega$, with characteristic functions

$$\mathsf{E}[e^{it\mathcal{L}(A)}] = \exp\left\{\iint_{\mathbb{R}\times A}(e^{it\beta}-1)\nu(d\beta\, d\omega)\right\} \tag{6}$$

with $\nu(d\beta\, d\omega)\equiv\nu^+\pi(d\beta\, d\omega)$. More generally, the "Lévy measure" $\nu(d\beta\, d\omega)$ need not be finite for the random measure $\mathcal{L}$ to be well defined, so long as the integral in (6) converges for all $t \in \mathbb{R}$; since the integrand is bounded on all of $\mathbb{R} \times \Omega$ and is of order $O(\beta)$ near $\beta \approx 0$, this will hold for any measure that satisfies the local L_1 integrability condition

$$\iint_{\mathbb{R}\times K} (1 \wedge |\beta|)\nu(d\beta\, d\omega) < \infty \tag{7}$$

for each compact $K \subset \Omega$. The mean and variance, when they exist, are given by $\mathsf{E}[\mathcal{L}(A)] = \iint_{\mathbb{R}\times A} \beta\nu(d\beta\, d\omega)$ and $\mathsf{Var}[\mathcal{L}(A)] = \iint_{\mathbb{R}\times A} \beta^2\nu(d\beta\, d\omega)$, respectively.

Khinchine and Lévy (1936) showed that the most general ID random variables [and hence the most general ID-valued random measures; see Rajput and Rosiński (1989), Proposition 2.1] have characteristic functions of the form

$$\begin{aligned} \mathsf{E}[e^{it\mathcal{L}(A)}] = \exp\biggl\{ & it\delta(A) - \frac{1}{2}t^2\Sigma(A) \\ & + \iint_{\mathbb{R}\times A} (e^{it\beta} - 1 - ith_0(\beta))\nu(d\beta\, d\omega)\biggr\}, \end{aligned} \tag{8}$$

where $h_0(\beta) \equiv \beta\mathbf{1}_{[-1,1]}(\beta)$, determined uniquely by the characteristic triplet of sigma-finite measures (δ, Σ, ν) consisting of a signed measure $\delta(d\omega)$ and a positive measure $\Sigma(d\omega)$ on Ω, and a positive measure $\nu(d\beta\, d\omega)$ on $\mathbb{R} \times \Omega$ that satisfies the local L_2 integrability condition

$$\iint_{\mathbb{R}\times K} (1 \wedge \beta^2)\nu(d\beta\, d\omega) < \infty \tag{9}$$

for each compact $K \subset \Omega$ and $\nu(\{0\}, \Omega) = 0$ (for more details on this nonstationary version of the classic Lévy–Khinchine formula see Jacod and Shiryaev [(1987), page 75], Cont and Tankov [(2004), pages 457–459] or Wolpert and Taqqu (2005)).

The role of the *compensator* function $h_0(\beta)$ is to make the last integrand in (8) bounded and $O(\beta^2)$ near $\beta \approx 0$, permitting the replacement of (7) with the weaker condition (9); in this case $\mathcal{L}(d\omega)$ may have countably-many points of support $\{\omega_j\} \subset \Omega$ whose magnitudes $\{\beta_j\}$ are not absolutely summable, precluding a representation of the form (5). The compensator $h_0(\beta)$ may be replaced by any bounded measurable function satisfying

$$h(\beta) = \beta + O(\beta^2), \qquad \beta \approx 0, \tag{10}$$

with a corresponding replacement of $\delta(d\omega)$ with $\delta_h(d\omega) = \delta(d\omega) + \int_{\mathbb{R}}[h(\beta) - h_0(\beta)]\nu(d\beta\, d\omega)$. Whenever (7) is satisfied, we may take $h(\beta) \equiv 0$ with the same adjustment to δ_0.

By (8) the random measure $\mathcal{L}$ may be written as the sum of two independent parts: a Gaussian portion, assigning independent normally-distributed random variables with mean $\delta_h(A_i)$ and variance $\Sigma(A_i)$ to disjoint sets A_i, and the remaining portion, with characteristic function

$$\mathsf{E}[e^{it\mathcal{L}(A)}] = \exp\biggl\{\iint_{\mathbb{R}\times A} (e^{it\beta} - 1 - ith(\beta))\nu(d\beta\, d\omega)\biggr\}. \tag{11}$$

We call a random signed measure $\mathcal{L}$ with no Gaussian component [i.e., an ID-valued measure with $\Sigma(\Omega) = \delta_h(\Omega) = 0$, that satisfies (11)] a *Lévy random measure*. Nonnegative Lévy random measures satisfying (7) were called "completely random measures" by Kingman (1967).

2.3. *Lévy random fields*. A Lévy random measure $\mathcal{L}$ satisfying (11) induces a linear mapping $\phi \mapsto \mathcal{L}[\phi]$ from functions $\phi : \Omega \to \mathbb{R}$ to random variables $\mathcal{L}[\phi] \equiv \int_\Omega \phi(\omega)\mathcal{L}(d\omega)$; such a mapping is called a *random field*. For simple functions $\phi(\omega) = \sum a_i \mathbf{1}_{A_i}(\omega)$ with each $\bar{A}_i \subset \Omega$ compact, we set $\mathcal{L}[\phi] \equiv \sum a_i \mathcal{L}(A_i)$ and verify that

$$\mathsf{E}\big[e^{it\mathcal{L}[\phi]}\big] = \exp\left\{\iint_{\mathbb{R}\times\Omega} \big(e^{it\phi(\omega)\beta} - 1 - it\phi(\omega)h(\beta)\big)\nu(d\beta\, d\omega)\right\}. \tag{12}$$

It is straightforward to extend this by continuity in probability to (at least) all bounded measurable compactly-supported $\phi : \Omega \to \mathbb{R}$. We now present a general construction based on Poisson random fields, the key to our approach to tractable posterior Bayesian inference.

2.3.1. *Poisson construction I*: *Uncompensated*. When $\nu(d\beta\, d\omega)$ satisfies (7) (i.e., $|\beta|$ is locally ν-integrable at zero) we may take $h(\beta) \equiv 0$ in (12) and construct $\mathcal{L}$ as follows. Begin with a Poisson random measure $\mathcal{N}(d\beta\, d\omega) \sim \mathsf{Po}(\nu)$ on $(\mathbb{R} \times \Omega)$ that assigns independent Poisson-distributed random variables $\mathcal{N}(C_i) \sim \mathsf{Po}(\nu(C_i))$ with means $\nu(C_i)$ to disjoint Borel sets $C_i \subset (\mathbb{R} \times \Omega)$. For any Borel set $A \subset \Omega$ with compact closure $\bar{A}$ and bounded measurable compactly-supported $\phi : \Omega \to \mathbb{R}$, set $J \equiv \mathcal{N}(\mathbb{R} \times A)$ and

$$\mathcal{L}(A) \equiv \iint_{\mathbb{R}\times A} \beta \mathcal{N}(d\beta\, d\omega) = \sum_{0 \le j < J} \mathbf{1}_A(\omega_j)\beta_j,$$

$$\mathcal{L}[\phi] \equiv \iint_{\mathbb{R}\times\Omega} \beta\phi(\omega)\mathcal{N}(d\beta\, d\omega) = \sum_{0 \le j < J} \phi(\omega_j)\beta_j, \tag{13}$$

where $\{(\beta_j, \omega_j)\}$ is the (random) set of $J \le \infty$ support points of $\mathcal{N}(d\beta\, d\omega)$. The integrals and sums in (12), (13) are well defined for all ϕ for which

$$\iint_{[-1,1]\times\Omega} |\beta\phi(\omega)|\nu(d\beta\, d\omega) < \infty,$$

which by (7) includes all bounded measurable compactly-supported functions.

For any Borel sets $A \subset \Omega$ and $B \subset \mathbb{R}$, the Poisson measure $\mathcal{N}$ assigns to the set $B \times A \subset \mathbb{R} \times \Omega$ the number $\mathcal{N}(B \times A)$ of $\mathcal{L}$'s support points $\omega_j \in A$ with mass of sizes $\beta_j \in B$. By (7) this is necessarily finite if A has compact closure and B is bounded away from zero, but if $\nu(\mathbb{R} \times \Omega) = \infty$ then $\mathcal{L}$ will have $J = \infty$ support points in Ω altogether with (almost surely) absolutely summable magnitudes $\sum_{0 \le j < J}\{|\beta_j| : \omega_j \in A\} < \infty$.

2.3.2. *Poisson construction II*: *Compensated*. The situation is more delicate in case the Lévy measure does not satisfy (7), but only the weaker bound in (9) (i.e., if β^2 is locally ν-integrable but $|\beta|$ is not). Begin again with the Poisson measure $\mathcal{N} \sim \mathsf{Po}(\nu)$ on $\mathbb{R} \times \Omega$, and introduce the *compensated* or centered Poisson measure $\tilde{\mathcal{N}}(d\beta\, d\omega) \equiv \mathcal{N}(d\beta\, d\omega) - \nu(d\beta\, d\omega)$ with mean zero [Sato (1999), page 38], inducing an isometry from $L_2(\mathbb{R} \times \Omega, \nu(d\beta\, d\omega))$ to the square-integrable zero-mean random variables. Following Wolpert and Taqqu (2005), set

$$\mathcal{L}(A) \equiv \iint_{\mathbb{R}\times A} [\beta - h(\beta)]\mathcal{N}(d\beta\, d\omega) + \iint_{\mathbb{R}\times A} h(\beta)\tilde{\mathcal{N}}(d\beta\, d\omega),$$

$$\begin{aligned}\mathcal{L}[\phi] \equiv & \iint_{\mathbb{R}\times\Omega} [\beta - h(\beta)]\phi(\omega)\mathcal{N}(d\beta\, d\omega) \\ & + \iint_{\mathbb{R}\times\Omega} h(\beta)\phi(\omega)\tilde{\mathcal{N}}(d\beta\, d\omega)\end{aligned} \tag{14}$$

for any measurable ϕ for which (14) converges. If (7) holds, one may simplify (14) to

$$\mathcal{L}[\phi] = \iint_{\mathbb{R}\times\Omega} \beta\phi(\omega)\mathcal{N}(d\beta\, d\omega) - \iint_{\mathbb{R}\times\Omega} h(\beta)\phi(\omega)\nu(d\beta\, d\omega) \tag{15a}$$

$$= \sum_{0\le j<J} \phi(\omega_j)\beta_j + \delta_h[\phi] \tag{15b}$$

showing that the role of the compensator is to add an h-dependent "drift" (or *offset*, in higher dimensions) term $\delta_h[\phi] = -\iint_{\mathbb{R}\times\Omega} h(\beta)\phi(\omega)\nu(d\beta\, d\omega)$ to (13). When (7) fails, however, both the uncompensated sum and $\delta_h[\phi]$ in (15) will be infinite, while the representation of (14) remains valid under the following conditions.

THEOREM 1. *Let ν be a Lévy measure on $\mathbb{R} \times \Omega$ satisfying* (9). *Then $\mathcal{L}[\phi]$ is well defined by* (14) *with characteristic function given by* (12) *for compensator $h_0(\beta) \equiv \beta\mathbf{1}_{\{|\beta|\le 1\}}$ if ϕ satisfies*

$$\iint_{[-1,1]^c\times\Omega} \big(1 \wedge |\beta\phi(\omega)|\big)\nu(d\beta\, d\omega) < \infty, \tag{16a}$$

$$\iint_{[-1,1]\times\Omega} \big(|\beta\phi(\omega)| \wedge |\beta\phi(\omega)|^2\big)\nu(d\beta\, d\omega) < \infty. \tag{16b}$$

If, in addition, ϕ satisfies

$$\iint_{\mathbb{R}\times\Omega} (1 \wedge \beta^2)|\phi(\omega)|\nu(d\beta\, d\omega) < \infty, \tag{16c}$$

then $\mathcal{L}[\phi]$ is well defined for any compensator $h(\beta)$ satisfying (10).

PROOF. Under these conditions, the integrands of the compensated and uncompensated Poisson integrals in (14) are in the Musielak–Orlicz spaces for which those integrals are well defined; see Rajput and Rosiński [(1989), page 9], Kwapień and Woyczyński (1992). □

In particular:

COROLLARY 1. *$\mathcal{L}[\phi]$ is well defined with characteristic function (12) for any function ϕ satisfying*

$$\iint_{\mathbb{R}\times\Omega}(1\wedge\beta^2)\big(|\phi(\omega)|\vee\phi^2(\omega)\big)\nu(d\beta\,d\omega)<\infty, \tag{17}$$

including [by (9)] all bounded measurable compactly-supported ϕ. Thus, $\mathcal{L}(A)=\mathcal{L}[\mathbf{1}_A]$ is always well defined for any Borel set $A\subset\Omega$ with compact closure $\bar{A}$.

Similarly:

PROPOSITION 1. *For a Lévy measure ν satisfying (7), take $h(\beta)\equiv 0$; then*

$$\mathcal{L}[\phi]\equiv\iint_{\mathbb{R}\times\Omega}\beta\phi(\omega)\mathcal{N}(d\beta\,d\omega)=\sum_{0\le j<J}\phi(\omega_j)\beta_j \tag{13}$$

[with $J\equiv\mathcal{N}(\mathbb{R}\times\Omega)\le\infty$] is well defined with characteristic function (12) for any ϕ satisfying

$$\iint_{\mathbb{R}\times\Omega}\big(1\wedge|\beta\phi(\omega)|\big)\nu(d\beta\,d\omega)<\infty. \tag{18}$$

2.4. *Constructing Lévy kernel integrals.* Denote by Φ the linear space of functions $\phi:\Omega\to\mathbb{R}$ for which $\mathcal{L}[\phi]$ has been defined; we have seen that this includes at least all bounded measurable compactly-supported functions ϕ. Denote by $\mathcal{G}$ the linear space of measurable functions $g:\mathcal{X}\to\Phi$, and simplify notation by writing "$g(x,\omega)$" for $g(x)(\omega)$. Each of the generating functions introduced in (4) lies in $\mathcal{G}$. For any $g\in\mathcal{G}$, we can construct a random function $f:\mathcal{X}\to\mathbb{R}$ by

$$\begin{aligned} f(x)&\equiv\mathcal{L}[g(x)] \\ &=\iint_{\mathbb{R}\times\Omega}g(x,\omega)[\beta-h(\beta)]\mathcal{N}(d\beta\,d\omega) \\ &\quad+\iint_{\mathbb{R}\times\Omega}g(x,\omega)h(\beta)\tilde{\mathcal{N}}(d\beta\,d\omega) \\ &=\sum_{0\le j<J}g(x,\omega_j)[\beta_j-h(\beta_j)] \\ &\quad+\iint_{\mathbb{R}\times\Omega}g(x,\omega)h(\beta)\tilde{\mathcal{N}}(d\beta\,d\omega)\quad\text{or} \end{aligned} \tag{19}$$

$$=\sum_{0\le j<J}g(x,\omega_j)\beta_j\qquad\text{if (7) holds so compensation is unneeded.} \tag{20}$$

Integer moments of $f(x)$ are easy to compute, when they exist, from the characteristic function given in (12), for example:

$$\mathsf{E}\{f(x)\} = \iint_{\mathbb{R}\times\Omega} \phi(x,\omega)[\beta - h(\beta)]\nu(d\beta\, d\omega), \tag{21a}$$

$$\mathsf{Cov}\{f(x_1), f(x_2)\} = \iint_{\mathbb{R}\times\Omega} \phi(x_1,\omega)\phi(x_2,\omega)\beta^2\nu(d\beta\, d\omega). \tag{21b}$$

2.5. *Examples of Lévy measures*. We now consider some specific examples of Lévy random fields and the corresponding kernel integrals. Familiar examples include Poisson, Gamma, Cauchy and more generally α-Stable random fields.

2.5.1. *Compound Poisson processes*. The simplest model to consider would be that of (2), with finite Lévy measure satisfying $\nu^+ \equiv \nu(\mathbb{R}\times\Omega) < \infty$, reproduced here:

$$f(x) \equiv \sum_{0\le j<J} \phi(x,\omega_j)\beta_j. \tag{2}$$

This has a Poisson-distributed number $J \sim \mathsf{Po}(\nu^+)$ of terms whose locations ω_j and magnitudes β_j are i.i.d. with an arbitrary distribution $\{\beta_j,\omega_j\} \overset{\text{i.i.d.}}{\sim} \pi(d\beta\, d\omega)$, hence Lévy measure of the form $\nu(d\beta\, d\omega) = \nu^+\pi(d\beta\, d\omega)$. The marginal distribution of $f(x)$ at each $x\in\mathcal{X}$ is compound Poisson.

2.5.2. *Gamma random fields*. The Lévy measure for the Gamma random field is infinite but satisfies the strong local L_1 integrability condition (7), obviating compensation; in the homogeneous case, it is

$$\nu(d\beta\, d\omega) = \beta^{-1}e^{-\beta\eta}\mathbf{1}_{\{\beta>0\}}\, d\beta\gamma(d\omega) \tag{22}$$

for some σ-finite measure $\gamma(d\omega)$ on Ω, giving $\mathcal{L}(A) \sim \mathsf{Ga}(\gamma(A),\eta)$ [with mean $\gamma(A)/\eta$] for Borel measurable $A\subset\Omega$ with $\gamma(A)<\infty$. Because ν is concentrated on $\mathbb{R}_+$, the mass β_j at each of the Gamma random measure's support points ω_j is positive, so all the coefficients in the expression $f(x) = \sum\phi(x,\omega_j)\beta_j$ are nonnegative. With a nonnegative generating function $\phi\in\mathcal{G}$, this provides a direct way to construct nonnegative mean functions $f\ge 0$ without having to transform the responses $\{Y_i\}$ as Gaussian methods would require. The mean $\mathsf{E}[f(x)] = \eta^{-1}\int g(x,\omega)\gamma(d\omega)$ is available from (21a), as is the covariance from (21b).

2.5.3. *Symmetric Gamma random fields*. A symmetric analogue of the Gamma random field (22) has Lévy measure

$$\nu(d\beta\, d\omega) = |\beta|^{-1}e^{-|\beta|\eta}\, d\beta\gamma(d\omega) \tag{23}$$

on all of $\mathbb{R}\times\Omega$, leading to random variables $\mathcal{L}(A)$ distributed as the difference of two independent $\mathsf{Ga}(\gamma(A),\eta)$ variables, with characteristic function $\mathsf{E}[e^{it\mathcal{L}(A)}] =$

$(1+t^2/\eta^2)^{-\gamma(A)}$. Both the standard positive Gamma random measure and this symmetric version satisfy the local L_1 bound (7), hence no compensation is required so we may take $h(\beta)\equiv 0$ and employ the simple construction (20) of $f(x)$. The mean $\mathsf{E}[f(x)]=0$ vanishes for the symmetric Gamma random field, or for any other Lévy random field with a symmetric (in $\pm\beta$) Lévy measure satisfying (7). Covariances are available from (21b). Nearly all of the commonly used isotropic geostatistical covariance functions [see Chilès and Delfiner (1999), Section 2.5] may be achieved by the choice of a suitable generating kernel $g(x,\cdot)$ and Lévy measure $\nu(d\beta\, d\omega)$; see Clyde and Wolpert (2007) for specific examples.

2.5.4. *Symmetric α-Stable random fields.* Symmetric α-Stable (SαS) Lévy random fields have Lévy measure

$$\nu(d\beta\, d\omega)=c_\alpha\alpha|\beta|^{-1-\alpha}\, d\beta\,\gamma(d\omega) \tag{24}$$

on $\mathbb{R}\times\Omega$ for some $0<\alpha<2$ and σ-finite positive measure $\gamma(d\omega)$, where $c_\alpha=(1/\pi)\Gamma(\alpha)\sin(\pi\alpha/2)$, giving $\mathcal{L}(A)\sim\mathsf{St}(\alpha,0,\gamma(A),0)$ [in parametrization (M) of Zolotarev (1986), page 11] with infinite variance (and thus no meaningful covariance function for $f(x)\equiv\mathcal{L}[g(x)]$). This infinite Lévy measure satisfies (9) for all $0<\alpha<2$, but satisfies the stronger local L_1 condition (7) only for $0<\alpha<1$; thus compensation is required to construct SαS random fields with $1\le\alpha<2$, including the Cauchy case of $\alpha=1$. One can show that $f(x)$ is well defined for any $\phi(x,\cdot)\in L_\alpha(\Omega,\gamma(d\omega))$, including the generating functions of (4). The SαS fields have heavier tails than, for example, the symmetric Gamma fields of Section 2.5.3, and may be more appropriate for problems where one might expect $f(\cdot)$ to include by a few heavily weighted kernels.

3. Approximations for implementing kernel integrals. Computer simulations of Lévy random measures $A\mapsto\mathcal{L}(A)$ and random fields $\phi\mapsto\mathcal{L}[\phi]$ associated with *finite* Lévy measures ν may be constructed as in (5), (13), simply by setting $\nu^+\equiv\nu(\mathbb{R}\times\Omega)$ and drawing $J\sim\mathsf{Po}(\nu^+)$ and $\{(\beta_j,\omega_j)\}_{0\le j<J}\overset{\text{i.i.d.}}{\sim}\pi(d\beta\, d\omega)\equiv\nu(d\beta\, d\omega)/\nu^+$. If $\nu(\mathbb{R}\times\Omega)=\infty$ however the sums in these equations will include countably infinitely-many terms, and may not be absolutely summable. We now construct an approximating set of finite Lévy measures $\{\nu_\varepsilon\}$ indexed by $\varepsilon>0$ and show that the approximate Lévy random fields $\mathcal{L}_\varepsilon[\phi]$ converge to the random field $\mathcal{L}[\phi]$ given in (14). Note that ε is *not* a model parameter. It is only a device used for two purposes: as a tool in the theorems constructing LARK models (in this section) and establishing their properties (in Section 4), and to enable the construction of practical numerical methods to approximate LARK models within specified error bounds (in Section 5).

THEOREM 2. *Let ν be a Lévy measure defined on $\mathbb{R}\times\Omega$ satisfying* (9) *and $\phi\in\Phi$ satisfying* (16). *Take $\{K_\varepsilon\}$ to be any family of compact sets increasing to Ω*

as $\varepsilon \to 0$, *and for any Borel sets* $A \subset \Omega$ *and* $B \subset \mathbb{R}$ *and let* ν_ε *be the unique Borel measure on* $\mathbb{R} \times \Omega$ *satisfying*

$$\nu_\varepsilon(B \times A) \equiv \nu\big((B \cap [-\varepsilon, \varepsilon]^c) \times (A \cap K_\varepsilon)\big) \tag{25}$$

for $B \subset \mathbb{R}$, $A \subset \Omega$ [*note* $\nu_\varepsilon^+ \equiv \nu_\varepsilon(\mathbb{R} \times \Omega) < \infty$]. *Let* $h(\cdot)$ *be any bounded measurable compensator function on* $\mathbb{R}$ *satisfying* $h(\beta) = \beta + O(\beta^2)$ *for* β *near zero. Then as* $\varepsilon \to 0$, *the random variables*

$$\begin{aligned}\mathcal{L}_\varepsilon[\phi] \equiv & \iint_{[-\varepsilon,\varepsilon]^c \times K_\varepsilon} \beta\phi(\omega)\mathcal{N}(d\beta\, d\omega) \\ & - \iint_{[-\varepsilon,\varepsilon]^c \times K_\varepsilon} h(\beta)\phi(\omega)\nu(d\beta\, d\omega)\end{aligned} \tag{26}$$

converge in probability to $\mathcal{L}[\phi]$ *of* (14).

PROOF. The error in approximating $\mathcal{L}[\phi]$ of (14) by $\mathcal{L}_\varepsilon[\phi]$ of (26) is

$$\begin{aligned}\mathcal{L}[\phi] - \mathcal{L}_\varepsilon[\phi] = & \iint_{N_\varepsilon} \big(\beta - h(\beta)\big)\phi(\omega)\mathcal{N}(d\beta\, d\omega) \\ & + \iint_{N_\varepsilon} h(\beta)\phi(\omega)\tilde{\mathcal{N}}(d\beta\, d\omega),\end{aligned} \tag{27}$$

where $N_\varepsilon \equiv \{(\beta, \omega) : |\beta| \le \varepsilon \text{ or } \omega \in K_\varepsilon^c\}$. The first term in (27) converges to zero almost surely, and the second in L_1, as $\varepsilon \to 0$; see the Appendix for details. $\square$

The approximation $\mathcal{L}_\varepsilon[\phi]$ is the sum of a Lévy random field with finite Lévy measure ν_ε [hence with simple representation (13)] and a deterministic drift term $\delta_\varepsilon[\phi]$ given by the second integral in (26). The drift vanishes whenever $\nu(d\beta\, d\omega)$ is symmetric in $\pm\beta$ and $h(\beta)$ is odd.

COROLLARY 2. *If either* (a) $\nu(d\beta\, d\omega)$ *satisfies* (7), *or* (b) $\nu(d\beta\, d\omega)$ *satisfies* (9) *and is even in* $\pm\beta$, *and also* $h(\beta)$ *is an odd function, then for each* $x \in \mathcal{X}$,

$$f_\varepsilon(x) \equiv \sum_{0 \le j < J_\varepsilon} g(x, \omega_j)\beta_j \tag{28}$$

with

$$J_\varepsilon \sim \mathsf{Po}(\nu_\varepsilon^+), \qquad \{\beta_j, \omega_j\}_{0 \le j < J_\varepsilon} \mid J_\varepsilon \overset{i.i.d.}{\sim} \nu_\varepsilon(d\beta\, d\omega)/\nu_\varepsilon^+$$

converges to $f(x)$ *in probability as* $\varepsilon \to 0$.

PROOF. With $f_\varepsilon(x) \equiv \mathcal{L}_\varepsilon[g(x)]$,

$$\begin{aligned}f_\varepsilon(x) &= \int_\Omega g(x, \omega)\mathcal{L}_\varepsilon(d\omega) \\ &= \iint_{\mathbb{R}\times\Omega} g(x, \omega)\beta\mathcal{N}_\varepsilon(d\beta\, d\omega) - \iint_{\mathbb{R}\times\Omega} g(x, \omega)h(\beta)\nu_\varepsilon(d\beta\, d\omega)\end{aligned} \tag{29}$$

with $\mathcal{N}_\varepsilon(d\beta\, d\omega) \sim \mathsf{Po}(\nu_\varepsilon(d\beta\, d\omega))$. If ν satisfies (7), then without loss of generality take the compensator function $h(\beta) \equiv 0$. In both cases (a) and (b), the second integral in (29) vanishes, leading to (28) [cf. (2)]. $\square$

Note that in case (b) the $\{g(x,\omega_j)\beta_j\}$ are not absolutely summable so "$\sum_{j=0}^{\infty} g(x,\omega_j)\beta_j$" does not converge in the Lebesgue sense. In each of our applications the conditions of Corollary 2 hold, allowing us to approximate ν by a finite Lévy measure ν_ε [and $\mathcal{L}$ by $\mathcal{L}_\varepsilon \sim \mathsf{Lévy}(\nu_\varepsilon)$], and exploit the resulting Poisson representation for inference.

4. Function spaces for LARK models. Theorem 2 and Corollary 2 establish pointwise convergence of $f_\varepsilon(x)$ to $f(x)$ as $\varepsilon \to 0$; in this section we provide conditions to ensure that $f_\varepsilon(\cdot) \to f(\cdot)$ in appropriate Besov or Sobolev norms if the generating functions lie in the same space.

For $s \geq 0$ and $d \in \mathbb{N}$ denote by $\mathbb{W}_2^s(\mathbb{R}^d)$ the Sobolev space of real-valued square-integrable functions $f(\cdot) \in L_2(\mathbb{R}^d)$ [Sobolev (1991), Section 1.7, Reed and Simon (1975), page 50] with finite Sobolev norm

$$\|f\|_{\mathbb{W}_2^s} = \left\{ \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} (1+|\xi|^2)^s |\hat{f}(\xi)|^2\, d\xi \right\}^{1/2} \tag{30}$$

with Fourier transforms defined for $f \in L_1(\mathbb{R}^d)$ by

$$\hat{f}(\xi) = \int_{\mathbb{R}^d} e^{i\xi\cdot x} f(x)\, dx$$

and by L_2 limits for $f \in L_2(\mathbb{R}^d)$; here $d\xi$ and dx denote the Lebesgue volume element in $\mathbb{R}^d$, and $\xi \cdot x$ denotes the Euclidean inner product. Each $\mathbb{W}_2^s$ is a Banach space, hence complete. By Plancherel's theorem, each $f \in \mathbb{W}_2^s$ with $s \geq 0$ has s distributional derivatives in $L_2(\mathbb{R})$, and by Sobolev's lemma has k continuous derivatives for each integer $0 \leq k < s - d/2$.

Besov spaces constitute a flexible family that includes elements with wide spatial irregularity. The Besov space $\mathbb{B}_{pq}^s$ consists of those $f \in L_p(\mathbb{R}^d)$ whose Besov semi-norms are finite. Several equivalent Besov semi-norms appear in the literature [Triebel (1992), Theorem 2.6.1, page 140]; we use the definition given as equation 2 of that theorem. For $p, q \geq 0$ and $s > d(1/p - 1)_+$ and for any integer $m > s$ ($m = 1 + \lfloor s \rfloor$ is easiest), set

$$|f|_{pq}^s = \left(\int_{|h|\leq 1} |h|^{-sq} \|\Delta_h^m f\|_p^q\, dh/|h|^d \right)^{1/q}$$

or, in dimension $d = 1$,

$$|f|_{pq}^s = \left(2\int_0^1 h^{-1-sq} \|\Delta_h^m f\|_p^q\, dh \right)^{1/q}, \tag{31}$$

where Δ_h^m denotes the mth forward finite difference,

$$\begin{aligned} \Delta_h^0 f(x) &= f(x), \\ \Delta_h^m f(x) &= [\Delta_h^{m-1} f(x+h) - \Delta_h^{m-1} f(x)] \\ &= \sum_{k=0}^{m} \binom{m}{k} (-1)^{m-k} f(x+kh). \end{aligned} \tag{32}$$

The Besov space $\mathbb{B}_{pq}^s$ is the Banach space completion of $L_p(\mathbb{R}^d)$ under norm

$$\|f\|_{pq}^s = \|f\|_p + |f|_{pq}^s. \tag{33}$$

For $p = q = 2$, $\mathbb{B}_{pq}^s$ coincides with the Sobolev space $\mathbb{W}_2^s$.

For fixed $\omega \in \Omega$, each of the kernel functions $g(\cdot, \omega)$ in (4) is in $\mathbb{B}_{pq}^s$ for all $p, q \geq 1$ and some $s > 0$, and hence each finite approximation of the form (28) lies in the same $\mathbb{B}_{pq}^s$. For example, the Gaussian kernel of (4a) (along with its d-dimensional generalization) satisfies $g_G(\cdot, \omega) \in \mathbb{B}_{pq}^s$ for every $s < \infty$ and $p, q \geq 1$, while in $\mathbb{R}^1$ the double-sided Laplace kernel of (4b) satisfies $g_L(\cdot, \omega) \in \mathbb{B}_{pp}^s$ for $s < 1 + 1/p < 2$ for integer p and the Haar wavelet of (4d) is in $\mathbb{B}_{pq}^s$ only for $s < 1/p$. To simplify proofs in Section 4.1, we will restrict attention to generating functions g on $\mathbb{R}^d$; these results may be extended to bounded domains the Besov semi-norms defined in terms of differences on bounded domains in Section 5.2.2 of Triebel (1992) may be used to extend these results.

We now provide conditions for LARK models to be in the same Besov space as their generating functions.

4.1. *Convergence of LARK models in Besov spaces.*

THEOREM 3. *Fix $g \in \mathbb{B}_{pq}^s(\mathbb{R}^d)$ for some $p, q \geq 1$ and $s > 0$ and a Lévy measure ν on $\mathbb{R} \times \Omega$ with $\Omega = (\mathcal{S}_+^d \times \mathbb{R}^d)$ of translation-invariant product form $\nu(d\beta\, d\omega) = \tilde{\nu}(d\beta\, d\Lambda)\, d\chi$ [here $\omega = (\Lambda, \chi)$] for a σ-finite measure $\tilde{\nu}(d\beta\, d\Lambda)$ on $\mathbb{R} \times \mathcal{S}_+^d$ that satisfies the integrability condition (7). Define a location-scale LARK model $f(\cdot)$ on $\mathcal{X} = \mathbb{R}^d$ by: $f(x) = \int_\Omega \phi(x, \omega)\mathcal{L}(d\omega)$ where $\phi(x, \omega) \equiv g(\Lambda(x - \chi))$ satisfies (18) for each fixed $x \in \mathcal{X}$. Then f has the almost surely convergent series expression*

$$f(x) = \sum_j g\big(\Lambda_j(x - \chi_j)\big)\beta_j \tag{34}$$

and $f \in \mathbb{B}_{pq}^s$ almost surely if $\tilde{\nu}$ satisfies

$$\iint_{\mathbb{R}\times\mathcal{S}_+^d} (1 \wedge |\beta||\Lambda|^{-1/p})\tilde{\nu}(d\beta\, d\Lambda) < \infty, \tag{35a}$$

$$\iint_{\mathbb{R}\times\mathcal{S}_+^d} (1 \wedge |\beta||\Lambda|^{s-1/p})\tilde{\nu}(d\beta\, d\Lambda) < \infty. \tag{35b}$$

PROOF. Equation (18) ensures that the sum in (34) will converge almost surely for each fixed $x \in \mathcal{X}$, with a finite number of terms $|g(\Lambda_j(x-\chi_j))\beta_j| > 1$ and infinitely many, but absolutely summable, terms with $|g(\Lambda_j(x-\chi_j))\beta_j| \le 1$. The L_p norm of f satisfies the bound

$$\|f\|_p \le \sum_j \big\|g\big(\Lambda_j(\cdot - \chi_j)\big)\big\|_p |\beta_j| = \|g\|_p \sum_j |\Lambda_j|^{-1/p} |\beta_j|$$

by the triangle inequality and Proposition 2 in Appendix A. This is finite almost surely by (35a) since $g \in \mathbb{B}^s_{pq} \subset L_p$. The Besov semi-norm of f is bounded by

$$\begin{aligned}
|f|^s_{pq} &\le \sum_j |\beta_j| \big|g\big(\Lambda_j(x-\chi_j)\big)\big|^s_{pq} \\
&= \sum_j |\beta_j| \bigg(\int_{|h|\le 1} |h|^{-d-sq} \big\|\Delta^m_h g\big(\Lambda_j(\cdot-\chi_j)\big)\big\|^q_p \, dh \bigg)^{1/q} \\
&= \sum_j |\beta_j| \bigg(\int_{|h|\le 1|} |h|^{-d-sq} |\Lambda_j|^{-q/p} \|\Delta^m_{\Lambda_j h} g\|^q_p \, dh \bigg)^{1/q}
\end{aligned}$$

by Proposition 2; changing variables $h \mapsto t = \Lambda h$, this is

$$= \sum_j |\beta_j| |\Lambda_j|^{s-1/p} \bigg(\int_{|\Lambda_j^{-1} t| \le 1} |t|^{-d-sq} \|\Delta^m_t g\|^q_p \, dt \bigg)^{1/q}. \tag{36}$$

The integral in (36) is bounded by

$$\begin{aligned}
\int_{\mathbb{R}^d} |t|^{-d-sq} \|\Delta^m_t g\|^q_p \, dt &= \int_{|t|\le 1} |t|^{-d-sq} \|\Delta^m_t g\|^q_p \, dt \\
&\quad + \int_{|t|>1} |t|^{-d-sq} \|\Delta^m_t g\|^q_p \, dt.
\end{aligned}$$

The first term is just $(|g|^s_{pq})^q$, and (32) implies $\|\Delta^m_t g\|_p \le 2^m \|g\|_p$, so

$$\begin{aligned}
&\le (|g|^s_{pq})^q + \int_{|t|>1} |t|^{-d-sq} (2^m \|g\|_p)^q \, dt \\
&= (|g|^s_{pq})^q + \frac{\pi^{d/2} 2^{1+mq}}{\Gamma(d/2) sq} \|g\|^q_p \\
&\le (c\|g\|^s_{pq})^q
\end{aligned}$$

for some $c < \infty$, so

$$|f|^s_{pq} \le c \|g\|^s_{pq} \sum_j |\beta_j| |\Lambda_j|^{s-1/p}, \tag{37}$$

which is almost surely finite by (35b). $\square$

Each of the kernels $g(\cdot, \omega)$ considered in the examples in Sections 7 and 8 may be shown to be in some Besov space $\mathbb{B}^s_{pq}$, and each is bounded by $\|g\|_\infty \le 1$. Corollary 3 establishes that each of our LARK models with a Lévy measure that satisfies (7) is in the same space $\mathbb{B}^s_{pq}$ as its generating function.

COROLLARY 3. *Let $f(x) = \int \phi(x, \omega)\mathcal{L}(d\omega)$ be a one-dimensional LARK model on a compact set $\mathcal{X} \subset \mathbb{R}^1$, with product Lévy measure $\nu(d\beta\, d\omega) = \nu_\beta(d\beta)\pi_\lambda(d\lambda)\, d\chi$ on $\mathbb{R} \times \mathbb{R}^+ \times \mathcal{X}$ satisfying (7) with Gamma probability measure $\pi_\lambda(d\lambda) = \mathsf{Ga}(a_\lambda, b_\lambda)$ and location-scale generator $\phi(x, \omega) = g(\lambda(x - \chi))$ with bounded $g \in \mathbb{B}^s_{pq}$. Then $f \in \mathbb{B}^s_{pq}$ almost surely if $\alpha_\lambda > 1/p$ for $p, q \ge 1$ and $s > 0$. In particular, if $a_\lambda \ge 1$ then $f \in \mathbb{B}^s_{pq}$ if $g \in \mathbb{B}^s_{pq}$ for all $p, q \ge 1$ and $s > 0$.*

PROOF. Equation (18) holds for bounded $g \in \mathbb{B}^s_{pq}$ with Lévy measures of the form indicated; the conditions on α_λ ensure that also $\int_{\mathbb{R}_+} \lambda^{-1/p}\pi_\lambda(d\lambda) < \infty$ and $\int_{\mathbb{R}_+} \lambda^{s-1/p}\pi_\lambda(d\lambda) < \infty$, so the bounds of (35) hold. $\square$

4.2. *Comparisons with Abramovich, Sapatinas and Silverman*. The stochastic wavelet expansion of Abramovich, Sapatinas and Silverman (2000) may be viewed as a LARK model using wavelet generator (4e), with coefficients that, when conditioned on the scale parameters $\{a_j\}$, have independent Gaussian distributions $\{\beta_j\} \overset{\text{ind}}{\sim} \mathsf{No}(0, c a_j^{-\delta})$ with $\omega = (a, b) \in [a_0, \infty) \times [0, 1)$ and $\nu_\omega(d\omega) \propto a^{-\xi}\mathbf{1}_{\{a \ge a_0\}}\, db\, da$ for some $c, \delta, \xi \ge 0$, $\delta + \xi > 0$ and $a_0 \ge 1$. The parameters δ and ξ control the size and frequency of wavelet coefficients and determine whether the expansion will have a well-defined limit. For a finite Lévy measure $\nu_\omega(d\omega)$ ($\xi > 1$), the expansion will be in the corresponding Besov space of the generating wavelet with probability one. For $\xi \le 1$, the Poisson mean is no longer finite; however, Abramovich, Sapatinas and Silverman (2000) provide conditions on δ and ξ so that f falls in the corresponding Besov space of the generating wavelet.

For "simplicity of exposition," Abramovich, Sapatinas and Silverman work with functions of unit period [i.e., satisfying $g(x) = g(x + 1)$] and regard them as functions on the unit torus $\mathbb{T}$, the interval [0, 1] with the endpoints identified. We now illustrate how the LARK theory may be used to prove that the resulting expansion lies in $\mathbb{B}^s_{pq}(\mathbb{T})$ if the generating function does. The Besov sequence norms used by Abramovich, Sapatinas and Silverman and others are natural for the Gaussian distributions and discrete wavelet expansions they study; we have found the (equivalent) function norms to be more convenient for continuous wavelet expansions using non-Gaussian (α-Stable, e.g.) distributions used for the coefficients in our expansions. We follow Nikol'skiĭ [(1975), Sections 1.1.1 and 4.3.5] in defining Besov norms on the torus by replacing the L_p norm on $\mathbb{R}$ with that over $\mathbb{T}$ in the definition of the Besov semi-norm and norm [see (31), (33)], and in denoting the corresponding spaces by $L^*_p(\mathbb{T})$ and $\mathbb{B}^{s*}_{pq}(\mathbb{T})$, respectively.

To simplify the proof, we will use the following lemma.

LEMMA 1. *Let $\pi_z(dz)$ denote the standard normal distribution on $\mathbb{R}$, let $g \in L_p^*(\mathbb{T})$ with $p \geq 1$ and let $r \in \{0, 1\}$. Then*

$$\iiint_{\mathbb{R}\times[1,\infty)\times\mathbb{T}} (1 \wedge |zg(u)^r|\lambda^{-a})\lambda^{-b}\pi_z(dz)\,d\lambda\,du < \infty$$

for any $a \in \mathbb{R}$ if $b > 1$, and for all $a > 1 - b$ if $b \leq 1$.

The proof is given in Appendix A.1.

THEOREM 4. *Let $g \in \mathbb{B}_{pq}^{s*}(\mathbb{T})$ for some $p, q \geq 1$ and $s > 0$. Let $\mathcal{L}(d\omega)$ be a random field on $\Omega = [1, \infty) \times \mathbb{T}$ with Lévy measure*

$$\nu(d\beta\,d\lambda\,d\chi) = \frac{1}{\sqrt{2\pi}}\lambda^{\delta/2-\zeta}e^{-\beta^2\lambda^\delta/2}\,d\beta\,d\lambda\,d\chi \tag{38}$$

on $\mathbb{R} \times \Omega$ with $\delta, \zeta \geq 0$. Then the LARK model $f(x) = \int_\Omega \lambda^{1/2}g(\lambda(x - \chi))\mathcal{L}(d\omega)$ has an absolutely convergent expansion

$$f(x) = \sum_j \beta_j\lambda_j^{1/2}g\big(\lambda_j(x - \chi_j)\big), \qquad 0 \leq x < 1, \tag{39}$$

provided that $\frac{\delta-1}{2} > 1 - \zeta$ for $0 \leq \zeta \leq 1$, or for any $\delta \geq 0$ if $\zeta > 1$. Also $f(\cdot) \in \mathbb{B}_{pq}^{s*}(\mathbb{T})$ almost surely for $\frac{\delta-1}{2} > s + 1 - \zeta$ if $0 \leq \zeta \leq 1$ or for any $\delta \geq 0$ if $\zeta > 1$.

PROOF. The absolute convergence of (39) for each x will follow from Proposition 1 if we can verify the conditions of (18), that is, finiteness of the integral

$$\iiint_{\mathbb{R}\times[1,\infty)\times\mathbb{T}} \big(1 \wedge \big|\beta\lambda^{1/2}g\big(\lambda(x - \chi)\big)\big|\big)\nu(d\beta\,d\lambda\,d\chi). \tag{40}$$

Applying the change of variables $\beta \mapsto z = \lambda^{\delta/2}\beta$,

$$= \iiint_{\mathbb{R}\times[1,\infty)\times\mathbb{T}} \big(1 \wedge |z|\lambda^{(1-\delta)/2}\big|g\big(\lambda(x - \chi)\big)\big|\big)\lambda^{-\zeta}\pi_z(dz)\,d\lambda\,d\chi, \tag{41}$$

where $\pi_z(dz)$ is the standard normal distribution. Since the term in parentheses is bounded by one, (41) is finite for all δ and g if $\zeta > 1$. For $0 \leq \zeta \leq 1$, apply another change of variables $\chi \mapsto u = \lambda(x - \chi)$ and apply periodicity

$$= \int_{\mathbb{R}}\int_1^\infty\int_{\lambda(x-1)}^{\lambda x} \big(1 \wedge |zg(u)|\lambda^{(1-\delta)/2}\big)\,du\lambda^{-1-\zeta}\,d\lambda\pi_z(dz)$$

which, due to periodicity, satisfies the bound

$$\leq \int_{\mathbb{R}}\int_1^\infty\int_0^1 \big(1 \wedge |zg(u)|\lambda^{(1-\delta)/2}\big)\,du\lceil\lambda\rceil\lambda^{-1-\zeta}\,d\lambda\pi_z(dz)$$

$$\leq 2\iiint_{\mathbb{R}\times[1,\infty)\times\mathbb{T}} \big(1 \wedge |zg(u)|\lambda^{(1-\delta)/2}\big)\,du\lambda^{-\zeta}\,d\lambda\pi_z(dz),$$

where $\lceil\lambda\rceil$ denotes the least integer $\ge \lambda$. By Lemma 1 this is finite for $0 \le \zeta \le 1$ if $\frac{\delta-1}{2} > 1-\zeta$ with $g \in \mathbb{B}^{s*}_{pq}$, so (18) holds and Proposition 1 ensures convergence.

The L^*_p norms of the mth forward differences of a periodic function $g(\cdot) \in \mathbb{B}^{s*}_{pq}(\mathbb{T})$ and their scaled translates $g(\lambda(\cdot-\chi))$ for $\chi \in \mathbb{T}$ and positive scale $\lambda \in [1,\infty)$ are related by

$$\|\Delta^m_h g(\lambda(\cdot-\chi))\|^*_p \le 2^{1/p}\|\Delta^m_{\lambda h} g\|^*_p \tag{42}$$

since, by a change of variables $x \mapsto u = \lambda(x-\chi)$,

$$\begin{aligned}
&\|\Delta^m_h g(\lambda(\cdot-\chi))\|^*_p \\
&\quad = \lambda^{-1/p}\left\{\int_{-\lambda\chi}^{\lambda(1-\chi)}\left|\sum_{k=0}^{m}\binom{m}{k}(-1)^{m-k}g(u+k\lambda h)\right|^p du\right\}^{1/p},
\end{aligned}$$

which, again from periodicity, satisfies

$$\begin{aligned}
&\le \left(\frac{\lceil\lambda\rceil}{\lambda}\right)^{1/p}\left\{\int_0^1\left|\sum_{k=0}^{m}\binom{m}{k}(-1)^{m-k}g(u+k\lambda h)\right|^p du\right\}^{1/p} \\
&= \left(\frac{\lceil\lambda\rceil}{\lambda}\right)^{1/p}\|\Delta^m_{\lambda h} g\|^*_p,
\end{aligned}$$

while $\lceil\lambda\rceil/\lambda \le 2$.

The Besov semi-norm of f is bounded by

$$\begin{aligned}
|f|^{s*}_{pq} &\le \sum_j |\beta_j|\lambda_j^{1/2}|g(\lambda_j(\cdot-\chi_j))|^{s*}_{pq} \\
&= \sum_j |\beta_j|\lambda_j^{1/2}\left(\int_{|h|\le 1}|h|^{-1-sq}\|\Delta^m_h g(\lambda_j(\cdot-\chi_j))\|^{*q}_p\,dh\right)^{1/q} \\
&\le \sum_j |\beta_j|\lambda_j^{1/2}\left(\frac{\lceil\lambda_j\rceil}{\lambda_j}\right)^{1/p}\left(\int_{|h|\le 1|}|h|^{-1-sq}\|\Delta^m_{\lambda_j h} g\|^{*q}_p\,dh\right)^{1/q} \\
&= \sum_j |\beta_j|\lambda_j^{s+1/2}\left(\frac{\lceil\lambda_j\rceil}{\lambda_j}\right)^{1/p}\left(\int_{|t|\le\lambda_j}|t|^{-1-sq}\|\Delta^m_t g\|^{*q}_p\,dt\right)^{1/q}. \qquad (43)
\end{aligned}$$

The integral in (43) is bounded by

$$\begin{aligned}
\int_{\mathbb{R}}|t|^{-1-sq}\|\Delta^m_t g\|^{*q}_p\,dt &= \int_{|t|\le 1}|t|^{-1-sq}\|\Delta^m_t g\|^{*q}_p\,dt \\
&\quad + \int_{|t|>1}|t|^{-1-sq}\|\Delta^m_t g\|^{*q}_p\,dt.
\end{aligned}$$

The first term is just $(|g|^{s*}_{pq})^q$, and (32) implies $\|\Delta^m_t g\|^*_p \le 2^m \|g\|^*_p$, so

$$\begin{aligned}
&\le (|g|^{s*}_{pq})^q + \int_{|t|>1} |t|^{-1-sq} (2^m \|g\|^*_p)^q \, dt \\
&= (|g|^{s}_{pq})^q + \frac{2^{1+mq}}{sq} \|g\|^{*q}_p \\
&\le (c\|g\|^{s*}_{pq})^q
\end{aligned}$$

for some $c < \infty$, so

$$|f|^{s*}_{pq} \le 2c\|g\|^{s*}_{pq} \sum_j |\beta_j| \lambda_j^{s+1/2} \tag{44}$$

is almost surely finite if and only if

$$\iiint_{\mathbb{R}\times[1,\infty)\times\mathbb{T}} (1 \wedge |\beta|\lambda^{s+1/2}) \nu(d\beta \, d\lambda \, d\chi)$$

is finite. Applying the change of variables $\beta \mapsto z = \lambda^{\delta/2}\beta$,

$$= \iiint_{\mathbb{R}\times[1,\infty)\times\mathbb{T}} (1 \wedge |z|\lambda^{s+(1-\delta)/2}) \lambda^{-\zeta} \pi_z(dz) \, d\lambda \, d\chi$$

is finite by Lemma 1 for all $\delta \ge 0$ if $\zeta > 1$ and for $\frac{\delta-1}{2} > s + 1 - \zeta$ if $0 \le \zeta \le 1$. A similar argument shows that the L^*_p norm of f satisfies a bound of the form

$$\|f\|^*_p \le c\|g\|^*_p \sum_j |\beta_j| \lambda_j^{1/2}$$

for some $c < \infty$. This is finite almost surely if

$$\begin{aligned}
&\iiint_{\mathbb{R}\times[1,\infty)\times\mathbb{T}} (1 \wedge |\beta|\lambda^{1/2}) \nu(d\beta \, d\lambda \, d\chi) \\
&\quad = \iiint_{\mathbb{R}\times[1,\infty)\times\mathbb{T}} (1 \wedge |z|\lambda^{(1-\delta)/2}) \lambda^{-\zeta} \pi_Z(dz) \, d\lambda \, d\chi
\end{aligned}$$

is finite, which follows from Lemma 1 for all $\delta \ge 0$ if $\zeta > 1$ and, if $\zeta \le 1$, for δ satisfying $\frac{\delta-1}{2} > 1 - \zeta$ since $g \in \mathbb{B}^{s*}_{pq} \subset L^*_p$. Combining conditions, the $\mathbb{B}^{s*}_{pq}$ norm of f is finite if $\delta/2 - 1/2 > s + 1 - \zeta$ for $0 \le \zeta \le 1$ and for all $\delta \ge 0$ if $\zeta > 1$. □

For Lévy measures $\nu(d\beta \, d\lambda \, d\chi)$ supported on $\mathbb{R} \times \mathbb{N} \times \mathbb{T}$ (i.e., for which λ is almost-surely integral) the function $f(x)$ of (39) would inherit periodicity from the generator $g(\lambda_j(x - \chi_j))$ but, for the absolutely-continuous measure of (38), it is the definition of $f(x)$ as a function on $\mathbb{T}$ [as in Abramovich, Sapatinas and Silverman (2000), equation (2)] that induces periodicity. The restriction to $\lambda \ge 1$ may be relaxed to the more natural $\lambda > 0$ in the LARK framework, but may require the use of compensation.

4.3. *Compensation*. For Lévy measures satisfying only the local-L_2 bound of (9) and not the local-L_1 bound of (7), we must use the definition of $f(x)$ in (14) and use (16) to establish conditions that ensure f will be well defined for $g \in \mathbb{B}^s_{pq}$. We verify these conditions for the existence of LARK models under symmetric α-Stable random fields.

THEOREM 5. *For a Symmetric α-Stable random field with Lévy measure of the form $\nu(d\beta\, d\omega) = c_\alpha \alpha |\beta|^{-1-\alpha}\, d\beta \pi(d\Lambda)\, d\chi$ on $\mathbb{R} \times \mathcal{S}^d_+ \times \mathbb{R}^d$ for $0 < \alpha < 2$, with $\pi(d\Lambda)$ a probability measure on $\mathcal{S}^d_+$ and $g \in \mathbb{B}^s_{pq}(\mathbb{R}^d) \cap L_1(\mathbb{R}^d)$ for $p, q \geq 1$ and $s > 0$, the conditions of (16) for $f(x)$ to be well defined by Theorem 1 are satisfied for $1 < \alpha \leq p$, $\alpha < 2$ if $\mathsf{E}[|\Lambda|^{-1}] < \infty$. For $\alpha = 1$, there is the additional requirement that*

$$\int_{\mathbb{R}^d} |g(u) \log|g(u)||\, du < \infty. \tag{45}$$

PROOF. Fix $x \in \mathcal{X}$. By the affine change of variables of $\chi \mapsto u \equiv \Lambda(x - \chi)$,

$$\begin{aligned}
&\iint_{[-1,1]^c \times \Omega} \bigl(1 \wedge |\beta \phi(x, \omega)|\bigr) \nu(d\beta\, d\omega) \\
&\quad = 2c_\alpha \alpha \int_{\mathcal{S}^d_+} |\Lambda|^{-1} \pi(d\Lambda) \iint_{[1,\infty) \times \mathbb{R}^d} \bigl(1 \wedge \beta |g(u)|\bigr) \beta^{-1-\alpha}\, d\beta\, du \\
&\quad = 2c_\alpha \alpha \mathsf{E}|\Lambda|^{-1} \int_{\mathbb{R}^d} \biggl\{ \int_1^{|g(u)|^{-1}} \beta^{-\alpha} |g(u)|\, d\beta + \int_{|g(u)|^{-1}}^\infty \beta^{-1-\alpha}\, d\beta \biggr\}\, du.
\end{aligned}$$

For $1 < \alpha < 2$,

$$= 2c_\alpha \alpha \mathsf{E}|\Lambda|^{-1} \biggl\{ \int_{\mathbb{R}^d} \frac{|g(u)| - |g(u)|^\alpha}{\alpha - 1}\, du + \int_{\mathbb{R}^d} \frac{|g(u)|^\alpha}{\alpha}\, du \biggr\},$$

which is finite for $1 < \alpha \leq p$ since $g \in L_1$ and $g \in \mathbb{B}^s_{pq} \subset L_p$. For $\alpha = 1$,

$$= 2c_1 \mathsf{E}|\Lambda|^{-1} \biggl\{ \int_{\mathbb{R}^d} -|g(u)| \log|g(u)|\, du + \int_{\mathbb{R}^d} |g(u)|\, du \biggr\}.$$

The first integral exists and is finite by (45) while the second is finite since $g \in L_1$. Similarly, the integral in (16b) is

$$\begin{aligned}
&\iint_{[-1,1] \times \Omega} \bigl(|\beta \phi(x, \omega)| \wedge |\beta \phi(x, \omega)|^2\bigr) \nu(d\beta\, d\omega) \\
&\qquad = 2c_\alpha \alpha \mathsf{E}|\lambda|^{-1} \biggl\{ \iint_{[0,1) \times \mathbb{R}^d} \bigl(|\beta g(u)| \wedge |\beta g(u)|^2\bigr) \beta^{-1-\alpha}\, d\beta\, du \biggr\}.
\end{aligned}$$

The integral in braces

$$\iint_{[0, 1 \wedge |g(u)|^{-1}] \times \mathbb{R}^d} \beta^{1-\alpha} g(u)^2\, d\beta\, du + \iint_{[1 \wedge |g(u)|^{-1}, 1] \times \mathbb{R}^d} \beta^{-\alpha} |g(u)|\, d\beta\, du$$

is finite for $1 < \alpha \le p, \alpha < 2$:

$$\le \int_{\mathbb{R}^d} \frac{|g(u)|^\alpha}{2-\alpha} + \frac{|g(u)|^\alpha - |g(u)|}{\alpha - 1}\, du$$
$$\le \frac{\|g\|_p^p}{(2-\alpha)(\alpha-1)} < \infty,$$

while for $\alpha = 1$,

$$\le \int_{\mathbb{R}^d} \{|g(u)| + |g(u) \log|g(u)||\}\, du < \infty$$

by (45). Finally, (16c) holds because

$$\begin{aligned}
&\iint_{\mathbb{R}\times\Omega} (1 \wedge \beta^2)|\phi(x,\omega)|\nu(d\beta\, d\omega) \\
&\quad = \mathsf{E}|\Lambda|^{-1} c_\alpha \alpha \iint_{\mathbb{R}\times\mathbb{R}^d} (1\wedge\beta^2)|\beta|^{-1-\alpha}|g(u)|\, d\beta\, du \\
&\quad = \mathsf{E}|\Lambda|^{-1}\|g\|_1 c_\alpha \alpha \int_{\mathbb{R}} (1\wedge\beta^2)|\beta|^{-1-\alpha}\, d\beta < \infty.
\end{aligned}$$

□

All of the generator functions in the examples in Section 7 satisfy the conditions of the theorem for the Cauchy random field ($\alpha = 1$), so the LARK models are well defined as $\varepsilon \to 0$ and for finite $\varepsilon > 0$, the approximations are in the same Besov space as g. We are able to show that this also holds for Sobolev $\mathbb{W}_2^s$ spaces (which are equivalent to $\mathbb{B}_{22}^s$) even when compensation is required, but this remains an open question for $\mathbb{B}_{pq}^s$ with general p and q.

4.4. *Convergence in* $\mathbb{W}_2^s$.

THEOREM 6. *Let $\{\phi(x,\omega)\}$ be a location-scale family of the form $\phi(x,\omega) \equiv g(\Lambda(x-\chi))$ for $\omega = (\chi, \Lambda)$ with $\chi \in \mathbb{R}^d$ and nonsingular $d\times d$ matrix $\Lambda \in \mathcal{S}_+^d$ for some function $g(\cdot) \in \mathbb{W}_2^s$ with $s \ge 0$. Let ν be a Lévy measure satisfying the condition*

$$\iint_{\mathbb{R}\times\Omega} |\Lambda|^{-1}[1+\rho(\Lambda)^{2s}](1\wedge\beta^2)\nu(d\beta\, d\omega) < \infty, \tag{46}$$

where $\rho(\Lambda)$ denotes the spectral radius (*largest eigenvalue*) *of Λ. Recall*

$$\begin{aligned}
f(x) \equiv &\iint_{\mathbb{R}\times\Omega} \phi(x,\omega)[\beta - h(\beta)]\mathcal{N}(d\beta\, d\omega) \\
&+ \iint_{\mathbb{R}\times\Omega} \phi(x,\omega) h(\beta)\tilde{\mathcal{N}}(d\beta\, d\omega)
\end{aligned} \tag{19}$$

and, for $\varepsilon > 0$, *define*

$$
\begin{aligned}
f_\varepsilon(x) &\equiv \iint_{[-\varepsilon,\varepsilon]^c\times\Omega} \phi(x,\omega)[\beta - h(\beta)]\mathcal{N}(d\beta\, d\omega) \\
&\quad + \iint_{[-\varepsilon,\varepsilon]^c\times\Omega} \phi(x,\omega)h(\beta)\tilde{\mathcal{N}}(d\beta\, d\omega) \qquad (47)\\
&= \sum_{\substack{0\le j<J_\varepsilon\\ \varepsilon<|\beta_j|}} \phi(x,\omega_j)\beta_j - \iint_{[-\varepsilon,\varepsilon]^c\times\Omega} \phi(x,\omega)h(\beta)\eta(d\beta\, d\omega).
\end{aligned}
$$

Then $f_\varepsilon(\cdot) \to f(\cdot)$ *in* $\mathbb{W}_2^s$ *almost surely as* $\varepsilon \to 0$.

PROOF. First, consider the case of compensator functions satisfying $h(\beta) = \beta$ for all $|\beta| \le 1$. Apply an affine change of variables to see that $\phi(x,\omega)$ has Fourier transform (in x)

$$\hat{\phi}(\xi,\omega) = e^{i\xi\cdot\chi}|\Lambda|^{-1}\hat{g}(\Lambda^{-1}\xi).$$

For $0 < \varepsilon_1 < \varepsilon_2 < 1$ and $x \in \mathbb{R}^d$, set $\Delta(x) \equiv f_{\varepsilon_1}(x) - f_{\varepsilon_2}(x)$ and let $A \equiv \{\varepsilon_1 < |\beta| \le \varepsilon_2\} \times \Omega$. Then

$$\Delta(x) = \sum_{\substack{0\le j<J_{\varepsilon_1}\\ \varepsilon_1<|\beta_j|\le\varepsilon_2}} \phi(x,\omega_j)\beta_j - \iint_A \phi(x,\omega)\beta\nu(d\beta\, d\omega)$$

is a zero-mean random function of x with Fourier transform

$$
\begin{aligned}
\widehat{\Delta}(\xi) = &\sum_{\substack{0\le j<J_{\varepsilon_1}\\ \varepsilon_1<|\beta_j|\le\varepsilon_2}} e^{i\xi\cdot\chi_j}|\Lambda_j|^{-1}\hat{g}(\Lambda_j^{-1}\xi)\beta_j \\
&- \iint_A e^{i\xi\cdot\chi}|\Lambda|^{-1}\hat{g}(\Lambda^{-1}\xi)\beta\nu(d\beta\, d\omega),
\end{aligned}
$$

a zero-mean L_2 random function of ξ with second moment

$$\mathsf{E}|\widehat{\Delta}(\xi)|^2 = \iint_A |\Lambda|^{-2}|\hat{g}(\Lambda^{-1}\xi)|^2\beta^2\nu(d\beta\, d\omega). \qquad (48)$$

Thus $\Delta(\cdot)$ has expected squared Sobolev norm $\mathsf{E}\|f_{\varepsilon_1} - f_{\varepsilon_2}\|^2_{\mathbb{W}_2^s}$:

$$
\begin{aligned}
&= (2\pi)^{-d}\iiint_{\mathbb{R}^d\times A} (1+|\xi|^2)^s|\Lambda|^{-2}|\hat{g}(\Lambda^{-1}\xi)|^2\beta^2\nu(d\beta\, d\omega)\, d\xi \\
&= (2\pi)^{-d}\iiint_{\mathbb{R}^d\times A} (1+|\Lambda\eta|^2)^s|\Lambda|^{-1}|\hat{g}(\eta)|^2\beta^2\nu(d\beta\, d\omega)\, d\eta \\
&\le (2\pi)^{-d}\iiint_{\mathbb{R}^d\times A} (1+|\eta|^2)^s\big[\big(1+\rho(\Lambda)\big)^{2s}\big]|\Lambda|^{-1}|\hat{g}(\eta)|^2\beta^2\nu(d\beta\, d\omega)\, d\eta
\end{aligned}
$$

$$= \|G\|^2_{\mathbb{W}_2^s} \iint_{\{\varepsilon_1 < |\beta| \le \varepsilon_2\} \times \Omega} \big[\big(1+\rho(\Lambda)\big)^{2s}\big] |\Lambda|^{-1} \beta^2 \nu(d\beta\, d\omega) \tag{49}$$

$$\to 0 \qquad \text{as } \varepsilon_1, \varepsilon_2 \to 0 \text{ by (46)},$$

so $\{f_{\varepsilon_k}\}$ is a Cauchy sequence in $\mathbb{W}_2^s$ for any $\varepsilon_k \to 0$ and $\|f - f_{\varepsilon_k}\|_{\mathbb{W}_2^s} \to 0$. Since f_ε is a finite linear combination of scaled translates of $g \in \mathbb{W}_2^s$, each f_ε (and hence f) lies in $\mathbb{W}_2^s$ almost surely and Theorem 6 is proved for compensator functions satisfying $h(\beta) = \beta$ for $|\beta| < 1$.

For an arbitrary bounded compensator $h(\beta)$ satisfying $|\beta - h(\beta)| \le c\beta^2$ for some $c > 0$, (48) has the additional nonrandom term

$$\left| \iint_A e^{i\xi\cdot\chi} \frac{\hat{g}(\Lambda^{-1}\xi)}{|\Lambda|} \big(\beta - h(\beta)\big) \nu(d\beta\, d\omega) \right|^2 \le c \left(\iint_A \frac{|\hat{g}(\Lambda^{-1}\xi)|}{|\Lambda|} \beta^2 \nu(d\beta\, d\omega) \right)^2$$

leading at most to an additional constant factor of $[1 + c \iint_{\mathbb{R}\times\Omega} (1 \wedge \beta^2) \nu(d\beta\, d\omega)]$ in (49), leading as before to $\|f - f_{\varepsilon_k}\|_{\mathbb{W}_2^s} \to 0$ and completing the proof. $\square$

COROLLARY 4. *If $\{\phi(x,\omega)\}$ is a location-scale family of the form considered in Theorem 6 and if a Lévy measure ν is of product form $\nu(d\beta\, d\omega) = \nu_\beta(d\beta)\pi_\omega(d\omega)$ for some σ-finite measure $\nu_\beta(d\beta)$ on $\mathbb{R}$ and probability measure $\pi_\omega(\cdot)$ on Ω that for some $s \ge 0$ satisfy*

$$\int_{\mathbb{R}} (1 \wedge \beta^2) \nu_\beta(d\beta) < \infty, \tag{50a}$$

$$\int_\Omega |\Lambda|^{-1} \big(\big(1 + \rho(\Lambda)\big)^{2s}\big) \pi_\omega(d\omega) < \infty, \tag{50b}$$

then $\nu(d\beta\, d\omega)$ also satisfies (46) and hence $f_\varepsilon(\cdot) \to f(\cdot)$ in $\mathbb{W}_2^s$ almost surely as $\varepsilon \to 0$.

For example, in one dimension, (50b) is satisfied for all $s > 0$ if $\Lambda = \lambda$ has the χ_ν distribution with $\nu > 1$ degrees of freedom, that is, if $\lambda^2 \sim \mathsf{Ga}(\alpha_\lambda, \beta_\lambda)$ with $\alpha_\lambda > \frac{1}{2}$. More generally, for any $m > 0$ (50b) is satisfied for all $s > 0$ if $\lambda^m \sim \mathsf{Ga}(\alpha_\lambda, \beta_\lambda)$ with $\alpha_\lambda > 1/m$ or, for $m < 0$, for $\alpha_\lambda > (1 - 2s)/m$.

Recall that the quantity ε introduced in the proof of Theorem 6 and the statement of Corollary 4 is *not* a model parameter and has no bearing on the Sobolov spaces to which the limiting function $f(\cdot)$ belongs; it is only a tool used in proofs and implementations, to which we now turn.

5. Inference for LARK models. The LARK model introduced in Section 1 may now be summarized as

$$\begin{aligned} \mathsf{E}[Y(x) \mid \mathcal{L}, \theta] = f(x) &\equiv \int_\Omega \phi(x,\omega) \mathcal{L}(d\omega), \\ \mathcal{L} \mid \theta &\sim \mathsf{Lévy}(\nu), \\ \theta &\sim \pi_\theta(d\theta) \end{aligned} \tag{51}$$

with implicit dependence of the Lévy measure $\nu(d\beta\, d\omega)$ and conditional distribution for $Y(x)$ on a hyperparameter vector θ. In all of our examples, we take ν to be a product measure $\nu(d\beta\, d\omega) = \nu_\beta(\beta)\, d\beta |\Omega| \pi_\omega(d\omega)$ satisfying the conditions of Corollary 2, with $\pi_\omega(\cdot)$ a probability measure on Ω, $|\Omega|$ a measure of the volume of Ω, and $\nu_\beta(\cdot) > 0$ a nonnegative density function on $\mathbb{R}$ satisfying $\int_{\mathbb{R}} (1 \wedge \beta^2) \nu_\beta(\beta)\, d\beta < \infty$ [so ν satisfies (9)], for which either (a) ν also satisfies (7) or (b) $\nu_\beta(\beta)$ is even and $h(\beta)$ is odd in β. Thus, we have the representation

$$\theta \sim \pi_\theta(d\theta), \tag{52a}$$

$$J \mid \theta \sim \mathsf{Po}(\nu_\varepsilon^+), \qquad \nu_\varepsilon^+ \equiv \nu_\varepsilon(\mathbb{R} \times \Omega), \tag{52b}$$

$$\begin{aligned} \{(\beta_j, \omega_j)\}_{0 \le j < J} \mid J, \theta &\overset{\text{i.i.d.}}{\sim} \pi_\beta(\beta_j)\, d\beta_j \pi_\omega(d\omega_j), \\ \pi_\beta(\beta) &\equiv \mathbf{1}_{\{|\beta| > \varepsilon\}} \nu_\beta(\beta) |\Omega| / \nu_\varepsilon^+, \end{aligned} \tag{52c}$$

$$\begin{aligned} Y_i \mid f &\overset{\text{ind}}{\sim} p_Y(y \mid f(x_i))\, dy, \\ f(x_i) &\equiv \sum_{0 \le j < J} \phi(x_i, \omega_j) \beta_j \end{aligned} \tag{52d}$$

for sampling model $p_Y(\cdot \mid \mu)$ parametrized by μ.

5.1. *Examples of Lévy random fields.* Motivated by the applications in Section 8, we now focus on LARK models built on approximations to Gamma, symmetric Gamma and Symmetric α-Stable (in particular, Cauchy) Lévy random fields, and quantify the approximation errors to facilitate the selection of ε and other prior hyperparameters.

5.1.1. *Gamma LARK models.* The Gamma random field of Section 2.5.2 has $\nu_\beta(d\beta) = \gamma \beta^{-1} e^{-\beta\eta} \mathbf{1}_{\{\beta > 0\}}\, d\beta$ for some constants $\gamma > 0$ and $\eta > 0$. The parameter η in (22) controls both the Poisson rate of mass points $\{(\beta_j, \omega_j)\}$ of magnitude $|\beta| > \varepsilon$ and the probability distribution of those magnitudes $\{\beta_j\}$. To facilitate elicitation we disentangle those two roles by truncating at $|\beta\eta| \ge \varepsilon$ (rather than $|\beta| \ge \varepsilon$); of course the limit as $\varepsilon \to 0$ is the same. The distributions of J and $\{\beta_j\}$ are now given by

$$\begin{aligned} J &\sim \mathsf{Po}(\nu_\varepsilon^+), \qquad \nu_\varepsilon^+ = \gamma |\Omega| \mathrm{E}_1(\varepsilon), \\ \beta_j &\overset{\text{i.i.d.}}{\sim} \pi_\beta(\beta_j)\, d\beta_j, \qquad \pi_\beta(\beta_j) = \frac{\beta_j{}^{-1} e^{-\beta_j \eta}}{\mathrm{E}_1(\varepsilon)} \mathbf{1}_{\{\beta_j \eta > \varepsilon\}}, \end{aligned}$$

where the exponential integral function [Abramowitz and Stegun (1964), page 228] is denoted as $\mathrm{E}_1(z) \equiv \int_z^\infty t^{-1} e^{-t}\, dt$. With this truncation, the expected square L_2 norm of the loss due to truncation for any $\phi \in L_2(\Omega, |\Omega| \pi_\omega(d\omega))$, such as $\phi(\omega) = \phi(x, \omega)$, is

$$\mathsf{E}|\mathcal{L}[\phi] - \mathcal{L}_\varepsilon[\phi]|^2 = \iint_{\mathbb{R}\times\Omega} \phi(\omega)^2 |\beta|^2 \mathbf{1}_{\{|\beta\eta|\le\varepsilon\}} \nu(d\beta\, d\omega)$$

$$\text{(53a)} \qquad = \|\phi\|_2^2 \int_0^{\varepsilon/\eta} \beta^2 \nu_\beta(\beta)\, d\beta$$

$$= \gamma\eta^{-2}\|\phi\|_2^2[1 - (1+\varepsilon)e^{-\varepsilon}],$$

showing the rate at which $\mathcal{L}_\varepsilon[\phi] \to \mathcal{L}[\phi]$ in L_2 as $\varepsilon \to 0$. This is used in Section 5.2 to guide the elicitation of hyperparameters.

5.1.2. *Symmetric Gamma LARK models*. The symmetric Gamma random field of Section 2.5.3 has Lévy measure $\nu_\beta(d\beta) = \gamma|\beta|^{-1}e^{-|\beta|\eta}\, d\beta$ for some constants $\gamma > 0$ and $\eta > 0$. Once again truncation at $|\beta\eta| > \varepsilon$ leads to

$$J \sim \mathsf{Po}(\nu_\varepsilon^+), \qquad \nu_\varepsilon^+ = 2\gamma|\Omega|\mathrm{E}_1(\varepsilon)$$

$$\beta_j \overset{\text{i.i.d.}}{\sim} \pi_\beta(\beta_j)\, d\beta_j, \qquad \pi_\beta(\beta_j) = \frac{|\beta_j|^{-1}e^{-|\beta_j|\eta}}{2\mathrm{E}_1(\varepsilon)} \mathbf{1}_{\{|\beta_j\eta|>\varepsilon\}}$$

and expected squared discrepancy (used for elicitation)

$$\text{(53b)} \qquad \mathsf{E}|\mathcal{L}[\phi] - \mathcal{L}_\varepsilon[\phi]|^2 = 2\gamma\eta^{-2}\|\phi\|_2^2[1 - (1+\varepsilon)e^{-\varepsilon}].$$

5.1.3. *Symmetric α-Stable LARK models*. The $\mathsf{S}\alpha\mathsf{S}$ Lévy random field of Section 2.5.4 has $\nu_\beta(d\beta) = \frac{\dot\gamma\alpha}{\pi}\Gamma(\alpha)\sin\frac{\pi\alpha}{2}|\beta|^{-\alpha-1}\, d\beta$ for some constants $\dot\gamma > 0$ and $0 < \alpha < 2$. To facilitate elicitation and posterior inference, we write $\dot\gamma = \gamma\eta^{-\alpha}$ and (again) truncate at $|\beta_j\eta| > \varepsilon$. This leads to

$$J \sim \mathsf{Po}(\nu_\varepsilon^+), \qquad \nu_\varepsilon^+ = \gamma|\Omega|\frac{2}{\pi}\Gamma(\alpha)\sin\frac{\pi\alpha}{2}\varepsilon^{-\alpha}$$

$$\beta_j \overset{\text{i.i.d.}}{\sim} \pi_\beta(\beta_j)\, d\beta_j, \qquad \pi_\beta(\beta_j) = \frac{\alpha\varepsilon^\alpha}{2\eta^\alpha}|\beta_j|^{-\alpha-1}\mathbf{1}_{\{|\beta_j\eta|>\varepsilon\}}$$

with symmetric Pareto distributions for the coefficients $\{\beta_j\}$. For the Cauchy ($\alpha = 1$), these simplify to $\nu_\varepsilon^+ = 2\gamma|\Omega|/(\pi\varepsilon)$, with

$$\pi_\beta(\beta_j) = \frac{\varepsilon}{2\eta}|\beta_j|^{-2}\mathbf{1}_{\{|\beta_j\eta|>\varepsilon\}}.$$

Although the total variation $|\mathcal{L}|$ is almost surely infinite, and even $|\mathcal{L} - \mathcal{L}_\varepsilon|$ will be infinite for $\alpha \ge 1$, still for $\phi \in L_2(\Omega, |\Omega|\pi_\omega(d\omega))$ the expected squared discrepancy is finite:

$$\mathsf{E}|\mathcal{L}[\phi] - \mathcal{L}_\varepsilon[\phi]|^2 = \iint_{\mathbb{R}\times\Omega} \phi(\omega)^2 |\beta|^2 \mathbf{1}_{\{|\beta\eta|\le\varepsilon\}} \nu(d\beta\, d\omega)$$

$$\text{(53c)} \qquad = 2\gamma\eta^{-2}\|\phi\|_2^2\left[\frac{\Gamma(\alpha+1)}{\pi(2-\alpha)}\sin\frac{\pi\alpha}{2}\varepsilon^{2-\alpha}\right]$$

or $2\gamma\eta^{-2}\|\phi\|_2^2[\varepsilon/\pi]$ for the Cauchy case $\alpha = 1$.

5.2. *Prior elicitation of hyperparameters*. We now turn to the selection of $\varepsilon > 0$, the vector $\theta \in \Theta$ of (52), and the Lévy measure $\nu(d\beta\, d\omega)$. In each of our examples $\theta \equiv (\gamma, \eta)$ for rate parameters γ and η governing the frequency and magnitude of coefficients $\{\beta_j\}$, respectively, and the expected squared truncation error for $\mathcal{L}_\varepsilon[\phi]$ for $\phi(\omega) = \phi(x, \omega)$ is of the form $\mathsf{E}|\mathcal{L}[\phi(x, \cdot)] - \mathcal{L}_\varepsilon[\phi(x, \cdot)]|^2 = \gamma\eta^{-2}\|\phi(x, \cdot)\|_2^2 c(\varepsilon)$ for some $c(\varepsilon) > 0$ with $c(\varepsilon) \to 0$ as $\varepsilon \to 0$ [see (53)].

We choose prior distributions to attain three goals: (1) desired range of number J of terms in the stochastic expansion; (2) desired range of coefficient magnitudes $\{\beta_j\}$; and (3) tolerable expected truncation error. We first select a Lévy family (Gamma, α-Stable, etc.) to meet the needs of a particular problem for symmetry or positivity, sharp or heavy tails, etc. Each of our Lévy measures is of the product form $\nu(d\beta\, d\omega) = \nu_\beta(d\beta)\pi_\omega(d\omega)$ considered in Theorem 6 and Corollary 4, with location, scale, and perhaps other location-specific (and hence adaptive) attributes encoded in $\omega \in \Omega$ in problem-specific ways.

Hyperparameters in the Lévy measure $\nu_\beta(d\beta)$ govern sparseness for LARK models, that is, the number J of terms in the stochastic expansion. In each LARK model, J has a Poisson distribution with mean proportional to γ. The coefficient of variation under the Poisson distribution falls to zero as the mean increases, overstating the prior certainty for large values of $\mathsf{E}J$. To ameliorate this, we introduce an additional layer of hierarchy by placing a Gamma prior distribution on the parameter $\gamma \sim \mathsf{Ga}(a_\gamma, b_\gamma)$, leading to the overdispersed negative binomial prior distribution for $J \sim \mathsf{NB}(a_J, p_J)$. The parameter η governs the scale of the coefficients $\{\beta_j\}$, and hence the range of the regression function $f(\cdot)$. We employ a Gamma distribution for the scale parameter $\eta^{-1} \sim \mathsf{Ga}(a_\eta, b_\eta)$. Together the hyperparameters ε, a_γ, b_γ, a_η, b_η determine the prior distributions for J, for the coefficients $\{\beta_j\}$ (and hence the range of $f(\cdot)$), and for the expected mean-square truncation error. We select values for these five parameters to meet five criteria: attain two specified quantiles (such as a central 99% interval) for each of J and $\{\beta_j\}$, and a specified bound on the expected truncation error $\mathsf{E}\gamma\eta^{-2}\|\phi(x, \cdot)\|_2^2 c(\varepsilon)$. Typically this involves an iterative numerical solution.

As a default choice, we take $\pi(d\omega) = \pi_\chi(d\chi)\pi_\lambda(d\lambda)$ to be the product of the uniform distribution for locations $\chi \sim \mathsf{Un}(\mathcal{X})$ and a Gamma distribution for inverse (distance) scale parameters $\lambda \sim \mathsf{Ga}(a_\lambda, b_\lambda)$. The shape and rate hyperparameters a_λ and b_λ govern the range of probable values for the location-specific inverse scale parameters $\{\lambda_j\}$ and hence for the smoothness of $f(x)$, similar to how bandwidth selection governs smoothness in other kernel methods. A kernel at $\omega_j = (\chi_j, \lambda_j)$ will represent a feature located at χ_j of width $1/\lambda_j$, so large values of λ_j are needed to fit a very "spiky" part of a curve, while a smoother part of a curve may be fit most parsimoniously using small values of λ_j. The prior distribution for λ_j must support an adequate range of values in order to fit a spatially inhomogeneous curve. Values of $a_\lambda > 1$ will ensure $\mathsf{E}[\lambda] < \infty$ and a finite covariance function; we choose (a_λ, b_λ) to attain two specified quantiles, such as a central 99% interval.

5.3. *Posterior inference.* The joint posterior density of all parameters under the LARK model of (52), given observations $\mathbf{Y} = \{Y_i\}$, is

$$\begin{aligned} &p(\gamma, \eta, J, \boldsymbol{\beta}, \boldsymbol{\omega} \mid \mathbf{Y}) \\ &\quad \propto \pi_\gamma(\gamma)\pi_\eta(\eta)\frac{\exp[-\nu_\varepsilon(\mathbb{R} \times \Omega)]}{J!} \\ &\qquad \times \left\{\prod_{0 \le j < J} \nu_\varepsilon(\beta_j, \omega_j)\right\}\left\{\prod_{i \in I} p_Y\left(Y_i \,\Big|\, \sum_{0 \le j < J} \phi(x_i, \omega_j)\beta_j\right)\right\}. \end{aligned} \tag{54}$$

The posterior (and full conditional) distributions of the parameters are not available in closed form. Since some of our parameters ($\boldsymbol{\beta}$ and $\boldsymbol{\omega}$) have varying dimension, some form of trans-dimensional Markov chain Monte Carlo, such as a reversible jump (RJ-MCMC) algorithm [Green (1995), Wolpert, Ickstadt and Hansen (2003), Sisson (2005)] must be used to provide samples from (54) for posterior inference. See Appendix B for a sketch of the RJ-MCMC algorithm.

6. Relation of LARK to other models.

6.1. *Gaussian processes or random fields.* For any positive Borel measure $\Sigma(d\omega)$ on a complete separable metric space Ω, there exists a Gaussian random measure $\mathcal{Z}(d\omega)$ on Ω that assigns to disjoint Borel sets $A_i \subset \Omega$ of finite measure $\Sigma(A_i) < \infty$ independent mean-zero Gaussian random variables $\mathcal{Z}(A_i) \sim \mathsf{No}(0, \Sigma(A_i))$ of variance $\mathsf{E}\mathcal{Z}(A_i)^2 = \Sigma(A_i)$. For any kernel function g on $\mathcal{X} \times \Omega$ with $\phi(x, \cdot) \in L_2(\Omega, \Sigma(d\omega))$ for each $x \in \mathcal{X}$, this induces a mean-zero Gaussian random field through the Wiener stochastic integral

$$f(x) = \int_\Omega \phi(x, \omega)\mathcal{Z}(d\omega)$$

with covariance $C(x, y) = \mathsf{E}[f(x)f(y)] = \int_\Omega \phi(x, \omega)\phi(y, \omega)\Sigma(d\omega)$. The Gaussian random measure $\mathcal{Z}(d\omega)$ is the special case of a Lévy random measure $\mathcal{L}(d\omega)$ defined earlier in (8) with $\delta(d\omega) \equiv 0$ and $\nu(d\beta\, d\omega) \equiv 0$.

A wide variety of Gaussian processes are available in this form. For example, those with stationary covariance $C(x, y) = c(x - y)$ may be written in the above form if the spectral measure has a density function $\hat{c}(\omega) = \int_{\mathcal{X}} e^{-i\omega \cdot x} c(x)\, dx$ whose square root is Lebesgue integrable, for example, the Matérn class [Stein (1999), page 31] in $\mathbb{R}^d$ with smoothness parameter $\nu > d/2$. The Gaussian random field model above may also be obtained as the limit as $\alpha \to 2$ of the symmetric α-Stable LARK models considered herein, providing an alternative method for inference that avoids the need for large matrix inversions. To maintain a unified computational approach, we have limited our attention in this article to LARK models with pure-jump Lévy random measures, that is, $\Sigma(\cdot) \equiv 0$.

6.2. *Compound Poissons and mixtures of Gaussian random fields*. Mixtures of Gaussian random fields may be constructed as LARK models with Lévy measure of the form

$$\nu(d\beta\, d\omega) = (2\pi\sigma_\omega^2)^{-1/2} e^{-\beta^2/2\sigma_\omega^2}\, d\beta\, \nu_\omega(d\omega) \tag{55}$$

leading to mean functions of the form $f(x_i) = \sum_{0\le j<J} \phi(x, \omega_j)\beta_j$ with normally-distributed coefficients $\beta_j|\omega \sim \mathsf{No}(\mu_\omega, \sigma_\omega^2)$. For finite measures ν_ω, the expansion has a Poisson-distributed number of terms, hence, is a Poisson mixture of Gaussian processes (or for hierarchical models with a Gamma distributed Poisson mean, a negative binomial mixture of Gaussian processes). In Section 4.2, we showed that the stochastic wavelet expansion of Abramovich, Sapatinas and Silverman (2000), an example of (55), may be viewed as a LARK model. Chu, Clyde and Liang (2009) extend the compound Poisson (or LARK with finite ν) model to include mixtures of normals distributions for β_ω and develop methods for Bayesian inference for such OverComplete Wavelet expansions (OCW); we compare the OCW method to other LARK models in the simulation study of Section 7.

For automatic curve fitting using splines and wavelets, Denison et al. [(2002), Chapter 3] used a similar hierarchical model with common $\sigma_\omega \equiv \sigma$, but truncated the (Poisson-distributed) number of terms in the basis expansions at some fixed upper bound J_u. Taking $J_u \to \infty$ leads to the Gaussian LARK model of (55) with a common variance. Gaussian processes have sharp tails, of course, leading to concerns about robustness when they are used as prior distributions in problems with likelihood functions that fall off more slowly. Specifying variances for Gaussian prior distributions is nontrivial, with large "noninformative" choices leading to the so-called Lindley paradox. Denison et al. recommend an inverse Gamma prior on σ^2 to avoid this well-known problem. This leads to a multivariate Student t distribution on the expansion coefficients and, since the prior now has bounded influence, provides robustness. The limiting model (as $J_u \to \infty$) may be viewed as a mixture of Lévy random fields.

Rather than using a multivariate Student t for the coefficients, one might use "ridge" priors and model the uncertain function $f(\cdot) = \sum_{0\le j<J} \beta_j \phi(\cdot;\omega_j)$ as the sum of a Poisson (or negative binomial)-distributed number J of kernel functions $\phi(\cdot;\omega_j)$ with coefficients $\beta_j \overset{\text{i.i.d.}}{\sim} \mathsf{C}(0,\tau)$ drawn from a centered Cauchy distributions with scale τ. To accommodate rough functions $f(\cdot)$, one must be willing to consider large numbers of terms, most of which will have small coefficients—under these priors, one must consider large $\mathsf{E}J$ and small τ. But how small? And what happens if τ is made a bit smaller and $\mathsf{E}J$ a bit larger? As $\tau \to 0$, if one scales the expected number $\mathsf{E}J$ of terms (as a function of τ) properly, this model converges to a LARK model with infinite Lévy measure (and so is *not* sensitive to the cut-off ε, which merely quantifies how close is this approximation). If $\mathsf{E}J$ is not scaled properly to converge to a LARK model, the limiting results may depend critically on arbitrary and unintentional choices.

This may be implemented explicitly in LARK form by placing independent $\mathsf{Ga}(\alpha/2, \varepsilon/2)$ prior distributions on σ_ω^{-2} in (55) to achieve independent univariate Student $t_\alpha(0, \varepsilon)$ distributions for the coefficients $\{\beta_j\}$ and (approximately, as the parameter $\varepsilon \to 0$) the heavy-tailed Symmetric α-Stable process for $f(x)$ of Sections 2.5.4 and 5.1.3 [this also illustrates that truncating the support of β_ω is not the only way to construct suitable approximating sequences of finite Lévy measures $\nu_\varepsilon(d\beta\, d\omega) \Rightarrow \nu(d\beta\, d\omega)$ for which the integrals in (27) converge]. An important feature of our infinitely divisible construction (in contrast to a compound Poisson approach from other distributional families) is that in each case, as $\varepsilon \to 0$ the approximating model converges to one with a well-defined prior (with infinite Lévy measure) and a proper posterior distribution.

6.3. *Finite dimensional frames*. LARK may be viewed as a limit of Bayesian variable selection methods with finite frames or dictionaries. Wolfe, Godsill and Ng (2004) consider frames based on discretizing Ω as a fine grid with $|G|$ elements. They place i.i.d. prior distributions $\pi_G(\beta)\, d\beta$ on the nonzero coefficients and i.i.d. Bernoulli kernel inclusion indicators with inclusion probability ρ_G. If $|G|\rho_G\pi_G(\beta) \to \nu(\beta)$ as $|G| \to \infty$, then the result converges to a LARK model on the infinite-dimensional frame. The representation in Wolfe, Godsill and Ng (2004) uses a point mass at zero to provide sparsity. Similarly, one may view the prior distributions in LARK under the ε-truncation approach as assigning zero mass to a neighborhood around zero, also leading to sparse representations. One benefit of LARK is its provision of a formal method for coherent prior specification for continuous dictionaries; a second is its provision of a proper prior specification in the limit as $\varepsilon \to 0$, ensuring insensitivity to the choice of ε.

Standard stochastic search algorithms using finite-dimensional frames may exhibit poor mixing when the correlations between grid elements tend to ± 1. To illustrate, suppose that two possible kernel parameters ω_0 and ω_1 are close in parameter space, leading to two highly correlated columns in the design matrix. In addition, assume that inclusion of either column leads to nearly-maximal likelihood. With the standard one-at-a-time deletion or addition moves in many stochastic search algorithms, to move from a model including a kernel indexed by ω_0 to one indexed by ω_1 would require an extremely unlikely deletion followed by an addition (or unlikely addition followed by a deletion). LARK avoids this difficulty by allowing the continuous parameter ω indexing dictionary elements to move incrementally from ω_0 to ω_1 by a series of update steps, avoiding some of the poor mixing problems associated with highly correlated frame elements in a fine-grid based method.

6.4. *Dirichlet processes*. The Dirichlet process [Ferguson (1973, 1974), Antoniak (1974)] has received widespread use as a prior distribution on probability distribution functions. Its popularity is due in large part to its analytic tractability in many problems; simulation is straightforward, and Bayesian MCMC inference methods are available [Escobar (1994), MacEachern (1994), Escobar and West (1995), MacEachern (1998), Müller and Quintana (2004)]. Liang, Mukherjee and West (2007) consider nonlinear regression and classification models $\mathsf{E}[Y_i \mid X_i] = f(X_i)$ for data $\{(Y_i, X_i)\}$ using kernel expansions of the form

$$f(x) = \int k(x,u)\gamma(du) = \int k(x,u)w(u)F(du) \tag{56}$$

with random signed measure $\gamma(du)$ expressed as the integral of a weight function $w(u)$ with respect to a probability distribution F, modeled as a Dirichlet process $F \sim \mathsf{DP}(F_0, \alpha)$ with base measure F_0 and scale $\alpha > 0$. If observed points $\{X_i\}$ are viewed as a random sample from F, then updating the posterior for F solely on the basis of the observed $\{X_i\}$ would lead in the limit as $\alpha \to 0$ to a degenerate posterior for F concentrated at the empirical distribution for X, justifying the finite-dimensional expansion

$$f(x) = \sum_{i=1}^{n} k(x, x_i)w(x_i)$$

with kernels evaluated only at the observed data locations. The generalized g-prior of West (2003) for the coefficients $\{w_i = w(x_i)\}$ leads to dependent Cauchy distributions for the $\{f(x_i)\}$. This approach (like the SVM, RVM and related approaches) has as many coefficients as there are data points, but avoids over-fitting through shrinkage. Asymptotic properties of $f(x)$ as $n \to \infty$ are difficult to study in the absence of a limiting structure such as that provided by LARK.

The Dirichlet measure $F(du)$ does not assign independent random variables to disjoint sets and so (56) is not a LARK model, but it can be constructed from one. In fact it is exactly the *normalized* LARK model

$$\begin{aligned} f(x) &= \int_\Omega k(x,u)w(u)\mathcal{L}(du)/\mathcal{L}(\Omega) \\ &= \sum_j k(x,u_j)w_j\beta_j/\beta_+ \end{aligned} \tag{57}$$

with $F(du) = \mathcal{L}(du)/\mathcal{L}(\Omega)$ for a Gamma random field $\mathcal{L}(du)$ with infinite Lévy measure

$$\nu(d\beta\, du) = \alpha\beta^{-1}e^{-\beta}\mathbf{1}_{\{\beta>0\}}\, d\beta\, F_0(du),$$

where $\beta_+ := \sum \beta_j$ [note that $w(u)$ could be absorbed into $k(x,u)$].

Well-known disadvantages of Dirichlet process models include their inflexibility (the single parameter α determines the prior dispersion *everywhere*, precluding prior specifications with more uncertainty in some regions than in others), their discreteness, and the limited variability of the masses assigned to the countably-many support points. The normalized Gamma representation (57) of DP's offers the opportunity to overcome some of these disadvantages—for example, the Gamma process may be given a variable rate parameter $b(u)$ by taking

$$\nu(d\beta\, du) = \beta^{-1} e^{-b(u)\beta} \mathbf{1}_{\{\beta>0\}}\, d\beta\, F_0(du)$$

leading to a precision that can vary with location $u \in \Omega$, or the Gamma random field may be replaced with another nonnegative Lévy random field with wider dispersion, such as the fully-skewed Stable process of index $\alpha < 1$. Other nonnegative Lévy random fields are beginning to be used in machine learning [Jordan (2010)] and other fields.

7. Simulation study. We now turn our attention to simulated and real examples to illustrate the performance of LARK models in practice. We conducted a simulation study using four spatially varying functions introduced by Donoho and Johnstone (1994) that are now standard in the wavelet literature: Blocks, Bumps, Doppler and Heavysine. Data were generated for each test function by adding independent Gaussian random noise $\mathsf{No}(0, \sigma^2)$ to the true target function $f(\cdot)$ at $n = 1024$ equally-spaced points on $\mathcal{X} = [0, 10]$. As in Abramovich, Sapatinas and Silverman (1998), the value of σ was chosen to attain a root signal-to-noise ratio (RSNR) of $\sqrt{\int_{\mathcal{X}} (f(x) - \bar{f})^2\, dx / \sigma^2} = 7.0$, where $\bar{f} \equiv \frac{1}{|\mathcal{X}|} \int_{\mathcal{X}} f(x)\, dx$. Each target function $f(\cdot)$ has a range of approximately $0 \le f(x) \le 25$. For each function, we generated 100 replicate data sets to evaluate the performance of LARK and other methods on the basis of mean squared error

$$\text{MSE} \equiv n^{-1} \sum_{i=1}^{n} \bigl(\widehat{f(x_i)} - f(x_i)\bigr)^2. \tag{58}$$

7.1. *Hyperparameters.* In Table 1, we report the kernel functions used for the four simulation examples, chosen to illustrate the flexibility of LARK to use a wide range of kernels that may be adapted to anticipated features (smoothness, spikiness, jumps, curvature, covariation, etc.) of applications. In each case, we take

TABLE 1
Kernel functions used for four test functions

Test function	**Kernel $\phi(x_i; \chi_j, \lambda_j)$**
Blocks	$\mathbf{1}_{\{0<\lambda_j(x_i-\chi_j)\le 1\}}$
Bumps	$e^{-\lambda_j \lvert x_i-\chi_j\rvert}$
Doppler	$e^{-0.5\lambda_j^2 (x_i-\chi_j)^2}$
Heavysine	$e^{-0.5\lambda_j^2 (x_i-\chi_j)^2} \mathbf{1}_{\{\lvert x_i-\chi_j\rvert<2.0\}}$

TABLE 2
Hyperparameters used in examples of Section 7.1

Lévy measure	ε	a_γ	b_γ	a_η	b_η	a_λ	b_λ
Symmetric Gamma	0.0041	2.53	6.45	13.01	0.71	1.117	0.1965
Cauchy	0.0029	2.53	14.2	0.50	1.00	1.117	0.1965

$\Omega = [0, 10] \times \mathbb{R}_+$ (and $|\Omega| = 10$), with elements denoted $\omega = (\chi, \lambda)$, comprising a location parameter $\chi \in \mathcal{X} = [0, 10]$ and a shape parameter $\lambda > 0$. As described in Section 5.2, we take $\{\chi_j\} \overset{\text{i.i.d.}}{\sim} \mathsf{Un}(\Omega)$ and $\{\lambda_j\} \overset{\text{i.i.d.}}{\sim} \mathsf{Ga}(a_\lambda, b_\lambda)$ with a_λ, b_λ chosen (see Table 2) to achieve a 95% prior interval of [0.20, 20.0] for λ to attain dilated kernels covering from half a percent up to fifty percent of $\mathcal{X}$.

Our choice of the remaining hyperparameters was guided by three objectives: to achieve a 95% prior predictive interval of [5, 100] for J, to achieve a 95% prior predictive interval of $[-25, 25]$ for the $\{\beta_j\}$, and to achieve a limit on the mean squared truncation error of $\|\mathcal{L}[\phi] - \mathcal{L}_\varepsilon[\phi]\|_2 = (\mathsf{E}|\mathcal{L}[\phi] - \mathcal{L}_\varepsilon[\phi]|^2)^{1/2} \le 0.05 \cdot \|\phi\|_2$ (see Section 5.2). While these objectives could be met for the LARK model with symmetric Gamma prior with the values given in Table 2, they are not quite attainable for the Cauchy model—the competing goals of an extremely wide distribution for the $\{\beta_j\}$ and a low mean squared truncation error cannot be reconciled. Upon relaxing the prior predictive distribution requirement on $\{\beta_j\}$ to a 99.9% interval of $[-33, 33]$, adequate for this problem with a flat Pareto-tailed distribution for $\{\beta_i\}$, the remaining objectives for the distribution of J and the mean square truncation error were attained using the values given in Table 2. See Figure 5, Appendix C for realizations from the prior distribution.

7.1.1. *Performance*. We compared LARK with two of the best wavelet methods currently available for inhomogeneous function estimation using overcomplete representations: the empirical Bayes approach ("EBayesThresh") of Johnstone and Silverman (2004, 2005a, 2005b) using translational-invariant wavelets, and the continuous over-complete wavelet ("OCW") approach of Chu, Clyde and Liang (2009) based on the stochastic wavelet expansions of Abramovich, Sapatinas and Silverman (2000). We replicated the results of Johnstone and Silverman (2005b) under the beta-Laplace prior using their `R` package `EBayesThresh` [Johnstone and Silverman (2005a)] with Daubechies' "least asymmetric" (`la8`) wavelets [see Section 4 of Daubechies (1988) or Section 6.4 of Daubechies (1992)]. OCW uses the same `la8` wavelet as EBayesThresh except for the Blocks example, where both LARK and OCW use the Haar wavelet. The OCW method may be viewed as a special case of LARK with a finite nonseparable Lévy measure, where coefficients β_j have independent Laplace distributions conditional on scale parameters λ_j, which in turn have truncated Pareto distributions. As in LARK, OCW assigns

TABLE 3
Average and (standard errors) over 100 replications of mean square errors of the four test functions using the Lévy Adaptive Regression Kernels (LARK) using the symmetric Gamma and Cauchy priors, the OCW approach using a Laplace prior [Chu, Clyde and Liang (2009)], and the EBayesThresh approach using a Laplace prior [Johnstone and Silverman (2005a)]

Method	**Blocks**	**Bumps**	**HeavySine**	**Doppler**
LARK-Gamma	0.030 (0.0013)	0.111 (0.0019)	0.038 (0.0010)	0.152 (0.0030)
LARK-Cauchy	0.026 (0.0011)	0.105 (0.0017)	0.036 (0.0010)	0.157 (0.0028)
OCW	0.060 (0.0023)	0.285 (0.0025)	0.082 (0.0010)	0.152 (0.0019)
EBayesThresh	0.096 (0.0013)	0.307 (0.0032)	0.118 (0.00098)	0.202 (0.0027)

independent uniform locations, with a negative binomial distribution for the number of terms in the expansion.

The performance of each method was measured by its average mean square error (AMSE), defined as the average value of the MSE given in (58) over the 100 replicated simulations. Overall, the performance of the LARK model is excellent (Table 3). Both LARK versions generated lower AMSE values than did EBayesThresh for all four test functions. LARK also has smaller AMSE than OCW, except for Doppler, where the methods are comparable. For Blocks, both LARK and OCW use the Haar wavelet, thus any difference in results is due to the prior distribution on the function; LARK leads to a 50% reduction in AMSE compared to OCW. For the other examples, both OCW and EBayesThresh uses a Laplace prior distribution for each coefficient in the expansion and the same wavelet; in all cases it is clear that using a continuous dictionary is better than the finite-dimensional dictionary (frame) with the nondecimated wavelets. Lark reconstructions (right column, Figure 1) consistently show less ringing and fewer artifacts than EBayesThresh (left column).

8. Applications.

8.1. *Motorcycle crash data.* To further illustrate the method, we explore the motorcycle crash experiment data of Schmidt, Mattern and Schüler (1981) considered by Silverman (1985), shown in Figure 2. The 133 observations are unequally spaced, with repeated observations at some time points. Our focus in this example is to illustrate how a single wide class of generating functions may be used in LARK, with the data (through the likelihood) influencing the choice of kernels present in the posterior distribution. We use the power exponential family of kernel functions $\phi(x;\chi,\lambda,\rho) = \exp\{-\lambda|x-\chi|^{\rho}\}$, but here (in contrast with the examples in Section 7) we treat ρ as an uncertain parameter and make inference about it from the data. We take the power ρ to be common for all kernels, and use a relatively concentrated Gamma prior distribution $\rho \sim \mathsf{Ga}(2.0, 0.75)$ with a 50% HPD

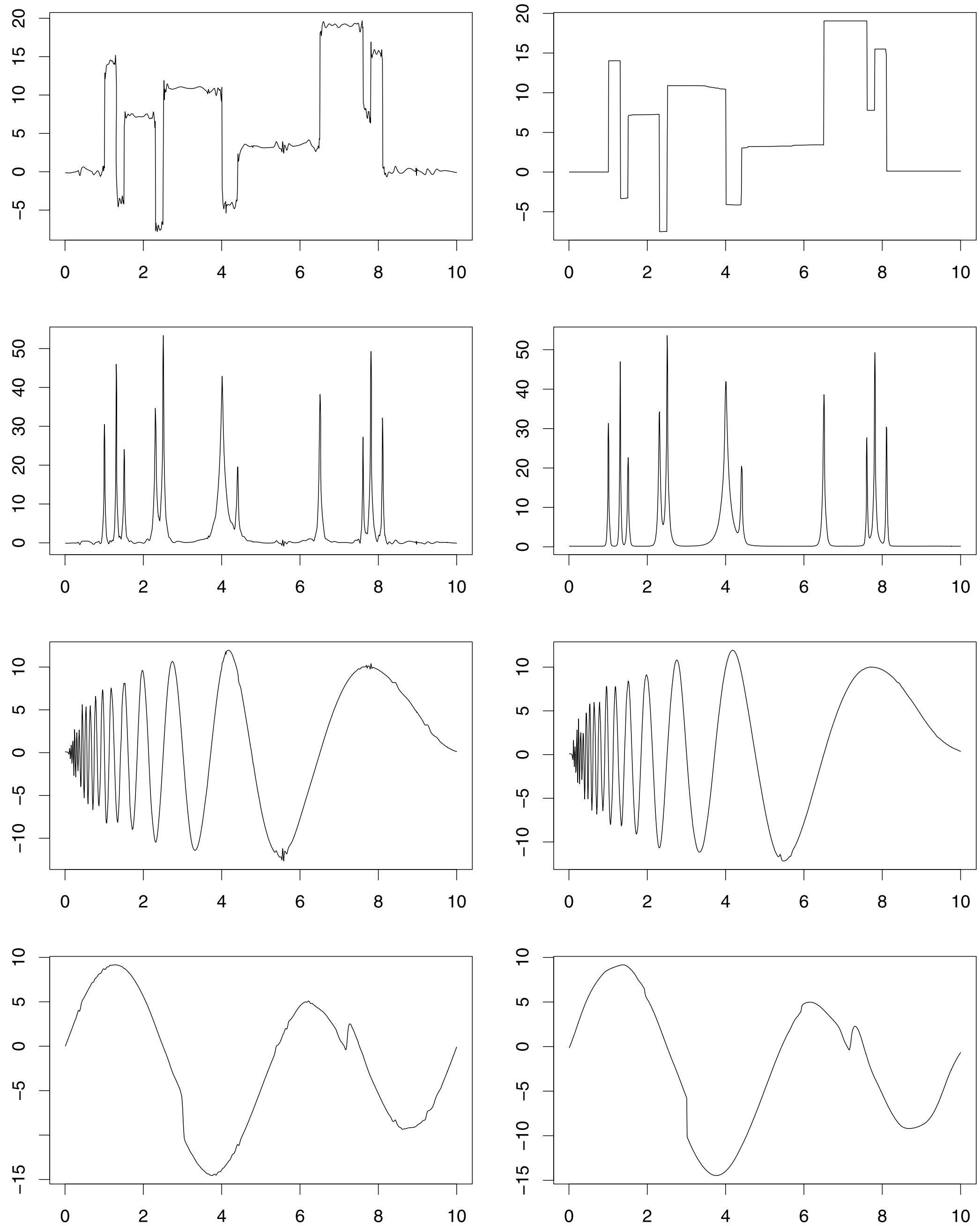


FIG. 1. *Comparison of fitted functions using EBayesThresh* `beta.laplace` [*Johnstone and Silverman* (2005a)] (*left column*) *and Lévy Adaptive Regression Kernels* (*LARK-Gamma*) (*right column*) *for the four test functions. From top to bottom*, *the test functions are Blocks*, *Bumps*, *Doppler and Heavysine*, *respectively.*

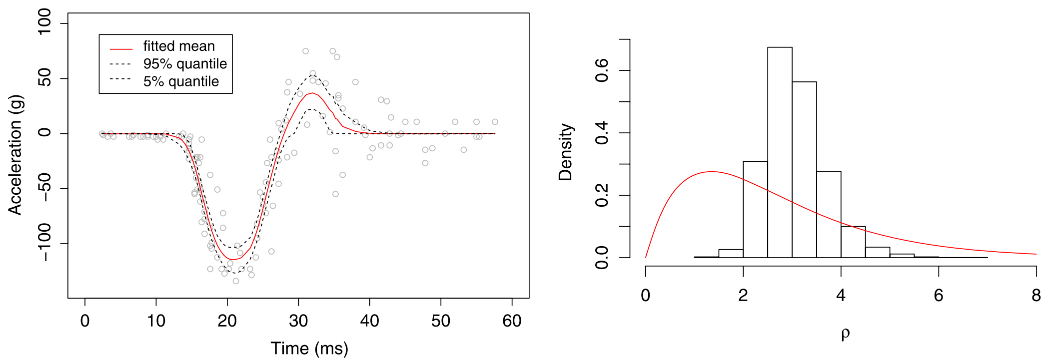


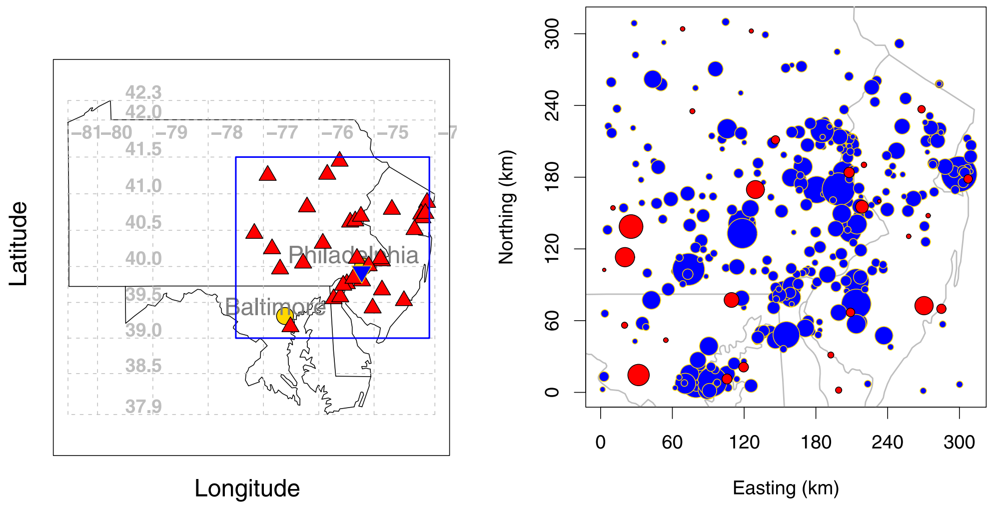
FIG. 2. *Left*: *results of the LARK model for the motorcycle crash data. Circles represent the observations*; *solid line is the posterior mean*; *dotted lines are pointwise* 90% *Bayesian credible interval for the mean function. Right*: *histogram of posterior samples of the exponential power parameter* ρ, *with prior density* (*solid line*) *for comparison*.

interval of $[0.58, 2.56]$ which comfortably includes both the Laplace ($\rho = 1$) and Gaussian ($\rho = 2$) kernels as special cases.

The results are summarized in Figure 2. It is apparent that the fitted mean captures the general trend of the data very well, with minimal boundary effects. The model is parsimonious in the sense we only need 4 kernels on average to fit the data. The posterior mean for ρ is approximately 3 with most of the posterior mass well above the values ($\rho = 1, 2$) for the Laplace and Gaussian kernels.

8.2. *Spatial temporal model.* In this section, we explore the performance of the LARK approach for modeling hourly SO_2 concentration levels (measured in ppm) in Pennsylvania, New Jersey, Delaware and Maryland [U.S. EPA (2007)]. The locations of the 33 monitoring stations are shown in Figure 3; the study region $\mathcal{S}$, delineated by a rectangle in the figure, covers a 310 km $\times$ 310 km area. We used rescaled coordinates from a Lambert (conformal conic) projection to reduce the distortion caused by the earth's curvature. For demonstration purposes, we restrict analysis to measurements taken during a 144 hour period $\mathcal{T}$ from September of 2002. About 5% of SO_2 readings are missing (at random) from the data set, which is not a problem for the LARK model. While Gaussian random field models are popular for modeling spatial-temporal data, the log transformation typically used in the Gaussian approach (because the mean function is strictly positive) eliminates many of the (important) spiky features of the data. Our Gamma random field prior distribution allows us to model the data in the original units.

The model can be written in the same simple form as (52), but now the SO_2 concentration $Y(x)$ is indexed by points $x \in \mathcal{X} = \mathcal{S} \times \mathcal{T}$ in space–time and the Lévy random measure $\mathcal{L}(d\omega)$ assigns Gamma-distributed random variables to Borel sets of a space Ω of points $\omega = (\sigma, \tau, \Lambda, \lambda)$ that include a location $(\sigma, \tau) \in \mathcal{S} \times \mathcal{T}$ in space–time, a positive-definite 2×2 spatial dispersion matrix $\Lambda \in \mathcal{S}_2^+$, and a tem-



(a) (b)

FIG. 3. (a) *Thirty-three monitors used by EPA to measure hourly* SO_2 *concentration in year* 2002. *The inverted triangle denotes Site* 31. *The study area is delineated by a rectangle that includes parts of Pennsylvania*, *Maryland*, *New Jersey and Delaware*, *and is blown up in* (b) *which illustrates locations of kernels from a draw from the posterior distribution*. *The blue circles represent aperiodic points and the red circles represent daily periodic point sources*. *Circle areas are proportional to the magnitudes of the point sources they represent*.

poral decay rate $\lambda > 0$. We employ a separable kernel of the form

$$\phi(x, \omega) = \exp\{-(s-\sigma)'\Lambda(s-\sigma)/2 - \lambda|t-\tau|\}$$

and in the spirit of Higdon [(1998), Section 3.2] and Higdon, Swall and Kern [(1999), Section 2.2], we employ a novel parametrization for Λ in terms of its eigenvalues and the orientation of its major axis [see Tu (2006), Section 4.2.6, for details on prior specifications]. In variations also described in Tu [(2006), Chapter 4] accommodation is made for partial periodicity (due to diurnal patterns associated with daily variation in ambient temperature, traffic levels, etc.), still within the framework described by (51) but now with more elaborate choices for Ω and $\phi(x, \omega)$.

The locations of latent point sources from one iteration of the RJ-MCMC algorithm are presented in Figure 3(b). Larger latent points appear to be clustered in the Baltimore metropolitan area and near the New Jersey/Pennsylvania border. The model's support points are more than a mere modeling device—they can help analysts identify possible underlying sources of pollution, or support future decisions on monitor locations.

The predictive power of the model is validated through out-of-sample prediction. The model was fit excluding data from Site 31 [the inverted triangle in Figure 3(a)], and then its predictions were compared with reported measurements from that site for the entire 144 hours. The result shown in Figure 4 is promising. The major peak was captured clearly, and 90% pointwise Bayesian credible

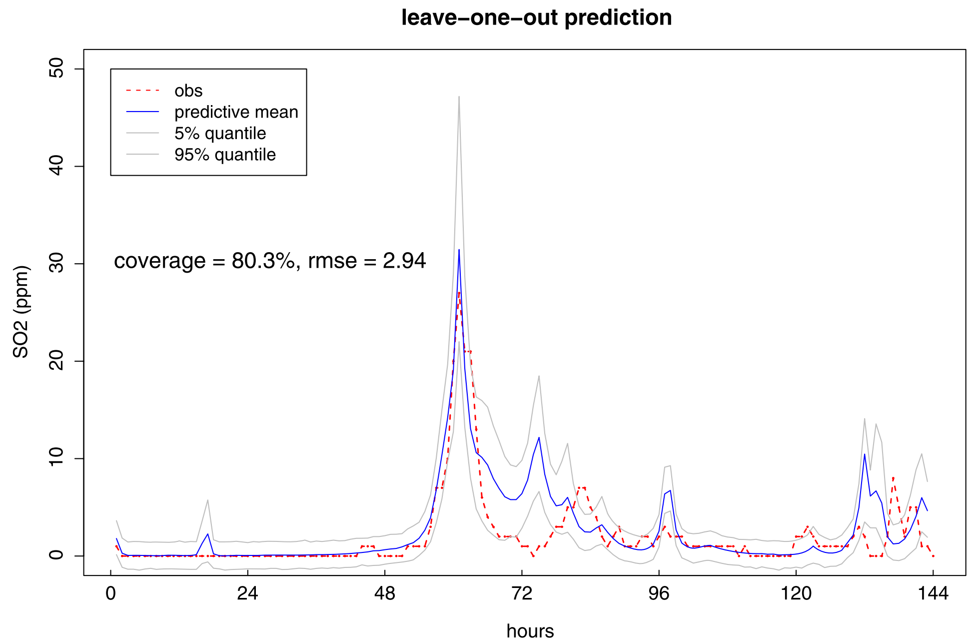


FIG. 4. *Out-of-sample predictions for Site* 31. *Dashed line represents observed time series*, *solid line represents predictive mean curve. Gray lines are* 90% *posterior predictive intervals.*

intervals cover in excess of 80% of the true observations. This was a challenging out-of-sample prediction problem due to low cross-correlations among sites. We are currently refining features of the prior distributions to incorporate known point sources.

9. Discussion. In this article, we have developed a fully Bayesian adaptive kernel method, LARK, for nonparametric function estimation. The LARK model is based on a stochastic expansion of functions in a continuous overcomplete dictionary, and may be expressed as a stochastic integral of a kernel or other generating function with respect to a Lévy random field. When (7) is satisfied (so compensation is unnecessary), the Lévy field is a random signed measure. By using a *positive* random measure and positive kernel family, LARK models provide natural constructions for nonnegative functions (as in Section 8.2); with *signed* measures, unconstrained functions may be modeled (as in Sections 7 and 8.1). The kernel parameters are location-specific and thus adapt to local features of the data. As with wavelets, the adaptive smoothing using LARK preserves local features such as discontinuities and high peaks and is especially useful for modeling inhomogeneous functions. The LARK approach does not require that the data be equally-spaced without missing observations nor that the sample size be a dyadic power as is a commonly required of many wavelet methods.

The RJ-MCMC algorithm developed for fitting LARK provides an automatic stochastic search mechanism for finding sparse representations of a function. The algorithm is computationally efficient [requiring only $O(n \cdot M)$ operations for data including n observations and an MCMC stream of length M], as dictionary elements are calculated only when needed. Kernel methods such as Support Vector Machines (SVMs) and Bayesian Relevance Vector Machines [or RVMs, Tipping (2001)] employ all data points as kernel locations, but attain sparsity by shrinking coefficients to zero. LARK provides additional flexibility by not restricting kernel locations. Many competing sparse methods, including the Dantzig Selector and Lasso, require the a priori selection of a pre-specified number of dictionary elements. Evaluating these kernels on a sufficiently fine grid will exceed the computational cost of LARK. Fine grids also lead to extreme multicollinearity in these approaches, that may lead both to numerical instability and violation of the conditions needed for sparse solutions.

9.1. *Extensions*. It is straightforward to implement LARK with wide classes of generating functions including wavelets, structural elements in texture analysis, and splines. Unlike support vector machines or other methods based on Mercer kernels [Pillai et al. (2007)], the LARK approach does not require symmetry, continuity or simple functional forms. While it is often convenient to use kernels based on some distance metric, arbitrary generating functions may be tailored to the problem at hand as illustrated in the space–time example of Section 8.2. The LARK modeling approach adapts readily to problems in any number of dimensions.

In Section 4, we present conditions for LARK models to belong to the same Besov space as their generating functions, for Lévy measures and generating functions that satisfy the stringent local L_1-bound of (18). In the more general case, where (18) fails and compensation is required, we are able to establish similar results only for $\mathbb{B}^s_{pq}$ with $p = q = 2$ (equivalent to $\mathbb{W}^s_2$). We are exploring extensions to the general case, but the additional drift term that arises in compensation complicates confirming the convergence of f_ε to f in $\mathbb{B}^s_{pq}$ for general p, q.

Work is also on-going in establishing conditions for posterior consistency for function estimation. Extending methods of Choudhuri, Ghosal and Roy (2004), Ghosal and van der Vaart (2007) and Choi and Schervish (2007), Pillai (2008) has verified posterior consistency for certain LARK models with Gaussian measurement errors in work that will be reported elsewhere.

APPENDIX A: DETAILS OF PROOFS

PROPOSITION 2. *For a function $g(\cdot) \in L_p(\mathbb{R}^d)$ and its scaled translate $g(\Lambda(\cdot - \chi))$ with $\chi \in \mathbb{R}^d$ and positive definite matrix $\Lambda \in \mathcal{S}^d_+$, the L_p norm of $g(\Lambda(\cdot - \chi))$ and the L_p norm of its mth forward differences are given by*

$$\|g(\Lambda(\cdot - \chi))\|_p = |\Lambda|^{1/p}\|g\|_p \|\Delta^m_h g(\Lambda(\cdot - \chi))\|_p = |\Lambda|^{1/p}\|\Delta^m_{\lambda h} g\|_p, \tag{59}$$

where $|\Lambda|$ denotes the determinant of Λ.

PROOF. By a change of variables $\chi \mapsto u = \Lambda(x - \chi)$,

$$\begin{aligned}
\|\Delta_h^m g(\Lambda(\cdot - \chi))\|_p &= \left\{\int |\Delta_h^m g(\Lambda(x - \chi))|^p \, dx\right\}^{1/p} \\
&= \left\{\int \left|\sum_{k=0}^{m} \binom{m}{k} (-1)^{m-k} g(\Lambda(x + kh - \chi))\right|^p dx\right\}^{1/p} \\
&= |\Lambda|^{-1/p} \left\{\int \left|\sum_{k=0}^{m} \binom{m}{k} (-1)^{m-k} g(u + k\Lambda h)\right|^p du\right\}^{1/p} \\
&= |\Lambda|^{-1/p} \|\Delta_{\Lambda h}^m g\|_p.
\end{aligned}$$

The proof for the L_p norm of $g(\Lambda(\cdot - \chi))$ follows by the same change of variables. $\square$

A.1. Proof of Lemma 1. First, consider the case $b > 1$ and $a \in \mathbb{R}$. Then

$$\begin{aligned}
&\iiint_{\mathbb{R}\times[1,\infty)\times\mathbb{T}} (1 \wedge |zg(u)^r|\lambda^{-a})\lambda^{-b} \pi_z(dz)\, d\lambda\, du \\
&\quad < \iiint_{\mathbb{R}\times[1,\infty)\times\mathbb{T}} \lambda^{-b} \pi_z(dz)\, d\lambda\, du \\
&\quad = \frac{1}{b-1} < \infty.
\end{aligned}$$

Next, consider the case of $b < 1$ and $a > 1 - b$ (which imply $a > 0$):

$$\begin{aligned}
&\iiint_{\mathbb{R}\times[1,\infty)\times\mathbb{T}} (1 \wedge |zg(u)^r|\lambda^{-a})\lambda^{-b} \pi_z(dz)\, d\lambda\, du \\
&\quad = \iint_{|zg(u)^r|>1} \int_1^{|zg(u)^r|^{1/a}} \lambda^{-b} \, d\lambda \pi_z(dz)\, du \\
&\qquad + \iint_{\mathbb{R}\times\mathbb{T}} |zg(u)^r| \int_{1\vee|zg(u)^r|^{1/a}}^{\infty} \lambda^{-a-b} \, d\lambda \pi_z(dz)\, du \\
&\quad = \iint_{|zg(u)^r|>1} \frac{\lambda^{1-b}}{1-b}\bigg|_{\lambda=1}^{\lambda=|zg(u)^r|^{1/a}} \pi_z(dz)\, du \\
&\qquad + \iint_{\mathbb{R}\times\mathbb{T}} |zg(u)^r| \frac{\lambda^{(1-a-b)}}{1-a-b}\bigg|_{\lambda=1\vee|zg(u)^r|^{1/a}}^{\lambda=\infty} \pi_z(dz)\, du \\
&\quad = \iint_{|zg(u)^r|>1} \frac{1 - |zg(u)^r|^{(1-b)/a}}{b-1} \pi_z(dz)\, du \\
&\qquad + \iint_{|zg(u)^r|>1} \frac{|zg(u)^r|^{(1-b)/a}}{a+b-1} \pi_z(dz)\, du
\end{aligned}$$

$$
\begin{aligned}
&+\iint_{|zg(u)^r|\le 1} \frac{|zg(u)^r|}{a+b-1}\pi_z(dz)\,du\\
&\le \frac{1}{b-1}+\iint_{\mathbb{R}\times\mathbb{T}} |zg(u)^r|^{(1-b)/a}\frac{a}{(a+b-1)(1-b)}\pi_z(dz)\,du\\
&= \frac{1}{b-1}+\frac{a}{(a+b-1)(1-b)}\int_{\mathbb{R}}|z|^{(1-b)/a}\pi_z(dz)\int_{\mathbb{T}}|g(u)^r|^{(1-b)/a}\,du\\
&<\infty
\end{aligned}
$$

for $a+b>1$ if $r=0$, and for $ap+b\ge 1$ if $r=1$ [since $g\in L_p^*(\mathbb{T})$], which is implied by $a>1-b$.

Now consider the case of $b=1$ and $a>0$:

$$
\begin{aligned}
&\iiint_{\mathbb{R}\times[1,\infty)\times\mathbb{T}} (1\wedge|zg(u)^r|\lambda^{-a})\lambda^{-b}\pi_z(dz)\,d\lambda\,du\\
&\quad=\iint_{|zg(u)^r|>1}\int_1^{|zg(u)^r|^{1/a}}\lambda^{-1}\,d\lambda\pi_z(dz)\,du\\
&\qquad+\iint_{\mathbb{R}\times\mathbb{T}}|zg(u)^r|\int_{1\vee|zg(u)^r|^{1/a}}^{\infty}\lambda^{-a-1}\,d\lambda\pi_z(dz)\,du\\
&\quad=\iint_{|zg(u)^r|>1}\log\lambda\Big|_{\lambda=1}^{\lambda=|zg(u)^r|^{1/a}}\pi_z(dz)\,du\\
&\qquad+\iint_{\mathbb{R}\times\mathbb{T}}|zg(u)^r|\frac{\lambda^{-a}}{-a}\Big|_{\lambda=1\vee|zg(u)^r|^{1/a}}^{\lambda=\infty}\pi_z(dz)\,du\\
&\quad=\iint_{|zg(u)^r|>1}\frac{1}{a}\log|zg(u)^r|\pi_z(dz)\,du+\iint_{|zg(u)^r|>1}\frac{1}{a}\pi_z(dz)\,du\\
&\qquad+\iint_{|zg(u)^r|\le 1}\frac{|zg(u)^r|}{a}\pi_z(dz)\,du\\
&\quad\le\frac{1}{a}\iint_{\mathbb{R}\times\mathbb{T}}\log_+|zg(u)^r|\pi_z(dz)\,du+\frac{1}{a}\\
&\quad<\infty
\end{aligned}
$$

since $\log_+(zg^r)=(0\vee\log|zg^r|)\le|z|+|g|^r$ and $g\in L_1^*(\mathbb{T})$.

A.2. Proof of Theorem 2. For any compensator function $h(\beta)$ satisfying (10) there are numbers $c_j\in(0,\infty)$ such that

$$
|h(\beta)|\le c_0,\qquad |\beta-h(\beta)|\le c_1(|\beta|\wedge\beta^2),\qquad |h(\beta)|\le c_2(1\wedge|\beta|)
$$

for all $\beta\in\mathbb{R}$. Fix $0<\varepsilon\le 1$ and a function $\phi:\mathbb{R}\times\Omega\to\mathbb{R}$ satisfying (16); let B_a, B_b and B_c be the values of the integrals from (16a)–(16c), respectively. To complete the proof of Theorem 2 it suffices to show that each of the two terms from (27),

$$
\begin{aligned}
X &\equiv \iint_{N_\varepsilon} \big(\beta - h(\beta)\big)\phi(\omega)\mathcal{N}(d\beta\, d\omega) \quad \text{and} \\
Y &\equiv \iint_{N_\varepsilon} h(\beta)\phi(\omega)\tilde{\mathcal{N}}(d\beta\, d\omega),
\end{aligned}
\tag{60}
$$

converges to zero in probability as $\varepsilon \to 0$. Write the first integral in (60) as the sum of two parts:

$$
X \equiv \iint_{N_\varepsilon} \big(\beta - h(\beta)\big)\phi(\omega)\mathcal{N}(d\beta\, d\omega) = X_1 + X_2
$$

with

$$
\begin{aligned}
X_1 &\equiv \iint_{N_\varepsilon \cap [|\beta\phi| \le 1]} \big(\beta - h(\beta)\big)\phi(\omega)\mathcal{N}(d\beta\, d\omega), \\
X_2 &\equiv \iint_{N_\varepsilon \cap [|\beta\phi| > 1]} \big(\beta - h(\beta)\big)\phi(\omega)\mathcal{N}(d\beta\, d\omega).
\end{aligned}
$$

Then

$$
\begin{aligned}
\mathsf{E}|X_1| &\le c_1 \iint_{N_\varepsilon \cap [|\beta\phi| \le 1]} (|\beta| \wedge \beta^2)|\phi(\omega)|\nu(d\beta\, d\omega) \\
&= c_1 \iint_{N_\varepsilon \cap [|\beta| \le 1] \cap [|\beta\phi| \le 1]} (1 \wedge \beta^2)|\phi(\omega)|\nu(d\beta\, d\omega) \\
&\quad + c_1 \iint_{N_\varepsilon \cap [|\beta| > 1] \cap [|\beta\phi| \le 1]} \big(1 \wedge |\beta\phi(\omega)|\big)\nu(d\beta\, d\omega) \\
&\le c_1(B_c + B_a) < \infty,
\end{aligned}
$$

so $X_1 \to 0$ in L_1 as $\varepsilon \to 0$ by Lebesgue's dominated convergence theorem since the indicator function $\mathbf{1}_{\{N_\varepsilon\}}(\beta, \omega)$ tends to zero a.e. (ν) as $\varepsilon \to 0$. Now consider X_2:

$$
\begin{aligned}
\nu\big(\{(\beta, \omega) : |\beta\phi(\omega)| > 1\}\big) &= \iint_{[|\beta| \le 1] \cap [|\beta\phi| > 1]} 1\nu(d\beta\, d\omega) \\
&\quad + \iint_{[|\beta| > 1] \cap [|\beta\phi| > 1]} 1\nu(d\beta\, d\omega) \\
&\le \iint_{[|\beta| \le 1] \cap [|\beta\phi| > 1]} \big(|\beta\phi(\omega)| \wedge |\beta\phi(\omega)|^2\big)\nu(d\beta\, d\omega) \\
&\quad + \iint_{[|\beta| > 1] \cap [|\beta\phi| > 1]} \big(1 \wedge |\beta\phi(\omega)|\big)\nu(d\beta\, d\omega) \\
&\le B_b + B_a < \infty,
\end{aligned}
$$

so almost surely the random support of $\mathcal{N}(d\beta\,d\omega)$ in $[|\beta\phi| > 1]$ is a finite set disjoint from $\bigcap_{\varepsilon>0} N_\varepsilon$; it follows that $\mathcal{N}(N_\varepsilon \cap [|\beta\phi(\omega)| > 1]) \to 0$ and hence $X_2 \to 0$ almost surely as $\varepsilon \to 0$.

Similarly, we write the second integral in (60) as the sum of four parts:

$$Y \equiv \iint_{N_\varepsilon} h(\beta)\phi(\omega)\tilde{\mathcal{N}}(d\beta\,d\omega) = Y_1 + Y_2 + Y_3 + Y_4$$

with

$$\begin{aligned}
Y_1 &\equiv \iint_{N_\varepsilon\cap[|\beta|\le 1]\cap[|\beta\phi|\le 1]} h(\beta)\phi(\omega)\tilde{\mathcal{N}}(d\beta\,d\omega),\\
Y_2 &\equiv \iint_{N_\varepsilon\cap[|\beta|\le 1]\cap[|\beta\phi|> 1]} h(\beta)\phi(\omega)\tilde{\mathcal{N}}(d\beta\,d\omega),\\
Y_3 &\equiv \iint_{N_\varepsilon\cap[|\beta|> 1]\cap[|\beta\phi|\le 1]} h(\beta)\phi(\omega)\tilde{\mathcal{N}}(d\beta\,d\omega),\\
Y_4 &\equiv \iint_{N_\varepsilon\cap[|\beta|> 1]\cap[|\beta\phi|> 1]} h(\beta)\phi(\omega)\tilde{\mathcal{N}}(d\beta\,d\omega).
\end{aligned}$$

Now

$$\begin{aligned}
\mathsf{E}|Y_1|^2 &= \iint_{N_\varepsilon\cap[|\beta|\le 1]\cap[|\beta\phi|\le 1]} h(\beta)^2\phi(\omega)^2\nu(d\beta\,d\omega)\\
&\le c_2^2 \iint_{N_\varepsilon\cap[|\beta|\le 1]\cap[|\beta\phi|\le 1]} |\beta\phi(\omega)|^2\nu(d\beta\,d\omega)\\
&= c_2^2 \iint_{N_\varepsilon\cap[|\beta|\le 1]\cap[|\beta\phi|\le 1]} \bigl(|\beta\phi(\omega)| \wedge |\beta\phi(\omega)|^2\bigr)\nu(d\beta\,d\omega)\\
&\le c_2^2 B_b < \infty,
\end{aligned}$$

so $Y_1 \to 0$ in L_2 (and hence also in L_1) as $\varepsilon \to 0$ by LDCT,

$$\begin{aligned}
Y_2 &\equiv \iint_{N_\varepsilon\cap[|\beta|\le 1]\cap[|\beta\phi|> 1]} h(\beta)\phi(\omega)\tilde{\mathcal{N}}(d\beta\,d\omega)\\
&= \iint_{N_\varepsilon\cap[|\beta|\le 1]\cap[|\beta\phi|> 1]} h(\beta)\phi(\omega)\mathcal{N}(d\beta\,d\omega)\\
&\quad - \iint_{N_\varepsilon\cap[|\beta|\le 1]\cap[|\beta\phi|> 1]} h(\beta)\phi(\omega)\nu(d\beta\,d\omega),\\
\mathsf{E}|Y_2| &\le 2\iint_{N_\varepsilon\cap[|\beta|\le 1]\cap[|\beta\phi|> 1]} |h(\beta)||\phi(\omega)|\nu(d\beta\,d\omega)\\
&\le 2c_2 \iint_{N_\varepsilon\cap[|\beta|\le 1]\cap[|\beta\phi|> 1]} |\beta\phi(\omega)|\nu(d\beta\,d\omega)\\
&= 2c_2 \iint_{N_\varepsilon\cap[|\beta|\le 1]\cap[|\beta\phi|> 1]} \bigl(|\beta\phi(\omega)| \wedge |\beta\phi(\omega)|^2\bigr)\nu(d\beta\,d\omega)\\
&\le 2c_2 B_b < \infty,
\end{aligned}$$

so $Y_2 \to 0$ in L_1 as $\varepsilon \to 0$ by dominated convergence,

$$\begin{aligned}
Y_3 &\equiv \iint_{N_\varepsilon \cap [|\beta|>1] \cap [|\beta\phi| \le 1]} h(\beta)\phi(\omega)\tilde{\mathcal{N}}(d\beta\, d\omega) \\
&= \iint_{N_\varepsilon \cap [|\beta|>1] \cap [|\beta\phi| \le 1]} h(\beta)\phi(\omega)\mathcal{N}(d\beta\, d\omega) \\
&\quad - \iint_{N_\varepsilon \cap [|\beta|>1] \cap [|\beta\phi| \le 1]} h(\beta)\phi(\omega)\nu(d\beta\, d\omega), \\
\mathsf{E}|Y_3| &\le 2 \iint_{N_\varepsilon \cap [|\beta|>1] \cap [|\beta\phi| \le 1]} |h(\beta)||\phi(\omega)|\nu(d\beta\, d\omega) \\
&\le 2c_2 \iint_{N_\varepsilon \cap [|\beta|>1] \cap [|\beta\phi| \le 1]} |\beta\phi(\omega)|\nu(d\beta\, d\omega) \\
&= 2c_2 \iint_{N_\varepsilon \cap [|\beta|>1] \cap [|\beta\phi| \le >1]} \bigl(1 \wedge |\beta\phi(\omega)|\bigr)\nu(d\beta\, d\omega) \\
&\le 2c_2 B_a < \infty,
\end{aligned}$$

so $Y_3 \to 0$ in L_1 as $\varepsilon \to 0$. Finally, for Y_4,

$$\begin{aligned}
Y_4 &\equiv \iint_{N_\varepsilon \cap [|\beta|>1] \cap [|\beta\phi| > 1]} h(\beta)\phi(\omega)\tilde{\mathcal{N}}(d\beta\, d\omega) \\
&= \iint_{N_\varepsilon \cap [|\beta|>1] \cap [|\beta\phi| > 1]} h(\beta)\phi(\omega)\mathcal{N}(d\beta\, d\omega) \\
&\quad - \iint_{N_\varepsilon \cap [|\beta|>1] \cap [|\beta\phi| > 1]} h(\beta)\phi(\omega)\nu(d\beta\, d\omega), \\
\mathsf{E}|Y_4| &\le 2 \iint_{N_\varepsilon \cap [|\beta|>1] \cap [|\beta\phi| > 1]} |h(\beta)\phi(\omega)|\nu(d\beta\, d\omega) \\
&\le 2c_0 \iint_{N_\varepsilon \cap [|\beta|>1] \cap [|\beta\phi| > 1]} |\phi(\omega)|\nu(d\beta\, d\omega) \\
&\le 2c_0 \iint_{N_\varepsilon \cap [|\beta|>1] \cap [|\beta\phi| > 1]} (1 \wedge \beta^2)|\phi(\omega)|\nu(d\beta\, d\omega) \\
&\le 2c_0 B_c < \infty
\end{aligned}$$

so $Y_4 \to 0$ in L_1 as $\varepsilon \to 0$, completing the proof of Theorem 2.

APPENDIX B: REVERSIBLE-JUMP MCMC PROCEDURES

A typical RJ-MCMC procedure for sampling varying-dimensional parameters involves at least three types of moves (Birth, Death and Update); we use Metropolis–Hastings steps for each of these. Our trans-dimensional update steps entail altering the value (β_j^*, ω_j^*) of one point (β_j, ω_j). We select $j \sim \mathsf{Un}(0 : J-1)$

for proposed updating, then take Gaussian random walk steps successively in the coefficient β_j, the location parameter χ_j, and the log kernel shape parameter, $\log \lambda_j$. Step sizes are chosen to achieve approximately 30% acceptance rates for each class of updates. One novel feature is that when the proposed update of some coefficient β_j falls in the truncated region $\beta_j^* \eta \in (-\varepsilon, \varepsilon)$, the move is treated as a Death, the point (β_j, ω_j) is removed and J is decremented. This is advantageous as it automatically focuses on small magnitude coefficients for removal (rather than a random selection as in the typical RJ-MCMC Death step). A Birth step entails generating a new point (β^*, ω^*) to be included among the $\{(\beta_j, \omega_j)\}$ and incrementing J by one. We use a double exponential birth distribution with rate η/ε, conditioned to exceed $|\beta_j|\eta > \varepsilon$ so that proposed coefficients are small, balancing the "Death" of small coefficients in the Update step to attain the target acceptance rates. The fixed-dimensional parameters are sampled using a conventional Metropolis–Hastings approach [Gilks, Richardson and Spiegelhalter (1996), Section 1.3.3]. Each of these inexpensive update steps requires only $O(n)$ operations [in contrast to Gaussian methods, which may require $O(n^3)$], so the method scales well in the number n of observations. Further details of the RJ-MCMC are available in [Tu (2006), Appendix A.1, pages 116 and 117]. An R package [R Development Core Team (2004)] implementing LARK is under development by the authors and will be made publicly available.

APPENDIX C: EXAMPLES OF LARK PRIOR REALIZATIONS

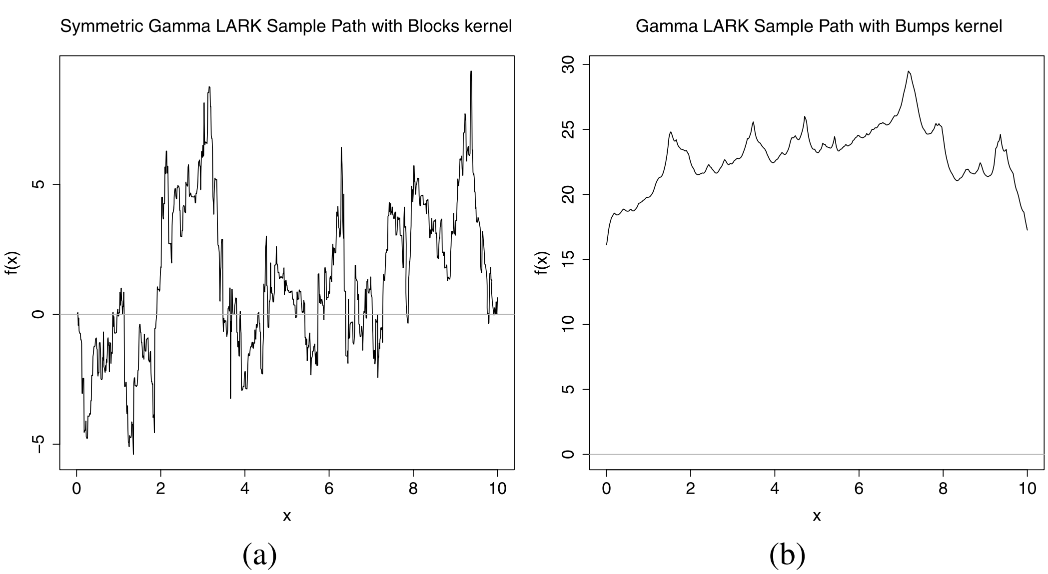
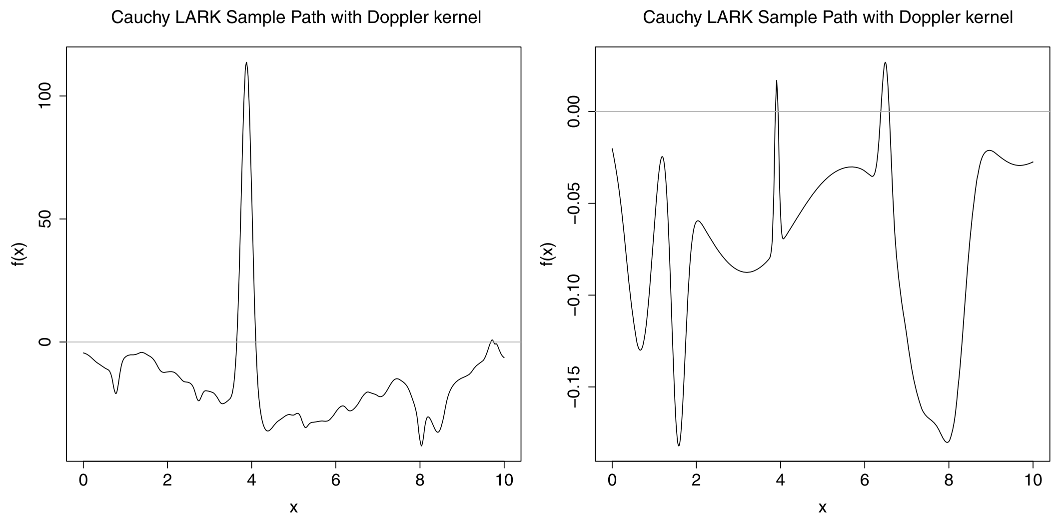


FIG. 5. *Four realizations from LARK prior distribution with* (a) *Blocks kernel and Symmetric Gamma Lévy measure*; (b) *Bumps kernel and Gamma Lévy measure*; (c), (d) *Doppler kernel and Cauchy Lévy measure*, *with* $J = 1000$ *for* (a)–(c) *and* $J = 10$ *for* (d) *components*. *Hyperparameters* $a_\lambda, b_\lambda, a_\gamma, b_\gamma, a_\eta, b_\eta$ *and* ε *are given in Table* 2.



(c) (d)

FIG. 5. (*Continued*.)

Acknowledgments. The authors would like to thank Natesh Pillai, three referees, the Associate Editor and the Editor for helpful comments and suggestions.

REFERENCES

ABRAMOVICH, F., SAPATINAS, T. and SILVERMAN, B. W. (1998). Wavelet thresholding via a Bayesian approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60** 725–749. MR1649547

ABRAMOVICH, F., SAPATINAS, T. and SILVERMAN, B. W. (2000). Stochastic expansions in an overcomplete wavelet dictionary. *Probab. Theory Related Fields* **117** 133–144. MR1759511

ABRAMOWITZ, M. and STEGUN, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. National Bureau of Standards Applied Mathematics Series* **55**. U.S. Government Printing Office, Washington, DC. MR0167642

ANTONIAK, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* **2** 1152–1174. MR0365969

CANDÈS, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *Ann. Statist.* **35** 2313–2351. MR2382644

CHEN, S. S., DONOHO, D. L. and SAUNDERS, M. A. (1998). Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **20** 33–61. MR1639094

CHILÈS, J.-P. and DELFINER, P. (1999). *Geostatistics*: *Modeling Spatial Uncertainty*. Wiley, New York. MR1679557

CHOI, T. and SCHERVISH, M. J. (2007). On posterior consistency in nonparametric regression problems. *J. Multivariate Anal.* **98** 1969–1987. MR2396949

CHOUDHURI, N., GHOSAL, S. and ROY, A. (2004). Bayesian estimation of the spectral density of a time series. *J. Amer. Statist. Assoc.* **99** 1050–1059. MR2109494

CHU, J.-H., CLYDE, M. A. and LIANG, F. (2009). Bayesian function estimation using continuous wavelet dictionaries. *Statist. Sinica* **19** 1419–1438. MR2589190

CHU, C.-K. and MARRON, J. S. (1991). Choosing a kernel regression estimator (with discussion). *Statist. Sci.* **6** 404–436. MR1146907

CLYDE, M. A. and WOLPERT, R. L. (2007). Nonparametric function estimation using overcomplete dictionaries. In *Bayesian Statistics* 8 (J. M. Bernardo, M. J. Bayarri, J. O. Berger,

A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West, eds.) 91–114. Oxford Univ. Press, Oxford. MR2433190

CONT, R. and TANKOV, P. (2004). *Financial Modelling with Jump Processes*. Chapman & Hall/CRC, Boca Raton, FL. MR2042661

CRISTIANINI, N. and SHAWE-TAYLOR, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge Univ. Press, Cambridge.

DAUBECHIES, I. (1988). Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.* **41** 909–996. MR0951745

DAUBECHIES, I. (1992). *Ten Lectures on Wavelets. CBMS-NSF Regional Conference Series in Applied Mathematics* **61**. SIAM, Philadelphia, PA. MR1162107

DENISON, D. G. T., MALLICK, B. K. and SMITH, A. F. M. (1998). Automatic Bayesian curve fitting. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60** 333–350. MR1616029

DENISON, D. G. T., HOLMES, C. C., MALLICK, B. K. and SMITH, A. F. M. (2002). *Bayesian Methods for Nonlinear Classification and Regression*. Wiley, Chichester. MR1962778

DIMATTEO, I., GENOVESE, C. R. and KASS, R. E. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika* **88** 1055–1071. MR1872219

DONOHO, D. L. and ELAD, M. (2003). Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization. *Proc. Natl. Acad. Sci. USA* **100** 2197–2202 (electronic). MR1963681

DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455. MR1311089

ESCOBAR, M. D. (1994). Estimating normal means with a Dirichlet process prior. *J. Amer. Statist. Assoc.* **89** 268–277. MR1266299

ESCOBAR, M. D. and WEST, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90** 577–588. MR1340510

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230. MR0350949

FERGUSON, T. S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2** 615–629. MR0438568

GHOSAL, S. and VAN DER VAART, A. (2007). Convergence rates of posterior distributions for non-i.i.d. observations. *Ann. Statist.* **35** 192–223. MR2332274

GILKS, W. R., RICHARDSON, S. and SPIEGELHALTER, D. J., eds. (1996). *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London. MR1397966

GREEN, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82** 711–732. MR1380810

HIGDON, D. M. (1998). A process-convolution approach to modeling temperatures in the North Atlantic ocean. *Environ. Ecol. Stat.* **5** 173–190.

HIGDON, D., SWALL, J. and KERN, J. (1999). Non-stationary spatial modeling. In *Bayesian Statistics* 6 (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) 761–768. Oxford Univ. Press, Oxford.

JACOD, J. and SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **288**. Springer, Berlin. MR0959133

JOHNSTONE, I. M. and SILVERMAN, B. W. (2004). Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences. *Ann. Statist.* **32** 1594–1649. MR2089135

JOHNSTONE, I. M. and SILVERMAN, B. W. (2005a). EBayesThresh: R programs for empirical Bayes thresholding. *Journal of Statistical Software* **12** 1–38.

JOHNSTONE, I. M. and SILVERMAN, B. W. (2005b). Empirical Bayes selection of wavelet thresholds. *Ann. Statist.* **33** 1700–1752. MR2166560

JORDAN, M. I. (2010). Hierarchical models, nested models and completely random measures. In *Frontiers of Statistical Decision Making and Bayesian Analysis*: *In Honor of James O. Berger* (M.-H. Chen, D. K. Dey, P. Müller, D. Sun and K. Ye, eds.) 207–217. Springer, New York. MRMR2766461

KHINCHINE, A. Y. and LÉVY, P. (1936). Sur les lois stables. *C. R. Math. Acad. Sci. Paris* **202** 374–376.

KINGMAN, J. F. C. (1967). Completely random measures. *Pacific J. Math*. **21** 59–78. MR0210185

KWAPIEŃ, S. and WOYCZYŃSKI, W. A. (1992). *Random Series and Stochastic Integrals*: *Single and Multiple*. Birkhäuser, Boston, MA. MR1167198

LAW, M. H. and KWOK, J. T. (2001). Bayesian support vector regression. In *Proceedings of the Eighth International Workshop on Artificial Intelligence and Statistics* (*AISTATS*) 239–244. Key West, FL.

LIANG, F., MUKHERJEE, S. and WEST, M. (2007). The use of unlabeled data in predictive modeling. *Statist. Sci*. **22** 189–205. MR2408958

MACEACHERN, S. N. (1994). Estimating normal means with a conjugate style Dirichlet process prior. *Comm. Statist. Simulation Comput*. **23** 727–741. MR1293996

MACEACHERN, S. N. (1998). Computational methods for mixture of Dirichlet process models. In *Practical Nonparametric and Semiparametric Bayesian Statistics* (D. K. Dey, P. Müller and D. Sinha, eds.). *Lecture Notes in Statist*. **133** 23–43. Springer, New York. MR1630074

MALLAT, S. G. and ZHANG, Z. (1993). Matching pursuit with time-frequency dictionaries. *IEEE Trans. Signal Process* **41** 3397–3415.

MÜLLER, P. and QUINTANA, F. A. (2004). Nonparametric Bayesian data analysis. *Statist. Sci*. **19** 95–110. MR2082149

NIKOL'SKIĬ, S. M. (1975). *Approximation of Functions of Several Variables and Imbedding Theorems*. *Die Grundlehren der Mathematischen Wissenschaften* **205** Springer, New York. Translated from the Russian by John M. Danskin, Jr. MR0374877

PILLAI, N. S. (2008). Lévy random measures: Posterior consistency and applications. Ph.D. dissertation, Dept. Statist. Sci., Duke Univ. Available at http://stat.duke.edu/people/theses/PillaiNS.html.

PILLAI, N. S., WU, Q., LIANG, F., MUKHERJEE, S. and WOLPERT, R. L. (2007). Characterizing the function space for Bayesian kernel models. *J. Mach. Learn. Res*. **8** 1769–1797 (electronic). MR2332448

R DEVELOPMENT CORE TEAM (2004). R: A language and environment for statistical computing. R foundation for statistical computing. Available at http://www.R-project.org.

RAJPUT, B. S. and ROSIŃSKI, J. (1989). Spectral representations of infinitely divisible processes. *Probab. Theory Related Fields* **82** 451–487. MR1001524

REED, M. C. and SIMON, B. (1975). *Methods of Modern Mathematical Physics*, *Vol. II*: *Fourier Analysis*, *Self-Adjointness*. Academic Press, New York.

SATO, K.-I. (1999). *Lévy Processes and Infinitely Divisible Distributions*. *Cambridge Studies in Advanced Mathematics* **68**. Cambridge Univ. Press, Cambridge. Translated from the 1990 Japanese original. Revised by the author. MR1739520

SCHMIDT, G., MATTERN, R. and SCHÜLER, F. (1981). Biomechanical investigation to determine physical and traumatological differentiation criteria for the maximum load capacity of head and vertebral column with and without protective helmet under the effects of impact. EEC research program on biomechanics of impacts, final report, phase III, Project 65, Institut für Rechtsmedizin, Univ. Heidelberg, Germany.

SILVERMAN, B. W. (1985). Some aspects of the spline smoothing approach to nonparametric regression curve fitting. *J. R. Stat. Soc. Ser. B Stat. Methodol*. **47** 1–52. MR0805063

SISSON, S. A. (2005). Transdimensional Markov chains: A decade of progress and future perspectives. *J. Amer. Statist. Assoc*. **100** 1077–1089. MR2201033

SMITH, M. and KOHN, R. (1996). Nonparametric regression using Bayesian variable selection. *J. Econometrics* **75** 317–343.

SOBOLEV, S. L. (1991). *Some Applications of Functional Analysis in Mathematical Physics*. *Translations of Mathematical Monographs* **90**. Amer. Math. Soc., Providence, RI. MR1125990

SOLLICH, P. (2002). Bayesian methods for support vector machines: Evidence and predictive class probabilities. *Machine Learning* **46** 21–52.

STEIN, M. L. (1999). *Interpolation of Spatial Data*: *Some Theory for Kriging*. Springer, New York. MR1697409

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol*. **58** 267–288. MR1379242

TIPPING, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res*. **1** 211–244. MR1875838

TRIEBEL, H. (1992). *Theory of Function Spaces*. *II*. *Monographs in Mathematics* **84**. Birkhäuser, Basel. MR1163193

TU, C. (2006). Nonparametric modelling using Lévy process priors with applications for function estimation, time series modeling and spatio-temporal modeling. Ph.D. dissertation, Dept. Statist. Sci., Duke Univ. Available at http://www.stat.duke.edu/people/theses/TuC.html.

U.S. EPA. (2007). Air Quality System (AQS). Available at http://www.epa.gov/ttn/airs/airsaqs/.

VIDAKOVIC, B. (1999). *Statistical Modeling by Wavelets*. Wiley, New York. MR1681904

WAHBA, G. (1992). Multivariate function and operator estimation, based on smoothing splines and reproducing kernels. In *Nonlinear Modeling and Forecasting*: *Proceedings of the Workshop on Nonlinear Modeling and Forecasting held September*, 1990, *in Santa Fe*, *New Mexico* (M. Casdagli and S. G. Eubank, eds.). *SFI Studies in the Sciences of Complexity* **XII** 95–112. Addison-Wesley, Redwood, CA.

WEST, M. (2003). Bayesian factor regression models in the "large p, small n" paradigm. In *Bayesian Statistics* 7 (J. M. Bernardo et al., eds.) 733–742. Oxford Univ. Press, New York. MR2003537

WOLFE, P. J., GODSILL, S. J. and NG, W.-J. (2004). Bayesian variable selection and regularization for time-frequency surface estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol*. **66** 575–589. MR2088291

WOLPERT, R. L., ICKSTADT, K. and HANSEN, M. B. (2003). A nonparametric Bayesian approach to inverse problems. In *Bayesian Statistics* 7 (J. M. Bernardo et al., eds.) 403–417. Oxford Univ. Press, New York. MR2003186

WOLPERT, R. L. and TAQQU, M. S. (2005). Fractional Ornstein–Uhlenbeck Lévy processes and the Telecom process: Upstairs and downstairs. *Signal Processing* **85** 1523–1545.

ZOLOTAREV, V. M. (1986). *One-dimensional Stable Distributions*. *Translations of Mathematical Monographs* **65**. Amer. Math. Soc., Providence, RI. MR0854867

R. L. WOLPERT
M. A. CLYDE
DEPARTMENT OF STATISTICAL SCIENCE
DUKE UNIVERSITY
DURHAM, NORTH CAROLINA 27708-0251
USA
E-MAIL: wolpert@stat.duke.edu
clyde@stat.duke.edu
URL: http://www.stat.duke.edu

C. TU
PIMCO
1345 AVENUE OF THE AMERICAS
NEW YORK, NEW YORK 10105-4800
USA
E-MAIL: Chong.Tu@pimco.com