Journal of Multivariate Analysis 123 (2014) 172–183

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Parameter estimation for operator scaling random fields

C.Y. Lim [a], M.M. Meerschaert [a,*], H.-P. Scheffler [b]

[a] Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, United States
[b] Fachbereich Mathematik, Universität Siegen, 57068 Siegen, Germany

ARTICLE INFO

Article history:
Received 11 September 2012
Available online 27 September 2013

AMS subject classifications:
primary 62F12
secondary 62H11
60G60

Keywords:
Random field
Self-similar
Operator scaling
Hurst index

ABSTRACT

Operator scaling random fields are useful for modeling physical phenomena with different scaling properties in each coordinate. This paper develops a general parameter estimation method for such fields which allows an arbitrary set of scaling axes. The method is based on a new approach to nonlinear regression with errors whose mean is not zero.

1. Introduction

Random fields are useful models for many natural phenomena (e.g., see Adler [1]). Self-similar random fields capture the fractal properties observed in applications (e.g., see Embrechts and Maejima [9]). An application to ground water hydrology laid out in Benson et al. [3] notes that the Hurst index of self-similarity can be expected to vary with the coordinate. In a two-dimensional model of an alluvial aquifer, a Hurst index $H_1 \geq 0.5$ models the organization of a porous medium in the natural direction of ground water flow, and another Hurst index $H_2 < 0.5$ describes negative dependence in the vertical direction, which captures the layering effect of the fluvial deposition process that created the medium structure. The scaling axes of the model often differ from the usual spatial coordinates. For example, there is often a dipping angle that tilts the first coordinate downward. In applications to fracture flow, a set of non-orthogonal scaling axes represents fracture orientations (e.g., see Ponson et al. [18] or Reeves et al. [19]).

To address such applications, Biermé et al. [6] developed a mathematical theory of operator scaling stable random fields (OSSRFs), based on ideas from [3]. An OSSRF is a scalar-valued random field $\{B(\boldsymbol{x})\}_{\boldsymbol{x}\in\mathbb{R}^d}$ such that

$$\{B(c^E\boldsymbol{x})\}_{\boldsymbol{x}\in\mathbb{R}^d} \triangleq \{cB(\boldsymbol{x})\}_{\boldsymbol{x}\in\mathbb{R}^d} \quad \text{for all } c > 0, \tag{1.1}$$

where E is a $d \times d$ scaling matrix whose eigenvalues have real part greater than zero, $c^E = \exp(E \log c)$, with $\exp(A) = I + A + A^2/2! + \cdots$ the usual matrix exponential, and $\triangleq$ denotes equality of all finite-dimensional distributions. If the scaling matrix E has a basis of eigenvectors $E\,\boldsymbol{b}_i = a_i\boldsymbol{b}_i$ for $i = 1, \ldots, d$, then $c^E\boldsymbol{b}_i = c^{a_i}\boldsymbol{b}_i$ for $i = 1, \ldots, d$, and it follows immediately from (1.1) that the one-dimensional slice $B_i(t) := B(t\boldsymbol{b}_i)$ is self-similar with Hurst index $H_i = 1/a_i$, i.e.,

$$\{B_i(ct)\}_{t\in\mathbb{R}} \triangleq \{c^{H_i}B_i(t)\}_{t\in\mathbb{R}} \quad \text{for all } c > 0.$$

* Corresponding author.
E-mail addresses: lim@stt.msu.edu (C.Y. Lim), mcubed@stt.msu.edu (M.M. Meerschaert), scheffler@mathematik.uni-siegen.de (H.-P. Scheffler).
URLs: http://www.stt.msu.edu/users/lim/ (C.Y. Lim), http://www.stt.msu.edu/users/mcubed/ (M.M. Meerschaert), http://www.uni-siegen.de/fb6/src/scheffler/ (H.-P. Scheffler).

0047-259X/$ – see front matter
http://dx.doi.org/10.1016/j.jmva.2013.09.010

In particular, the Hurst index H_i of self-similarity varies with the coordinate, and the scaling axes $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d$ can be any basis for $\mathbb{R}^d$. The construction of the OSSRF in [6] ensures that the random field has stationary increments, i.e.,

$$\{B(\boldsymbol{x}+\boldsymbol{h}) - B(\boldsymbol{h})\}_{\boldsymbol{x}\in\mathbb{R}^d} \triangleq \{B(\boldsymbol{x})\}_{\boldsymbol{x}\in\mathbb{R}^d} \quad \text{for all } \boldsymbol{h} \in \mathbb{R}^d,$$

and then it follows that any one-dimensional slice $B_{\boldsymbol{x},i}(t) := B(\boldsymbol{x} + t\boldsymbol{b}_i) - B(\boldsymbol{x})$ is self-similar with Hurst index H_i. If the random field is Gaussian, then $B_i(t) := B(t\boldsymbol{b}_i)$ is a fractional Brownian motion with Hurst index $H_i = 1/a_i$, since this is the only self-similar Gaussian process with stationary increments [20, Corollary 7.2.3]. OSSRFs were applied to ground water hydrology by Hu et al. [12] to synthesize realistic porosity fields and hydraulic conductivity fields, consistent with aquifer data. The multi-scaling produces organized regions of high porosity (and/or conductivity) that create preferential flow paths, an important feature of realistic random field simulations that is not present in an isotropic model.

Practical applications of multi-scaling random fields require a method to estimate the parameters. For the special case where the scaling axes equal the original Euclidean coordinates, estimation methods have been developed by Beran et al. [4], Boissy et al. [7], and Guo et al. [10]. However, applications to geophysics require a more general approach, with an arbitrary set of scaling axes. This paper develops a general method for parameter estimation, which also estimates the appropriate scaling axes. These axes need not be orthogonal. Our approach is based on a new method for nonlinear regression with errors whose mean is not zero. This method for nonlinear regression may well have further applications in other areas.

In Section 2, we review OSSRFs and outline the parameter estimation problem, which involves a nonlinear regression where the errors do not have a zero mean. In Section 3, we propose a new nonlinear regression method to handle the nonzero mean error, and prove consistency and asymptotic normality for this estimator. In Section 4, we return to OSSRFs and apply the proposed nonlinear regression method to estimate parameters. Section 5 summarizes the results of a brief simulation study, to verify that the method gives reasonably accurate parameter estimates in practice. Some concluding remarks are contained in Section 6.

2. Operator scaling random fields

In this section, we recall the spectral method for constructing OSSRFs; see Biermé et al. [6] for complete details. Then we outline the proposed nonlinear regression method for parameter estimation.

Given a $d \times d$ scaling matrix E whose eigenvalues all have positive real part, we say that a continuous function $\psi : \mathbb{R}^d \to [0, \infty)$ is E^T-homogeneous if $\psi(c^{E^T}\boldsymbol{\xi}) = c \cdot \psi(\boldsymbol{\xi})$ for all $c > 0, \boldsymbol{\xi} \in \mathbb{R}^d$. Then Theorem 4.1 in [6] shows that there exists a stochastically continuous OSSRF

$$B(\boldsymbol{x}) = \mathrm{Re}\left[\int_{\boldsymbol{\xi}\in\mathbb{R}^d} \left(e^{i\langle \boldsymbol{x},\boldsymbol{\xi}\rangle} - 1\right)\psi(\boldsymbol{\xi})^{-1-q/\alpha} W_\alpha(d\boldsymbol{\xi})\right], \tag{2.1}$$

where $q = \mathrm{trace}(E)$, $\langle \boldsymbol{x}, \boldsymbol{\xi}\rangle = \sum_{i=1}^d x_i\xi_i$ and $W_\alpha(d\boldsymbol{\xi})$ is a complex isotropic symmetric stable random measure with index $0 < \alpha \le 2$ and control measure $m(d\boldsymbol{\xi}) = \sigma_0^2 d\boldsymbol{\xi}$. If $\alpha = 2$, then $B(\boldsymbol{x})$ is a Gaussian random field, and for any Borel subset A of $\mathbb{R}^d$ we have $W_2(A) = Z_1 + iZ_2$, where Z_1 and Z_2 are independent and identically distributed (i.i.d.) Gaussian random variables on $\mathbb{R}^1$ with mean zero and variance $\sigma_0^2|A|/2$, so $\mathbb{E}[W_2(A)^2] = |A|$, the Lebesgue measure of A. Corollary 4.2 in [6] shows that the OSSRF (2.1) has stationary increments, and that the operator scaling property (1.1) holds. See for example Samorodnitsky and Taqqu [20] for general details on stable stochastic integrals.

Next, we review a spectral method for simulating the OSSRF (2.1), using a fast Fourier transform (FFT); see Kegel [15] for complete details. This method yields a spatial regression model for OSSRFs that is the basis for our parameter estimation scheme. To simplify the discussion, we focus on the case of Gaussian OSSRFs with $\alpha = 2$ in two dimensions. However, everything extends easily to stable OSSRFs on $\mathbb{R}^d$ with index $0 < \alpha < 2$. First, we approximate the stochastic integral in (2.1) by a Riemann sum. Let $\mathcal{D} = [-A, A]^2 \setminus (-B, B)^2 \subset \mathbb{R}^2$ be a large square centered at the origin with radius A, with a much smaller square of radius B deleted to form an annular region, such that B/A is a rational number. Select a large integer M such that $(B/A)M$ is also an integer. Next we subdivide the region $\mathcal{D}$ into small squares of size A/M. Define $\mathcal{I} = \{-M, \ldots, M-1\}^2$ and $\mathcal{J} = \mathcal{I} \setminus \{-(B/A)M, \ldots, (B/A)M - 1\}^2$, a collection of integer grid points in $\mathbb{R}^2$, and set $\boldsymbol{\xi}_{\boldsymbol{k}} = (A/M)\boldsymbol{k}$ for $\boldsymbol{k} = (k_1, k_2) \in \mathcal{J}$. Now let $\Delta\boldsymbol{\xi}_{\boldsymbol{k}}$ be the square of side A/M with the point $\boldsymbol{\xi}_{\boldsymbol{k}}$ at its southwest corner, i.e., $\Delta\boldsymbol{\xi}_{\boldsymbol{k}} = [(A/M)k_1, (A/M)(k_1+1)] \times [(A/M)k_2, (A/M)(k_2+1)]$. Then we define

$$J_{\mathcal{D}}(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathcal{J}} \left(e^{i\langle \boldsymbol{x},\boldsymbol{\xi}_{\boldsymbol{k}}\rangle} - 1\right)\psi(\boldsymbol{\xi}_{\boldsymbol{k}})^{-1-q/2} W_2(\Delta\boldsymbol{\xi}_{\boldsymbol{k}}), \tag{2.2}$$

where the complex-valued random variables $W_2(\Delta\boldsymbol{\xi}_i)$ are i.i.d. with $(\sigma_0 A/M)(Z_1 + iZ_2)$, and Z_1 and Z_2 are i.i.d. $\mathcal{N}(0, 1/2)$. As $M \to \infty$, the approximating sum $J_{\mathcal{D}}(\boldsymbol{x})$ converges in probability to the stochastic integral

$$I_{\mathcal{D}}(\boldsymbol{x}) = \int_{\boldsymbol{\xi}\in\mathcal{D}} \left(e^{i\langle \boldsymbol{x},\boldsymbol{\xi}\rangle} - 1\right)\psi(\boldsymbol{\xi})^{-1-q/2}\, W_2(d\boldsymbol{\xi}),$$

since the integrand is continuous on the compact set $\mathcal{D}$ (e.g., see [16, Section 7.7]). Since the stochastic integral (2.1) exists, $I_{\mathcal{D}}(\boldsymbol{x})$ converges in probability to (2.1) as $A \to \infty$ and $B \to 0$.

Next we write $I_{\mathcal{D}}(\boldsymbol{x})$ in terms of discrete Fourier transforms. Set

$$g(\boldsymbol{k}) = \begin{cases} \psi(\boldsymbol{\xi_k})^{-1-q/2} & \text{for } \boldsymbol{k} \in \mathcal{J} \\ 0 & \text{for } \boldsymbol{k} \in \mathcal{I} \setminus \mathcal{J} \end{cases}$$

and $e_{\boldsymbol{k}} = W_2(\Delta\boldsymbol{\xi_k})$. Then we can write $J_{\mathcal{D}}(\boldsymbol{x}) = \bar{J}_{\mathcal{D}}(\boldsymbol{x}) - \bar{J}_{\mathcal{D}}(\boldsymbol{0})$, where

$$\bar{J}_{\mathcal{D}}(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathcal{I}} e^{i\langle \boldsymbol{x}, \boldsymbol{\xi_k}\rangle} g(\boldsymbol{k}) e_{\boldsymbol{k}}. \tag{2.3}$$

For an arbitrary set of spatial coordinates $\{\boldsymbol{x_k} : \boldsymbol{k} \in \mathcal{I}\}$, we can then view $\{\bar{J}_{\mathcal{D}}(\boldsymbol{x_k}) : \boldsymbol{k} \in \mathcal{I}\}$ as the two-dimensional inverse discrete Fourier transform of $\{g(\boldsymbol{k})e_{\boldsymbol{k}} : \boldsymbol{k} \in \mathcal{I}\}$. Then the FFT algorithm can be used to efficiently compute $\{\bar{J}_{\mathcal{D}}(\boldsymbol{x_k}) : \boldsymbol{k} \in \mathcal{I}\}$, yielding the approximation $B(\boldsymbol{x_k}) \approx \bar{J}_{\mathcal{D}}(\boldsymbol{x_k}) - \bar{J}_{\mathcal{D}}(\boldsymbol{0})$, for all $\boldsymbol{k} \in \mathcal{I}$. If we take $\psi(\boldsymbol{\xi}) = \|\boldsymbol{\xi}\|^H$ and $E = (1/H)I$, where I is the identity matrix, then $\psi(\boldsymbol{\xi})^{-1-q/2} = \|\boldsymbol{\xi}\|^{-H-d/2}$, and this reduces to the well-known spectral simulation method for a Lévy fractional Brownian field with Hurst index $0 < H < 1$ (e.g., see Voss [21]). Then (2.1) is the spectral representation for this isotropic random field. The OSSRF model is an extension of the Lévy fractional Brownian field that allows the Hurst index H to vary with the coordinate, in an arbitrary coordinate system.

Next, we outline our proposed parameter estimation method, which is based on the spectral simulation method described above. Suppose that we are given a set of $2M \times 2M$ spatial data $\{(\boldsymbol{x_k}, w_{\boldsymbol{k}}) : \boldsymbol{k} \in \mathcal{I}\}$, where the observation $w_{\boldsymbol{k}}$ is located at the spatial coordinates given by the vector $\boldsymbol{x_k}$. Since the random field model (2.1) has the property that $B(\boldsymbol{0}) = 0$, we suppose that $w_{\boldsymbol{k}} - w_{\boldsymbol{0}}$ comes from a realization of the OSSRF (2.1). Using the discrete approximation, we therefore suppose that $w_{\boldsymbol{k}} = \bar{J}_{\mathcal{D}}(\boldsymbol{x_k})$. Compute $\{z_{\boldsymbol{k}} : \boldsymbol{k} \in \mathcal{I}\}$ by taking the FFT of $\{w_{\boldsymbol{k}} : \boldsymbol{k} \in \mathcal{I}\}$. These Fourier transformed data $\{z_{\boldsymbol{k}} : \boldsymbol{k} \in \mathcal{I}\}$ satisfy a multiplicative model $z_{\boldsymbol{k}} = g(\boldsymbol{k}) \cdot e_{\boldsymbol{k}}$, where the coefficients $g(\boldsymbol{k}) := \psi(\boldsymbol{\xi_k})^{-1-q/2} I(\boldsymbol{k} \in \mathcal{J})$ depend on the Fourier filter $\psi(\boldsymbol{\xi})$, and $e_{\boldsymbol{k}}$ are i.i.d. complex-valued Gaussian with mean zero and $\mathbb{E}[|e_{\boldsymbol{k}}|^2] = \sigma_0^2 (A/M)^2 = |\Delta\boldsymbol{\xi_k}|$.

Next, we consider a simple parametric model for the Fourier filter. Suppose that the scaling matrix E has a basis of eigenvectors $\boldsymbol{b}_1, \boldsymbol{b}_2$ with associated eigenvalues $a_1, a_2 \in (0, \infty)$. To reduce the number of parameters, we adopt the polar representation $\boldsymbol{b}(v_j) = (\cos(v_j), \sin(v_j))$ with angle $v_j \in [0, \pi]$ for $j = 1, 2$ instead of $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$. Then it is not hard to check that

$$\psi(\boldsymbol{\xi}) = \left(C|\langle \boldsymbol{\xi}, \boldsymbol{b}(v_1)\rangle|^{2/a_1} + |\langle \boldsymbol{\xi}, \boldsymbol{b}(v_2)\rangle|^{2/a_2} \right)^{1/2} \tag{2.4}$$

is an E^T-homogeneous function. Note also that $q = \operatorname{trace}(E) = a_1 + a_2$. Substituting the Fourier filter (2.4) into the spectral representation (2.1) gives a five-parameter family of OSSRF models with parameters a_1, a_2, v_1, v_2, and C. More general models are developed in Clausel and Vedel [8], to allow complex and/or degenerate eigenvalues.

Assuming the Fourier filter (2.4), the Fourier transformed data follow a multiplicative model $z_{\boldsymbol{k}} = g(\boldsymbol{k}) \cdot e_{\boldsymbol{k}}$, which converts to an additive nonlinear regression model

$$y_{\boldsymbol{k}} = \log|z_{\boldsymbol{k}}| = \log|g(\boldsymbol{k})| + \log|e_{\boldsymbol{k}}|, \tag{2.5}$$

and we will estimate the parameters a_1, a_2, v_1, v_2, and C by solving this nonlinear regression problem. The nonlinear regression is complicated by the fact that the errors $\log|e_{\boldsymbol{k}}|$ do not have mean zero. Hence, we propose a new method for nonlinear regression in Section 3 that allows i.i.d. errors with a nonzero mean, and we prove that this method leads to consistent and asymptotically normal parameter estimates. Although the index $\boldsymbol{k}$ in (2.5) is a vector, this detail is irrelevant to the nonlinear regression problem, and hence without loss of generality we consider the traditional form in which the index is a positive integer.

3. Nonlinear regression

In this section, we develop a new method for nonlinear regression when the regression errors have a (possibly) nonzero mean. We consider the following nonlinear regression model.

$$y_i^{(n)} = f(\boldsymbol{x}_i^{(n)}; \boldsymbol{\theta}) + \epsilon_i^{(n)}, \quad \text{for } i = 1, 2, \ldots, n;\ n = 1, 2, 3, \ldots, \tag{3.1}$$

where $f(\boldsymbol{x}; \boldsymbol{\theta})$ is a nonlinear function on $\boldsymbol{x} \in \mathcal{D} \subset \mathbb{R}^d$ that depends on the parameter vector $(\theta_1, \ldots, \theta_p)^T := \boldsymbol{\theta}$. To simplify the notation, we suppress the index n throughout the remainder of the paper. We assume that, for every fixed n, the errors ϵ_i are independent and identically distributed (i.i.d.) with $\mathrm{E}(\epsilon_i) = \mu_n$ and $\mathrm{Var}(\epsilon_i) = \sigma^2 > 0$. We assume that the variance σ^2 is the same for all n, but we allow that μ_n varies with n. Next, we develop a new nonlinear regression estimator for the parameter vector $\boldsymbol{\theta}$. We show that the resulting parameter estimates $\hat{\boldsymbol{\theta}}_n$ are strongly consistent and asymptotically normal as $n \to \infty$. The model (3.1) arises naturally from a multiplicative model such as $z_i = g(\boldsymbol{x}_i, \boldsymbol{\theta}) \cdot e_i$, where the multiplicative errors e_i are i.i.d. By taking absolute values and then logarithms on both sides, we obtain $\log|z_i| = \log|g(\boldsymbol{x}_i, \boldsymbol{\theta})| + \log|e_i|$, which can be viewed as the model given in (3.1). The application to OSSRFs in Section 4 involves infill asymptotics, and $\mu_n = \mathrm{E}(\log|e_i|) \to -\infty$ as $n \to \infty$, which is permissible under our model.

Since the error ϵ_i in (3.1) has nonzero mean, we propose a new least square estimator

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in\Theta} S_n(\boldsymbol{\theta}), \tag{3.2}$$

where Θ is a parameter space, and

$$S_n(\boldsymbol{\theta}) = \sum_{i=1}^{n}\left(y_i - f(\boldsymbol{x}_i;\boldsymbol{\theta}) - \frac{1}{n}\sum_{j=1}^{n}\left(y_j - f(\boldsymbol{x}_j;\boldsymbol{\theta})\right)\right)^2 \tag{3.3}$$

is the objective function. The idea is to minimize errors induced only by the model, since the mean of ϵ_i is removed by subtracting its sample mean.

Remark 3.1. Bhattacharyya et al. [5] showed inconsistency of the least square estimator of a nonlinear regression model with multiplicative error. Note that the paper [5] considered the least square estimator of the original model, that is, without logarithmic transformation. That paper mentioned the strong consistency of the least square estimator after logarithmic transformation, but the objective function is different from ours.

3.1. Strong consistency

Now we will prove the strong consistency of the proposed estimator (3.2). Note that $S_n(\boldsymbol{\theta})$ can be written in the following matrix form:

$$S_n(\boldsymbol{\theta}) = (Y - \boldsymbol{f}(\boldsymbol{\theta}))^T\left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)(Y - \boldsymbol{f}(\boldsymbol{\theta})), \tag{3.4}$$

where $Y = (y_1, \ldots, y_n)^T$, $\boldsymbol{f}(\boldsymbol{\theta}) = (f(\boldsymbol{x}_1;\boldsymbol{\theta}), \ldots, f(\boldsymbol{x}_n;\boldsymbol{\theta}))^T$, I is the identity matrix, and $\mathbf{1}$ is a vector of 1s. Let $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^T$ and $\Sigma_n = I - \frac{1}{n}\mathbf{1}\mathbf{1}^T$. Define $d_i(\boldsymbol{\theta}, \boldsymbol{\theta}') = f(\boldsymbol{x}_i;\boldsymbol{\theta}) - f(\boldsymbol{x}_i;\boldsymbol{\theta}')$ and $D_n(\boldsymbol{\theta}, \boldsymbol{\theta}') = \boldsymbol{d}^T\Sigma_n\boldsymbol{d}$, where $\boldsymbol{d} = (d_1(\boldsymbol{\theta}, \boldsymbol{\theta}'), \ldots, d_n(\boldsymbol{\theta}, \boldsymbol{\theta}'))^T$. For the proof of strong consistency, we need the following assumption.

Assumption 1. Assume that

(i) $\mathcal{D}$ and Θ are compact,
(ii) f is continuous on $\mathcal{D} \times \Theta$,
(iii) there exist functions $B_\gamma(\boldsymbol{\theta}, \boldsymbol{\theta}')$ for $\gamma = 1, 2$ such that

$$\frac{1}{n}\sum_{i=1}^{n} d_i(\boldsymbol{\theta}, \boldsymbol{\theta}')^\gamma \to B_\gamma(\boldsymbol{\theta}, \boldsymbol{\theta}') \quad \text{uniformly in } \boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta,$$

with $B(\boldsymbol{\theta}, \boldsymbol{\theta}') := B_2(\boldsymbol{\theta}, \boldsymbol{\theta}') - B_1(\boldsymbol{\theta}, \boldsymbol{\theta}')^2 \geq 0$, and $B(\boldsymbol{\theta}, \boldsymbol{\theta}') = 0$ if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}'$.

Theorem 3.2. *For the least square estimator given in (3.2), under Assumption 1, $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ a.s., where $\boldsymbol{\theta}_0$ is the true parameter value.*

The following lemma from Wu [22] will be used in the proof of Theorem 3.2.

Lemma 3.3. *Let Θ be a parameter space, and let $\boldsymbol{\theta}_0 \in \Theta$ be the true parameter value. Under Assumption 1(i) and (ii), suppose that, for any $\delta > 0$,*

$$\liminf_{n\to\infty}\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta}(S_n(\boldsymbol{\theta}) - S_n(\boldsymbol{\theta}_0)) > 0 \quad a.s. \tag{3.5}$$

Then, for $\hat{\boldsymbol{\theta}}_n$ given in (3.2), $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ a.s. as $n \to \infty$.

Proof. Lemma 1 in Wu [22] is for the classic least square estimator of the nonlinear regression model with mean zero errors. That is, $S_n(\boldsymbol{\theta}) = \sum_{i=1}^{n}(y_i - f(x_i;\boldsymbol{\theta}))^2$. However, the lemma still holds for our setting, since the proof does not make use of any mean zero error assumption, and the lemma is applicable to any estimation procedure based on minimization of a certain function, as Wu [22] noted. □

Proof of Theorem 3.2. $S_n(\boldsymbol{\theta}) - S_n(\boldsymbol{\theta}_0)$ can be written in a matrix form:

$$S_n(\boldsymbol{\theta}) - S_n(\boldsymbol{\theta}_0) = \boldsymbol{d}^T\Sigma_n\boldsymbol{d} - 2\,\boldsymbol{d}^T\Sigma_n\boldsymbol{\epsilon}.$$

Let $u_i = \epsilon_i - \mu$ and $\boldsymbol{u} = (u_1, \ldots, u_n)^T$. Then, u_i are i.i.d. with $\mathrm{E}(u_i) = 0$ and $\mathrm{Var}(u_i) = \sigma^2$. Since $\Sigma_n\mathbf{1} = \mathbf{0}$, we have

$$\begin{aligned} S_n(\boldsymbol{\theta}) - S_n(\boldsymbol{\theta}_0) &= \boldsymbol{d}^T\Sigma_n\boldsymbol{d} - 2\,\boldsymbol{d}^T\Sigma_n\boldsymbol{u} \\ &= D_n(\boldsymbol{\theta}, \boldsymbol{\theta}_0)\left(1 - \frac{2}{D_n(\boldsymbol{\theta}, \boldsymbol{\theta}_0)}\boldsymbol{d}^T\Sigma_n\boldsymbol{u}\right) \end{aligned}$$

and

$$\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta}(S_n(\boldsymbol{\theta})-S_n(\boldsymbol{\theta}_0))=\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta}D_n(\boldsymbol{\theta},\boldsymbol{\theta}_0)\left(1-\frac{2}{D_n(\boldsymbol{\theta},\boldsymbol{\theta}_0)}\boldsymbol{d}^T\Sigma_n\boldsymbol{u}\right).$$

Then, by Lemma 3.3, it is enough to show that

$$\frac{\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta}|\boldsymbol{d}^T\Sigma_n\boldsymbol{u}|}{\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta}D_n(\boldsymbol{\theta},\boldsymbol{\theta}_0)}\to 0\quad\text{as } n\to\infty, \tag{3.6}$$

since $(1/n)D_n\to B$ by Assumption 1(iii).

Note that we have $\boldsymbol{d}^T\Sigma_n\boldsymbol{u}=\sum_{i=1}^n d_iu_i-\left(\frac{1}{n}\sum_{i=1}^n d_i\right)\left(\sum_{i=1}^n u_i\right)$. By the strong law of large numbers, $\frac{1}{n}\sum_i u_i\to 0$ a.s., and, by Assumption 1(iii), $\left|\frac{1}{n}\sum_i d_i\right|<\infty$ uniformly in $\boldsymbol{\theta},\boldsymbol{\theta}_0\in\Theta$. Thus,

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta}\left(\frac{1}{n}\sum_{i=1}^n d_i\right)\left(\frac{1}{n}\sum_{i=1}^n u_i\right)\to 0\quad\text{a.s.}$$

To show that $\frac{1}{n}\sum_{i=1}^n d_iu_i\to 0$ a.s., we apply a theorem of Jenrich [14, Theorem 4]. Note that $l\equiv(u_i)_{i=1,2,\ldots}$ satisfies assumption (a) on p. 633 of [14] and that $g\equiv(d_i(\boldsymbol{\theta},\boldsymbol{\theta}_0))_{i=1,2,\ldots}$ satisfies

$$[g,g]\equiv\lim_{n\to\infty}\frac{1}{n}\sum_i d_i(\boldsymbol{\theta},\boldsymbol{\theta}_0)^2\to B_2(\boldsymbol{\theta},\boldsymbol{\theta}_0)$$

uniformly in $\boldsymbol{\theta},\boldsymbol{\theta}_0\in\Theta$. Then, [14, Theorem 4] implies that, almost surely, $\frac{1}{n}\sum_{i=1}^n d_iu_i\to 0$ uniformly in $\boldsymbol{\theta},\boldsymbol{\theta}_0\in\Theta$ as $n\to\infty$. That is, almost surely, $\sup_{\boldsymbol{\theta},\boldsymbol{\theta}'\in\Theta}\frac{1}{n}\sum_{i=1}^n d_iu_i\to 0$, which implies that

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta}\frac{1}{n}\sum_{i=1}^n d_iu_i\to 0,\quad\text{a.s.}$$

Therefore, in view of Assumption 1(iii), (3.6) holds. □

3.2. Asymptotic normality

To show asymptotic normality, we make further assumptions on f. If f is twice differentiable, let $\boldsymbol{f}_k(\boldsymbol{\theta})=(\partial f(\boldsymbol{x}_1;\boldsymbol{\theta})/\partial\theta_k,\ldots,\partial f(\boldsymbol{x}_n;\boldsymbol{\theta})/\partial\theta_k)^T$, $\boldsymbol{f}_{ks}(\boldsymbol{\theta})=(\partial^2 f(\boldsymbol{x}_1;\boldsymbol{\theta})/\partial\theta_k\partial\theta_s,\ldots,\partial^2 f(\boldsymbol{x}_n;\boldsymbol{\theta})/\partial\theta_k\partial\theta_s)^T$, $\dot{\mathbf{F}}(\boldsymbol{\theta})=(\boldsymbol{f}_1,\ldots,\boldsymbol{f}_p)$, and $\ddot{\mathbf{F}}(\boldsymbol{\theta})=\text{Block}(\mathbf{f}_{ks})$; that is, $\ddot{\mathbf{F}}$ is a $pn\times p$ block matrix whose (k,s)th block is $\boldsymbol{f}_{ks}$. Also, let $\mathcal{N}(\boldsymbol{\mu},\Sigma)$ denote the Gaussian or normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ. Now, we consider the following assumption for asymptotic normality.

Assumption 2. Assume further the following.

(i) The true parameter $\boldsymbol{\theta}_0$ is in the interior of $\boldsymbol{\Theta}$, and $f(\boldsymbol{x}_i;\boldsymbol{\theta})$ is twice differentiable with respect to $\boldsymbol{\theta}$ near $\boldsymbol{\theta}_0$ for all i.
(ii) There exists a positive definite matrix Γ such that $\frac{1}{n}\dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T\Sigma_n\dot{\mathbf{F}}(\boldsymbol{\theta}_0)\to\boldsymbol{\Gamma}$ as $n\to\infty$.
(iii) $\dot{\mathbf{F}}(\boldsymbol{\theta}_1)^T\Sigma_n\dot{\mathbf{F}}(\boldsymbol{\theta}_1)\left(\dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T\Sigma_n\dot{\mathbf{F}}(\boldsymbol{\theta}_0)\right)^{-1}$ converges to the identity matrix uniformly as $n\to\infty$ and $\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_0\|\to 0$.
(iv) There exists a $\delta>0$ such that

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|\leq\delta}\left(\frac{\partial^2 f(\boldsymbol{x}_i;\boldsymbol{\theta})}{\partial\theta_k\partial\theta_s}\right)^2<\infty$$

for all $k,s=1,\ldots,p$.

Theorem 3.4. *Let $\hat{\boldsymbol{\theta}}_n$ be the strongly consistent estimator (3.2) for model (3.1). Then, under Assumption 2,*

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\sigma^2\boldsymbol{\Gamma}^{-1}\right). \tag{3.7}$$

Proof. Note that the first two derivatives of $S_n(\boldsymbol{\theta})$ are

$$\dot{S}_n(\boldsymbol{\theta})=2\dot{\mathbf{F}}(\boldsymbol{\theta})^T\Sigma_n(\boldsymbol{d}-\boldsymbol{\epsilon}) \tag{3.8}$$

and

$$\ddot{S}_n(\boldsymbol{\theta})=2\dot{\mathbf{F}}(\boldsymbol{\theta})^T\Sigma_n\dot{\mathbf{F}}(\boldsymbol{\theta})+2\ddot{\mathbf{F}}(\boldsymbol{\theta})^T(I\otimes(\Sigma_n(\boldsymbol{d}-\boldsymbol{\epsilon}))). \tag{3.9}$$

Since $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ a.s., by the mean value theorem, we have

$$\dot{S}_n(\boldsymbol{\theta}_0) = \ddot{S}_n(\tilde{\boldsymbol{\theta}}_n)(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_n), \tag{3.10}$$

where $\tilde{\boldsymbol{\theta}}_n$ lies on a line segment between $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}_n$. Eq. (3.10) can be rewritten as

$$(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \boldsymbol{Z}_n \left(\dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T \Sigma_n \dot{\mathbf{F}}(\boldsymbol{\theta}_0)\right)^{-1} \dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T \Sigma_n \boldsymbol{u}, \tag{3.11}$$

where

$$\boldsymbol{Z}_n = \left(\dot{\mathbf{F}}(\tilde{\boldsymbol{\theta}}_n)^T \Sigma_n \dot{\mathbf{F}}(\tilde{\boldsymbol{\theta}}_n) + \ddot{\mathbf{F}}(\tilde{\boldsymbol{\theta}}_n)^T (I \otimes (\Sigma_n(\boldsymbol{d} - \boldsymbol{u})))\right)^{-1} \left(\dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T \Sigma_n \dot{\mathbf{F}}(\boldsymbol{\theta}_0)\right). \tag{3.12}$$

Note that $\Sigma_n \boldsymbol{\epsilon}$ in (3.11) and (3.12) is replaced by $\Sigma_n \boldsymbol{u} = \Sigma_n(\boldsymbol{\epsilon} - \mu \mathbf{1})$, since $\Sigma_n \mathbf{1} = 0$. It is enough to show that

$$\boldsymbol{Z}_n \longrightarrow I \quad \text{a.s.} \tag{3.13}$$

and

$$\left(\dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T \Sigma_n \dot{\mathbf{F}}(\boldsymbol{\theta}_0)\right)^{-1/2} \dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T \Sigma_n \boldsymbol{u} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \sigma^2 I\right) \tag{3.14}$$

to complete the proof.

Since (3.13) is equivalent to $\boldsymbol{Z}_n^{-1} \longrightarrow I$ a.s., we consider $\boldsymbol{Z}_n^{-1}$, which can be rewritten as

$$\begin{aligned}\boldsymbol{Z}_n^{-1} &= \left(\dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T \Sigma_n \dot{\mathbf{F}}(\boldsymbol{\theta}_0)\right)^{-1} \dot{\mathbf{F}}(\tilde{\boldsymbol{\theta}}_n)^T \Sigma_n \dot{\mathbf{F}}(\tilde{\boldsymbol{\theta}}_n) + \left(\dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T \Sigma_n \dot{\mathbf{F}}(\boldsymbol{\theta}_0)\right)^{-1} \ddot{\mathbf{F}}(\tilde{\boldsymbol{\theta}}_n)^T (I \otimes (\Sigma_n \boldsymbol{d})) \\ &\quad - \left(\dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T \Sigma_n \dot{\mathbf{F}}(\boldsymbol{\theta}_0)\right)^{-1} \ddot{\mathbf{F}}(\tilde{\boldsymbol{\theta}}_n)^T (I \otimes (\Sigma_n \boldsymbol{u})).\end{aligned}$$

By condition (ii) of Assumption 2, the first term of $\boldsymbol{Z}_n^{-1}$ converges to I almost surely. By conditions (i), (ii) and (iv) of Assumption 2 and the Cauchy–Schwarz inequality, the second term of $\boldsymbol{Z}_n^{-1}$ converges to zero almost surely. To show that the third term of $\boldsymbol{Z}_n^{-1}$ converges to zero a.s., it is enough to show that, for all $k, s = 1, \ldots, p$,

$$\frac{1}{n} \boldsymbol{f}_{ks}(\boldsymbol{\theta})^T \Sigma_n \boldsymbol{u} \longrightarrow 0$$

uniformly on $\mathcal{S} = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \le \delta\}$ with probability 1, because of conditions (i) and (ii) of Assumption 2. Now $\frac{1}{n} \boldsymbol{f}_{ks}(\boldsymbol{\theta}) \Sigma_n \boldsymbol{u}$ is decomposed into two parts, so

$$\frac{1}{n} \sup_{\theta \in \mathcal{S}} |\boldsymbol{f}_{ks}(\boldsymbol{\theta}) \Sigma_n \boldsymbol{u}| \le \frac{1}{n} \sup_{\theta \in \mathcal{S}} \left| \sum_{i=1}^n \frac{\partial^2 f(i; \boldsymbol{\theta})}{\partial \theta_k \partial \theta_s} u_i \right| + \frac{1}{n} \sup_{\theta \in \mathcal{S}} \left| \sum_{i=1}^n \frac{\partial^2 f(i; \boldsymbol{\theta})}{\partial \theta_k \partial \theta_s} \right| \left| \frac{1}{n} \sum_{i=1}^n u_i \right|. \tag{3.15}$$

Then, by condition (iv) of Assumption 2, the first term of (3.15) converges to zero a.s., which can be shown in a manner similar to that of Wu [22, p. 509]. By the strong law of large numbers, condition (iv) of Assumption 2, and the Cauchy–Schwarz inequality, the second term of (3.15) converges to zero almost surely.

To show (3.14), we use a lemma of Huber [13, Lemma 2.1]. The condition of that lemma in our setting is

$$\left\| \left(\dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T \Sigma_n \dot{\mathbf{F}}(\boldsymbol{\theta}_0)\right)^{-1/2} \dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T \right\|_\infty \longrightarrow 0 \quad \text{as } n \to \infty, \tag{3.16}$$

where $\|A\|_\infty = \max_{1 \le i \le p, 1 \le j \le q} |a_{ij}|$ for a $p \times q$ matrix A. Note that for a $p \times q$ matrix A and a $q \times r$ matrix B, we have

$$\|AB\|_\infty \le q\, \|A\|_\infty \|B\|_\infty.$$

Since $\left(\dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T \Sigma_n \dot{\mathbf{F}}(\boldsymbol{\theta}_0)\right)^{-1/2}$ is a $p \times p$ matrix and $\dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T$ is a $p \times n$ matrix, we have

$$\begin{aligned}\left\| \left(\dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T \Sigma_n \dot{\mathbf{F}}(\boldsymbol{\theta}_0)\right)^{-1/2} \dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T \right\|_\infty &\le p \left\| \left(\dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T \Sigma_n \dot{\mathbf{F}}(\boldsymbol{\theta}_0)\right)^{-1/2} \right\|_\infty \left\| \dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T \right\|_\infty \\ &\le (p/\sqrt{n}) \left\| \left(\dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T \Sigma_n \dot{\mathbf{F}}(\boldsymbol{\theta}_0)/n\right)^{-1/2} \right\|_\infty \left\| \dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T \right\|_\infty.\end{aligned}$$

Then conditions (i) and (ii) of Assumption 2 imply that

$$\left\| \left(\dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T \Sigma_n \dot{\mathbf{F}}(\boldsymbol{\theta}_0)/n\right)^{-1/2} \right\|_\infty \quad \text{and} \quad \left\| \dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T \right\|_\infty$$

are bounded, so (3.16) holds, which leads to the asymptotic normality in (3.14). □

4. Parameter estimation for OSSRFs

Now, we apply the nonlinear regression estimator in Section 3 to develop a practical method for OSSRF parameter estimation. Theorem 3.4 requires a twice differentiable objective function, but the filter (2.4) contains absolute values. Hence, we consider a smoothed version of $\log|g(\boldsymbol{k})|$, by setting

$$f(\boldsymbol{\xi},\boldsymbol{\theta}) := \left(-\frac{1}{2}-\frac{a_1+a_2}{2}\right)\log\left(l_\varepsilon(C|\langle\boldsymbol{\xi},\boldsymbol{b}(v_1)\rangle|^{2/a_1})+l_\varepsilon(|\langle\boldsymbol{\xi},\boldsymbol{b}(v_2)\rangle|^{2/a_2})\right), \tag{4.1}$$

with parameter vector $\boldsymbol{\theta}=(a_1,a_2,v_1,v_2,C)^T$. Here, for a given small $\varepsilon>0$, the function $l_\varepsilon(\cdot)$ is a smooth cutoff function such that $l_\varepsilon(x)=x$ for $x>\varepsilon$ and $l_\epsilon(0)=0$. The smooth cutoff function is for mathematical convenience. It has no impact on the estimation method, since in practice M is fixed, so we can choose $\varepsilon<A/M$, and then $f(\boldsymbol{\xi}_{\boldsymbol{k}},\boldsymbol{\theta})=\log|g(\boldsymbol{k})|$ for all $\boldsymbol{k}\in\mathcal{I}$. To ease notation, index the grid points $\boldsymbol{k}=\boldsymbol{k}(i)$, where $i=1,2,\ldots,n:=(2M)^2$, and write $z_i=z_{\boldsymbol{k}(i)}$, and so forth. Then the nonlinear regression model (2.5) can be written in the form

$$y_i=f(\boldsymbol{\xi}_i,\boldsymbol{\theta})+\epsilon_i, \tag{4.2}$$

where $\epsilon_i=\log|e_i|$. Write $e_i=(\sigma_0A/M)Z_i$, where Z_i are i.i.d. complex-valued Gaussian with $\mathbb{E}[Z_i]=0$ and $\mathbb{E}[|Z_i|^2]=1$. Then $\mu_n=\mathrm{E}(\epsilon_i)=\log(\sigma_0A/M)+\mathrm{E}(\log|Z_i|)\to-\infty$ as $n=(2M)^2\to\infty$. However, the error variance $\sigma^2=\mathrm{Var}(\epsilon_i)=\mathrm{Var}(\log|Z_i|)$ does not depend on n. Hence we can estimate the parameter vector $\boldsymbol{\theta}$ of this OSSRF model by a nonlinear regression (3.2), and use the theory in Section 3 to get the asymptotics. For our OSSRF estimator, we consider the compact parameter space

$$\Theta=\{(a_1,a_2,v_1,v_2,C)|1\le a_1,a_2\le a_{\max},v_1,v_2\in[0,\pi],v_1<v_2,|v_1-v_2|\ge\delta_0,c_0^{-1}\le C\le c_0\},$$

where $a_{\max}>1$, and $\delta_0>0$ and $c_0^{-1}>0$ are small. These restrictions on the parameter space are for mathematical convenience, and they pose no limitations in terms of practical applications. Note that, since $H_i=a_i^{-1}\le 1$ is the Hurst scaling index in the ith coordinate direction, it suffices to consider $a_i\ge 1$.

Theorem 4.1. *Under the assumptions detailed above, the nonlinear regression estimator $\hat{\boldsymbol{\theta}}_n$ for the OSSRF parameter vector $\boldsymbol{\theta}=(a_1,a_2,v_1,v_2,C)$ is strongly consistent and asymptotically normal. That is, we have*

$$\hat{\boldsymbol{\theta}}_n\to\boldsymbol{\theta}_0\quad a.s.\ as\ n\to\infty,$$

where $\boldsymbol{\theta}_0$ is the vector of true parameter values, and also

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right)\stackrel{d}{\to}\mathcal{N}\left(\mathbf{0},\sigma^2\boldsymbol{\Gamma}^{-1}\right)$$

in distribution. The matrix $\boldsymbol{\Gamma}$ has (i,j)th entry given by (4.3) with $\boldsymbol{\theta}=\boldsymbol{\theta}_0$, and $\sigma^2=\mathrm{Var}(\log(Z_1^2+Z_2^2))$, where Z_1,Z_2 are i.i.d. $\mathcal{N}(0,1/2)$. The matrix $\boldsymbol{\Gamma}$ is positive definite, and $\frac{1}{n}\dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T\Sigma_n\dot{\mathbf{F}}(\boldsymbol{\theta}_0)\to\boldsymbol{\Gamma}$ as $n\to\infty$.

Proof. To prove strong consistency of the estimator $\hat{\boldsymbol{\theta}}$, we need to check Assumption 1(iii). This requires the following lemma.

Lemma 4.2. *For the parameter estimation of OSSRFs, we have, for $\gamma=1,2$,*

$$\frac{1}{n}\sum_{i=1}^n d_i(\boldsymbol{\theta},\boldsymbol{\theta}')^\gamma\to B_\gamma(\boldsymbol{\theta},\boldsymbol{\theta}')\quad uniformly\ in\ \boldsymbol{\theta},\boldsymbol{\theta}'\in\Theta,$$

as $n\to\infty$, where

$$B_\gamma(\boldsymbol{\theta},\boldsymbol{\theta}')=\frac{1}{(2A)^2}\int_{[-A,A]^2}\left(f(\boldsymbol{\xi},\boldsymbol{\theta})-f(\boldsymbol{\xi},\boldsymbol{\theta}')\right)^\gamma d\boldsymbol{\xi}.$$

Moreover, $B(\boldsymbol{\theta},\boldsymbol{\theta}')=B_2(\boldsymbol{\theta},\boldsymbol{\theta}')-B_1(\boldsymbol{\theta},\boldsymbol{\theta}')^2\ge 0$ and $B(\boldsymbol{\theta},\boldsymbol{\theta}')=0$ if and only if $\boldsymbol{\theta}=\boldsymbol{\theta}'$.

Proof. We have

$$\begin{aligned}\frac{1}{n}\sum_i d_i^2 &= \frac{1}{|\mathcal{I}|}\sum_{\boldsymbol{k}\in\mathcal{I}}\left(f(\boldsymbol{\xi}_{\boldsymbol{k}},\boldsymbol{\theta})-f(\boldsymbol{\xi}_{\boldsymbol{k}},\boldsymbol{\theta}')\right)^2\\ &= \frac{1}{(2A)^2}\left(\frac{A}{M}\right)^2\sum_{\boldsymbol{k}\in\mathcal{I}}\left(f(\boldsymbol{\xi}_{\boldsymbol{k}},\boldsymbol{\theta})-f(\boldsymbol{\xi}_{\boldsymbol{k}},\boldsymbol{\theta}')\right)^2.\end{aligned}$$

As $n \to \infty$, we have $M \to \infty$, and hence

$$\left(\frac{A}{M}\right)^2 \sum_{\boldsymbol{k}\in\mathcal{I}} \big(f(\boldsymbol{\xi}_{\boldsymbol{k}}, \boldsymbol{\theta}) - f(\boldsymbol{\xi}_{\boldsymbol{k}}, \boldsymbol{\theta}')\big)^2 \to \int_{[-A,A]^2} \big(f(\boldsymbol{\xi}, \boldsymbol{\theta}) - f(\boldsymbol{\xi}, \boldsymbol{\theta}')\big)^2 \, d\boldsymbol{\xi}.$$

Similarly,

$$\begin{aligned}\frac{1}{n}\sum_i d_i &= \frac{1}{|\mathcal{I}|}\sum_{\boldsymbol{k}\in\mathcal{I}} \big(f(\boldsymbol{\xi}_{\boldsymbol{k}}, \boldsymbol{\theta}) - f(\boldsymbol{\xi}_{\boldsymbol{k}}, \boldsymbol{\theta}')\big) \\ &\to \frac{1}{(2A)^2}\int_{[-A,A]^2} \big(f(\boldsymbol{\xi}, \boldsymbol{\theta}) - f(\boldsymbol{\xi}, \boldsymbol{\theta}')\big) \, d\boldsymbol{\xi}.\end{aligned}$$

Thus

$$\frac{1}{n}\sum_{i=1}^n d_i^2 - \left(\frac{1}{n}\sum_{i=1}^n d_i\right)^2 \to B$$

pointwise as $n \to \infty$. Uniform convergence in $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$ follows from the uniform continuity of $f(\boldsymbol{\xi}, \boldsymbol{\theta})$ in $\boldsymbol{\xi} \in [-A, A]^2$ and $\boldsymbol{\theta} \in \Theta$ by a standard argument. Also, observe that, by the Cauchy–Schwarz inequality, $B(\boldsymbol{\theta}, \boldsymbol{\theta}') \geq 0$ and $B(\boldsymbol{\theta}, \boldsymbol{\theta}') = 0$ if and only if $f(\boldsymbol{\xi}, \boldsymbol{\theta}) - f(\boldsymbol{\xi}, \boldsymbol{\theta}') =$ constant for all $\boldsymbol{\xi} \in \mathcal{D}$, which holds if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}'$. □

In order to apply the results of Section 3.2 in this particular case, observe that the function (4.1) is continuous over $(\boldsymbol{\xi}, \boldsymbol{\theta}) \in [-A, A]^2 \times \Theta$, and that $f(\boldsymbol{\xi}, \boldsymbol{\theta})$ is twice differentiable with respect to $\boldsymbol{\theta}$ near $\boldsymbol{\theta}_0$ for all $\boldsymbol{\xi} \in [-A, A]^2$ such that, for all $1 \leq i, j \leq p = 5$,

$$\begin{aligned}&\frac{\partial f}{\partial \theta_i}(\boldsymbol{\xi}, \boldsymbol{\theta}) : [-A, A]^2 \longrightarrow \mathbb{R} \\ &\frac{\partial^2 f}{\partial \theta_i \partial \theta_j}(\boldsymbol{\xi}, \boldsymbol{\theta}) : [-A, A]^2 \longrightarrow \mathbb{R}\end{aligned}$$

are continuous functions of $\boldsymbol{\xi} \in [-A, A]^2$ (and hence uniform continuous, by compactness of $[-A, A]^2$).

Next, we verify that the function (4.1) satisfies Assumption 2, so the least squares estimator is asymptotically normal, i.e., (3.7) holds.

For $\boldsymbol{\theta}$ near $\boldsymbol{\theta}_0$ and $i, j = 1, \ldots, p = 5$, define

$$\Gamma(\boldsymbol{\theta})_{i,j} = \frac{1}{(2A)^2}\int_{[-A,A]^2} \frac{\partial f(\boldsymbol{\xi}, \boldsymbol{\theta})}{\partial \theta_i}\frac{\partial f(\boldsymbol{\xi}, \boldsymbol{\theta})}{\partial \theta_j} \, d\boldsymbol{\xi} - \left(\frac{1}{(2A)^2}\int_{[-A,A]^2} \frac{\partial f(\boldsymbol{\xi}, \boldsymbol{\theta})}{\partial \theta_i} \, d\boldsymbol{\xi}\right)\left(\frac{1}{(2A)^2}\int_{[-A,A]^2} \frac{\partial f(\boldsymbol{\xi}, \boldsymbol{\theta})}{\partial \theta_j} \, d\boldsymbol{\xi}\right). \tag{4.3}$$

Lemma 4.3. *$\Gamma(\boldsymbol{\theta})$ is positive definite for all $\boldsymbol{\theta}$ near $\boldsymbol{\theta}_0$.*

Proof. Observe that, for $\lambda_1, \ldots, \lambda_p \in \mathbb{R}$, the quantity

$$\sum_{i,j=1}^p \lambda_i\lambda_j\Gamma(\boldsymbol{\theta})_{i,j} = \frac{1}{(2A)^2}\int_{[-A,A]^2}\left(\sum_{i=1}^p \lambda_i \frac{\partial f(\boldsymbol{\xi}, \boldsymbol{\theta})}{\partial \theta_i}\right)^2 d\boldsymbol{\xi} - \left(\frac{1}{(2A)^2}\int_{[-A,A]^2}\sum_{i=1}^p \lambda_i \frac{\partial f(\boldsymbol{\xi}, \boldsymbol{\theta})}{\partial \theta_i} \, d\boldsymbol{\xi}\right)^2$$

is greater than or equal to zero, by the Cauchy–Schwarz inequality. Moreover, if at least one $\lambda_i \neq 0$, then the function

$$\boldsymbol{\xi} \mapsto \sum_{i=1}^p \lambda_i \frac{\partial f(\boldsymbol{\xi}, \boldsymbol{\theta})}{\partial \theta_i}$$

is not constant on $\mathcal{D}$, and hence $\Gamma(\boldsymbol{\theta})$ is positive definite. □

By enlarging $\boldsymbol{\Theta}$ if necessary, we can assume that $\boldsymbol{\theta}_0$ is in the interior of $\boldsymbol{\Theta}$, and hence Assumption 2(i) holds true. For the proof of (ii), first note that, for $\boldsymbol{\theta}$ near $\boldsymbol{\theta}_0$, we have

$$\dot{\mathbf{F}}(\boldsymbol{\theta})^T \Sigma_n \dot{\mathbf{F}}(\boldsymbol{\theta}) = \dot{\mathbf{F}}(\boldsymbol{\theta})^T \dot{\mathbf{F}}(\boldsymbol{\theta}) - \frac{1}{n}\big(\dot{\mathbf{F}}(\boldsymbol{\theta})^T \mathbf{1}\big)\big(\mathbf{1}^T \dot{\mathbf{F}}(\boldsymbol{\theta})\big).$$

Now, for $i, j = 1, \ldots, p$ we have

$$\frac{1}{n}\big(\dot{\mathbf{F}}(\boldsymbol{\theta})^T \dot{\mathbf{F}}(\boldsymbol{\theta})\big)_{i,j} = \frac{1}{n}\sum_{l=1}^n \frac{\partial f(\boldsymbol{\xi}_l, \boldsymbol{\theta})}{\partial \theta_i}\frac{\partial f(\boldsymbol{\xi}_l, \boldsymbol{\theta})}{\partial \theta_j}$$

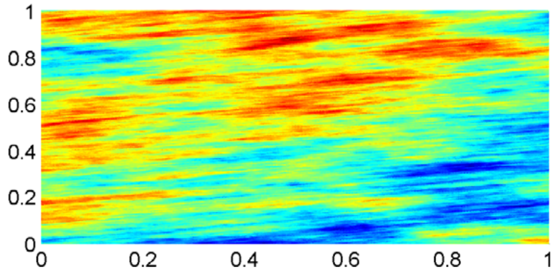

Fig. 1. Typical OSSRF realization with parameters $a_1 = 2.0$, $a_2 = 3.1$, $\upsilon_1 = 0.17$, $\upsilon_2 = 2.3$, and $C = 3.0$. The angles υ_1 and υ_2 determine the orientation of the random field, with Hurst index $H_1 = 1/a_1$ and $H_2 = 1/a_2$ in those directions.

$$= \frac{1}{(2A)^2}\left(\frac{A}{M}\right)^2 \sum_{k\in\mathcal{I}} \frac{\partial f(\boldsymbol{\xi}_k,\boldsymbol{\theta})}{\partial\theta_i}\frac{\partial f(\boldsymbol{\xi}_k,\boldsymbol{\theta})}{\partial\theta_j}$$
$$\to \frac{1}{(2A)^2}\int_{[-A,A]^2} \frac{\partial f(\boldsymbol{\xi},\boldsymbol{\theta})}{\partial\theta_i}\frac{\partial f(\boldsymbol{\xi},\boldsymbol{\theta})}{\partial\theta_j}\, d\boldsymbol{\xi}.$$

Similarly, we get

$$\left(\frac{1}{n}\dot{\mathbf{F}}(\boldsymbol{\theta})^T\mathbf{1}\right)_i \to \frac{1}{(2A)^2}\int_{[-A,A]^2}\frac{\partial f(\boldsymbol{\xi},\boldsymbol{\theta})}{\partial\theta_i}\, d\boldsymbol{\xi}.$$

Hence

$$\frac{1}{n}\dot{\mathbf{F}}(\boldsymbol{\theta})^T\varSigma_n\dot{\mathbf{F}}(\boldsymbol{\theta}) \to \varGamma(\boldsymbol{\theta}) \quad \text{as } n\to\infty. \tag{4.4}$$

Observe that, by compactness of $\mathcal{D}$ and uniform continuity, the convergence in (4.4) is uniform in $\boldsymbol{\theta}$ near $\boldsymbol{\theta}_0$. Hence Assumption 2(ii) holds.

To prove Assumption 2(iii), note that, in view of (4.4) and Lemma 4.3, we get

$$\dot{\mathbf{F}}(\boldsymbol{\theta}_1)^T\varSigma_n\dot{\mathbf{F}}(\boldsymbol{\theta}_1)\left(\dot{\mathbf{F}}(\boldsymbol{\theta}_0)^T\varSigma_n\dot{\mathbf{F}}(\boldsymbol{\theta}_0)\right)^{-1} \to \varGamma(\boldsymbol{\theta}_1)\varGamma(\boldsymbol{\theta}_0)^{-1} \quad \text{as } n\to\infty$$

uniformly in θ_1 near θ_0. Since $\theta \mapsto \varGamma(\theta)$ is continuous, Assumption 2 (iii) holds.

Finally, since $\boldsymbol{\theta}_0$ is assumed to be in the interior of $\boldsymbol{\Theta}$, there exists a $\delta > 0$ such that $\{\boldsymbol{\theta} : |\boldsymbol{\theta}-\boldsymbol{\theta}_0| \le \delta\}$ lies in the interior of $\boldsymbol{\Theta}$ and $f(\boldsymbol{\xi},\boldsymbol{\theta})$ is twice continuously differentiable in $\boldsymbol{\theta}$ on that set. For $k, s = 1,\ldots,p$ fixed, let

$$h(\boldsymbol{\xi}) = \sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|\le\delta}\left(\frac{\partial^2 f(\boldsymbol{\xi},\boldsymbol{\theta})}{\partial\theta_k\partial\theta_s}\right)^2,$$

which is a continuous function on $\mathcal{D}$. Then

$$\frac{1}{n}\sum_{i=1}^{n}\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|\le\delta}\left(\frac{\partial^2 f(\boldsymbol{\xi}_i,\boldsymbol{\theta})}{\partial\theta_k\partial\theta_s}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}h(\boldsymbol{\xi}_i) \to \frac{1}{(2A)^2}\int_{[-A,A]^2}h(\boldsymbol{\xi})\,d\boldsymbol{\xi} < \infty,$$

so Assumption 2(iv) holds true. Now the proof is complete. □

Remark 4.4. The spectral representation (2.1) can be extended to a broad range of infinitely divisible random measures (e.g., tempered stable) for suitable Fourier filter functions $\psi(\boldsymbol{\xi})$. Unless the measure is stable, the operator scaling is lost. However, the model may still be useful in some applications (e.g., see [11]). The parameter estimation scheme detailed here applies equally well to such models, since we only need that the random variables $\log|W_\alpha(\Delta\boldsymbol{\xi}_i)|$ are i.i.d. with the same variance for all n.

5. Simulation study

In this section, we present the results of a small simulation study of the proposed estimation method for OSSRFs. We simulated $R = 1000$ realizations of the random field (2.1) using method (2.3) with Fourier filter function $\psi(\boldsymbol{\xi})$ given by (2.4). The parameters $a_1 = 2.0$, $a_2 = 3.1$, $\upsilon_1 = 0.17$, $\upsilon_2 = 2.3$, and $C = 3.0$ were used, we took $\sigma_0 = 1$, and we set $M = 256$, so $n = 262144 = (2M)^2$ points were generated for the random field. A typical realization is shown in Fig. 1. The parameter values are within the range of applications in ground water hydrology. In those applications, the random field represents some scalar-valued physical property of an underground aquifer, such as *hydraulic conductivity*, the reciprocal of the resistance that a fluid encounters while passing through the porous medium. The connected regions of high or low

Table 1
Statistical summary of $R = 1000$ simulation results.

Statistic	a_1	a_2	v_1	v_2	C
True value	2.0	3.1	0.17	2.3	3.0
Sample mean	1.998	3.092	0.170	2.300	3.005
Standard deviation	0.013	0.047	0.008	0.004	0.045

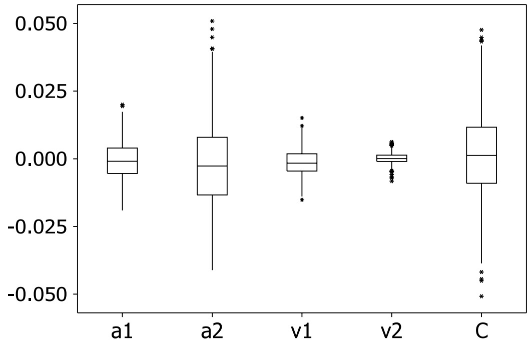

Fig. 2. Relative error for OSSRF parameter estimates, showing that the nonlinear regression method gives accurate estimates. A statistical summary for the same data is given in Table 1.

conductivity, called *facies*, control the ground water flow. Areas of high conductivity (red) form preferential paths for flow, leading to superdispersion. Areas of low conductivity (blue) lead to subdiffusion, as fluid that diffuses into the region cannot easily flow back out. Both are commonly observed in real-world experiments [2,3]. The vertical axis represents depth, and the horizontal axis is oriented in the mean flow direction. The Hurst index $H_2 = 1/a_2 = 0.323$ in the v_2 direction causes layering, since a fractional Brownian motion with Hurst index in the range $(0, 0.5)$ exhibits negative dependence. The "dipping angle" v_1 is typically nonzero, since the horizontal deposition of aquifer material is followed by geological changes that tilt or fold the aquifer. The structural features shown in Fig. 1 closely resemble a cut-out hillside exposed during road construction, or a river canyon, with strong anisotropy, and an orientation that differs from the gravity gradient. For the chosen parameter values, it was found that setting $A = 30.0$ and $B < A/(2M)$ (small enough so that the filter is set to zero only at the single point $\boldsymbol{k} = \mathbf{0}$) was sufficient to simulate the random field, i.e., increasing A or M or decreasing B made no apparent difference in the resulting graph. The nonlinear regression model (2.5) was then applied to estimate the parameters for each realization, treating them as unknown. The same value of A was used in the estimation procedure. The simulation and nonlinear regression were coded in MATLAB, using the command `fmincon` to solve the constrained optimization problem (3.2), where $S_n(\boldsymbol{\theta})$ is given by (3.3) and (4.1), over the parameter space Θ defined by $1 \le a_1, a_2 \le 40$, $0 \le v_1, v_2 \le \pi$, and $1/200 \le C \le 200$. Codes are available from the authors upon request.

Table 1 gives a statistical summary of the resulting parameter values from $R = 1000$ repeated simulations. The sample means of estimated parameter values are close to the true values, and the sample standard deviations are all less than 0.05. Fig. 2 shows a boxplot of relative error ((estimated − actual)/actual) for each parameter. The relative errors are small, indicating that the estimated parameter values are generally quite close to the assumed parameter values. This is also supported by the results in Table 1. The best accuracy is for the parameter estimates of the angles v_1 and v_2, which is gratifying, since identifying the coordinate axes $\boldsymbol{b}(v_1)$ and $\boldsymbol{b}(v_2)$ along which the Hurst indices $H_1 = 1/a_1$ and $H_2 = 1/a_2$ pertain was the main goal of this research.

Fig. 3 shows a histogram of simulation outputs for the v_1 estimates, with the best-fitting Gaussian density superimposed. Fig. 4 shows the corresponding probability plot (data versus model quantiles). Both plots strongly support the conclusion that the nonlinear regression estimates are normally distributed, consistent with the theory in Theorem 4.1. These data pass the Anderson–Darling test for normality ($p = 0.057$). Results for the remaining four parameters are similar (not shown).

6. Discussion

To our knowledge, the nonlinear regression approach developed in this paper is the first available method for estimating the parameters of anisotropic random fields whose Hurst index varies with the coordinate, in an arbitrary coordinate system that need not be orthogonal. Since such random fields are often encountered in practical applications (e.g., in ground water hydrology), the results of this paper can be applicable to those areas. In practical applications to real data, the parameters A and M should be chosen such that the approximation $J_{\mathcal{D}}(\boldsymbol{x}) - J_{\mathcal{D}}(\mathbf{0}) \approx B(\boldsymbol{x})$ that relates (2.1) to (2.2) is sufficiently accurate. The sample size of the data need not be as large as $n = (2M)^2$, but rather, its FFT needs to span a sufficient range of frequencies to accurately capture the shape of the Fourier filter $\psi(\boldsymbol{\xi})$ given by (2.4). The data need not be on a square grid, or evenly spaced. Theorem 4.1 can be used to determine confidence intervals for the parameter estimates. Of course, in real data applications, appropriate diagnostics should be employed to test the goodness of fit for model (2.5).

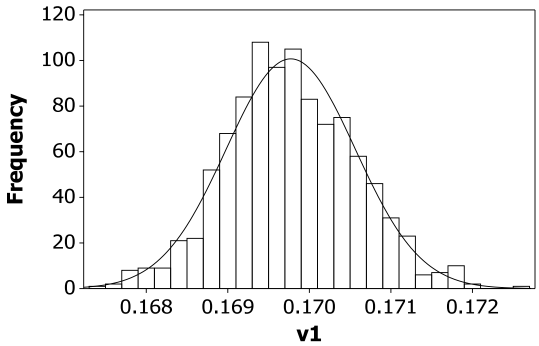

Fig. 3. Histogram of v_1 parameter estimates with the best-fitting Gaussian density function. The true parameter value was $v_1 = 0.17$.

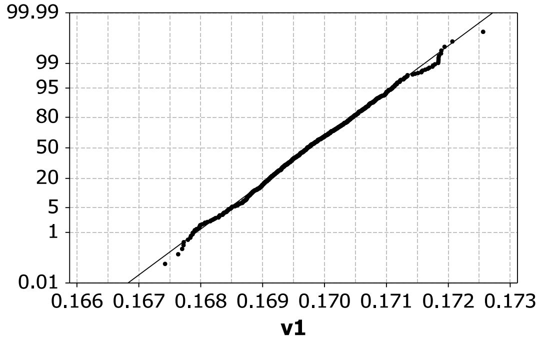

Fig. 4. Probability plot of v_1 parameter estimates, comparing data quantiles to model quantiles for the best-fitting Gaussian distribution. Since the plotted points show no systematic from the reference line (data = model), the fit is deemed adequate.

The methods of this paper can also be extended to a wider range of filters. The filter (2.4) is the Fourier symbol of an operator stable law with independent components in the eigenvector directions. Relaxing the independence assumption leads to a larger class of filters, discussed in [3] and applied to ground water hydrology in Monnig et al. [17], Reeves et al. [19], and Zhang et al. [23]. The approach in this paper applies to any parametric family of E^T-homogeneous filters $\psi(\boldsymbol{\xi})$, so all of those applications can be addressed. The explicit computation of those filters was discussed in Clausel and Vedel [8], and it would be interesting to apply the results of this paper with their filters.

Acknowledgments

MMM was partially supported by NSF grant DMS-1025486.

References

[1] R.J. Adler, The Geometry of Random Fields, John Wiley & Sons, 1981.
[2] D.A. Benson, M.M. Meerschaert, A simple and efficient random walk solution of multi-rate mobile/immobile mass transport equations, Adv. Water Resour. 32 (2009) 532–539.
[3] D.A. Benson, M.M. Meerschaert, B. Baeumer, H.P. Scheffler, Aquifer operator-scaling and the effect on solute mixing and dispersion, Water Resour. Res. 42 (1) (2006) W01415. http://dx.doi.org/10.1029/2004WR003755.
[4] J. Beran, S. Ghosh, D. Schell, On least squares estimation for long-memory lattice processes, J. Multivariate Anal. 100 (10) (2009) 2178–2194.
[5] B.B. Bhattacharyya, T.M. Khoshgoftaar, G.D. Richardson, Inconsistent *M*-estimators: nonlinear regression with multiplicative error, Statist. Probab. Lett. 14 (1992) 407–411.
[6] H. Biermé, M.M. Meerschaert, H.-P. Scheffler, Operator scaling stable random fields, Stochastic Process. Appl. 117 (2007) 312–332.
[7] Y. Boissy, B.B. Bhattacharyya, X. Li, G.D. Richardson, Parameter estimates for fractional autoregressive spatial processes, Ann. Statist. 33 (2005) 2553–2567.
[8] M. Clausel, B. Vedel, Explicit construction of operator scaling Gaussian random fields, Fractals 19 (1) (2011) 101.
[9] P. Embrechts, M. Maejima, Selfsimilar Processes, Princeton University Press, Princeton, NJ, 2002.
[10] H. Guo, C.Y. Lim, M.M. Meerschaert, Local Whittle estimator for anisotropic random fields, J. Multivariate Anal. 100 (5) (2009) 993–1028.
[11] C. Houdré, R. Kawai, On fractional tempered stable motion, Stochastic Process. Appl. 116 (8) (2006) 161–1184.
[12] B.X. Hu, M.M. Meerschaert, W. Barrash, D.W. Hyndman, C. He, X. Li, L. Guo, Examining the influence of heterogeneous porosity fields on conservative solute transport, J. Contam. Hydrol. 108 (3–4) (2009) 77–88.
[13] P.J. Huber, Robust regression: asymptotics, conjectures and Monte Carlo, Ann. Statist. 1 (1973) 799–821.
[14] R.I. Jenrich, Asymptotic properties of non-linear least squares estimators, Ann. Math. Statist. 40 (1969) 633–643.
[15] T. Kegel, On simulation and estimation of operator scaling stable random fields, Doctoral Dissertation, University of Siegen, 2011.
[16] M.M. Meerschaert, A. Sikorskii, Stochastic Models for Fractional Calculus, in: De Gruyter Studies in Mathematics, vol. 43, De Gruyter, Berlin, 2012.
[17] N.D. Monnig, D.A. Benson, M.M. Meerschaert, Ensemble solute transport in 2-D operator-scaling random fields, Water Resour. Res. 44 (2008) W02434. http://dx.doi.org/10.1029/2007WR005998.

[18] L. Ponson, D. Bonamy, E. Bouchaud, Two-dimensional scaling properties of experimental fracture surfaces, Phys. Rev. Lett. 96 (2006) 035506. http://dx.doi.org/10.1029/2008WR006858.
[19] D.M. Reeves, D.A. Benson, M.M. Meerschaert, H.P. Scheffler, Transport of conservative solutes in simulated fracture networks 2. Ensemble solute transport and the correspondence to operator-stable limit distributions, Water Resour. Res. 44 (2008) W05410.
[20] G. Samorodnitsky, M.S. Taqqu, Stable Non-Gaussian Random Processes, Chapman and Hall, New York, 1994.
[21] R.F. Voss, Random fractals—self-affinity in noise, music, mountains, and clouds, Physica D 38 (1989) 362–371.
[22] C.-F. Wu, Asymptotic theory of nonlinear least squares estimation, Ann. Statist. 9 (1981) 501–513.
[23] Y. Zhang, D.A. Benson, M.M. Meerschaert, E.M. LaBolle, H.P. Scheffler, Random walk approximation of fractional-order multiscaling anomalous diffusion, Phys. Rev. E 74 (2006) 026706. http://dx.doi.org/10.1103/PhysRevE.74.026706.