ON THE CONSISTENT SEPARATION OF SCALE AND VARIANCE FOR GAUSSIAN RANDOM FIELDS

By Ethan Anderes

University of California at Davis

We present fixed domain asymptotic results that establish consistent estimates of the variance and scale parameters for a Gaussian random field with a geometric anisotropic Matérn autocovariance in dimension $d > 4$. When $d < 4$ this is impossible due to the mutual absolute continuity of Matérn Gaussian random fields with different scale and variance (see Zhang [33]). Informally, when $d > 4$, we show that one can estimate the coefficient on the principle irregular term accurately enough to get a consistent estimate of the coefficient on the second irregular term. These two coefficients can then be used to separate the scale and variance. We extend our results to the general problem of estimating a variance and geometric anisotropy for more general autocovariance functions. Our results illustrate the interaction between the accuracy of estimation, the smoothness of the random field, the dimension of the observation space, and the number of increments used for estimation. As a corollary, our results establish the orthogonality of Matérn Gaussian random fields with different parameters when $d > 4$. The case $d = 4$ is still open.

1. Introduction. A common situation in spatial statistics is when one has observations on a single realization of a random field Y at a large number of spatial points $\boldsymbol{t}_1, \boldsymbol{t}_2, \ldots$ within some bounded region $\Omega \subset \mathbb{R}^d$. One is then is faced with the problem of predicting some quantity that depends on Y at unobserved points in Ω. For example, one may want to predict $\int_\Omega Y(\boldsymbol{t})d\boldsymbol{t}$ or the derivative $Y'(\boldsymbol{t}_0)$ where $\boldsymbol{t}_0$ is an unobserved point in Ω. A common technique is to first estimate the covariance structure of Y, then predict using the estimated covariance. Typically, fully nonparametric estimation of the covariance is difficult since the observations are from one realization of the random field. In this case, it is common to consider a class of covariance structures indexed by a finite number of parameters which are then estimated from the observations (see [12] or [9] for an introduction to spatial statistical techniques).

Two common parameters found in many covariance models are an overall scale α and an overall variance σ^2. The simplest example of this model stipulates that the random field Y is a scale and amplitude chance by an unknown α and σ of a known random field Z. In

AMS 2000 subject classifications: Primary 60G60; secondary 62M30, 62M40

Keywords and phrases: Matérn class, quadratic variations, Gaussian random fields, infill asymptotics, covariance estimation, principle irregular term

particular, for a spatial domain $\Omega \subset \mathbb{R}^d$, Y is modeled as

$$\{Y(\boldsymbol{t})\colon \boldsymbol{t} \in \Omega\} \overset{\mathcal{D}}{=} \{\sigma Z(\alpha \boldsymbol{t})\colon \boldsymbol{t} \in \Omega\} \tag{1}$$

where $\overset{\mathcal{D}}{=}$ denotes equality of the finite dimensional distributions. In this case, σ is an overall amplitude (in units of Y) and α is an overall spatial scale (in units of $\boldsymbol{t}$). For a nice discussion of the roll of α and σ in the Matérn autocovariance see Section 6.5 in [28].

A fundamental question is whether or not α and σ are consistently estimable when the number of the observations in Ω grows to infinity. Indeed, the answer is no in general. This is immediate from the existence of self similar random fields that satisfy $\{Z(\alpha \boldsymbol{t})\colon \boldsymbol{t} \in \Omega\} \overset{\mathcal{D}}{=} \{\alpha^\nu Z(\boldsymbol{t})\colon \boldsymbol{t} \in \Omega\}$ for any $\alpha > 0$ where ν is a fixed constant. For these self-similar processes, any two pairs (σ_1, α_1) and (σ_2, α_2) that satisfy $\sigma_1^2 \alpha_1^{2\nu} = \sigma_2^2 \alpha_2^{2\nu}$ give the same model in (1). This problem can also be present when Z is not self similar. For example, suppose Z is an isotropic Ornstein-Uhlenbeck process in dimension $d \leq 3$ (see Figure 1). In this case, if $\sigma_1^2 \alpha_1 = \sigma_2^2 \alpha_2$ (i.e. $\nu = 1/2$) the two models for Y yield mutually absolutely continuous measures (when $d = 1$ see [19], [32], when $d = 2, 3$ see [33], [28]) and therefore are impossible to discern with probability one when observing one realization of Y. We shall see, however, that in some cases it is possible to consistently estimate α and σ. Moreover, it will depend on dimension: typically the larger the dimension the more information there is to separate σ from α. Before we continue, we mention the work of Stein (see [25],[26]) which establishes that even if two models are mutually absolutely continuous, using the wrong model to make predictions may still yield asymptotically optimal estimates. In fact, this phenomenon can also occur for orthogonal measures when restricting to predictors that are linear combinations of the observations (see [27]).

To understand the condition $\sigma_1^2 \alpha_1^{2\nu} = \sigma_2^2 \alpha_2^{2\nu}$ one can look at what is called the principle irregular term of the autocovariance function (see [28]). Suppose, for exposition, that there exists constants $\delta_2 > \delta_1 > 0$ such that the covariance structure of Z satisfies

$$\mathrm{cov}(Z(\boldsymbol{t} + \boldsymbol{h}), Z(\boldsymbol{t})) \approx c_1 |\boldsymbol{h}|^{\delta_1} + c_2 |\boldsymbol{h}|^{\delta_2} + p(|\boldsymbol{h}|), \quad \text{as } |\boldsymbol{h}| \to 0 \tag{2}$$

where p is an even polynomial and both δ_1, δ_2 are not even integers. This model is not as restrictive as it seems and includes the Ornstein-Uhlenbeck process, the exponential autocovariance function $e^{-|\boldsymbol{s}-\boldsymbol{t}|^{\delta_1}}$ and the Matérn autocovariance function (see below). The term $c_1 |\boldsymbol{h}|^{\delta_1}$ is often referred to as the principle irregular term and is instrumental in determining the smoothness of Z. The second term, $c_2 |\boldsymbol{h}|^{\delta_2}$, is less influential but can have an observable effect depending on dimension and the magnitude of $\delta_2 - \delta_1$. Now, if we model Y by (1) and (2) we get

$$\mathrm{cov}(Y(\boldsymbol{t} + \boldsymbol{h}), Y(\boldsymbol{t})) \approx c_1 \sigma^2 \alpha^{\delta_1} |\boldsymbol{h}|^{\delta_1} + c_2 \sigma^2 \alpha^{\delta_2} |\boldsymbol{h}|^{\delta_2} + \tilde{p}(|\boldsymbol{h}|), \quad \text{as } |\boldsymbol{h}| \to 0. \tag{3}$$

Therefore for two pairs of parameters (σ_1, α_1) and (σ_2, α_2), the condition $\sigma_1^2 \alpha_1^{\delta_1} = \sigma_2^2 \alpha_2^{\delta_1}$ ensures that the covariance models for Y have the same principle irregular term. This

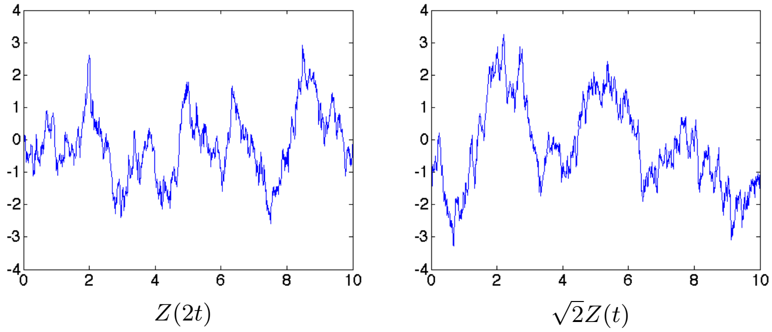

Fig 1: Independent simulations of $Z(2t)$ and $\sqrt{2}Z(t)$, observed on a dense grid in $[0, 10]$, where Z is the Ornstein-Uhlenbeck process with covariance structure $\mathrm{cov}(Z(s), Z(t)) = e^{-|s-t|}$. In 1, 2 and 3 dimensions these two processes (isotropically extended) are mutually absolutely continuous and therefore cannot be consistently distinguished under fixed domain asymptotics. Our results establish that when the dimension is greater than 4 one can distinguish the two with probability one under fixed domain asymptotics.

explains the importance of the quantity $\sigma^2\alpha^{\delta_1}$. In addition, if one can estimate both coefficients $c_1\sigma^2\alpha^{\delta_1}$ and $c_2\sigma^2\alpha^{\delta_2}$ then it is possible to get separate estimates of σ and α. In what follows we develop consistent estimators of these two coefficients which allow consistent estimation of σ and α.

The majority of this paper focuses on the case when Z is a mean zero, isotropic Gaussian random field which has a Matérn autocovariance. The reasons are twofold. First, the Matérn autocovariance has been used extensively in spatial statistics so that results on the Matérn autocovariance are of intrinsic interest alone. The second reason is that once one establishes the results for the Matérn it is relatively easy to see how to extend to other covariance functions. In Section 3, we give two examples that illustrate these extensions. Our Matérn assumption stipulates the existence of a known $\nu > 0$ such that

$$\mathrm{cov}(Z(\boldsymbol{s}), Z(\boldsymbol{t})) = \frac{|\boldsymbol{s}-\boldsymbol{t}|^\nu \mathcal{K}_\nu(|\boldsymbol{s}-\boldsymbol{t}|)}{2^{\nu-1}\Gamma(\nu)} \tag{4}$$

for all $\boldsymbol{s}, \boldsymbol{t} \in \Omega \subset \mathbb{R}^d$ where $|\cdot|$ denotes Euclidean distance and $\mathcal{K}_\mu$ is the modified Bessel function of the second kind of order $\nu > 0$ (see [1]). The parameter ν controls the mean square smoothness of the process: larger ν corresponds to smoother Z. The flexibility provided by the smoothness parameter ν along with the fact that it is positive definite in any dimension leads to its widespread use in spatial statistics.

In what follows we extend the basic model (1) to the case when there is an unknown invertible matrix M with determinant 1 (this class of matrices we denote by $SL(d,\mathbb{R})$) so that

$$\{Y(\boldsymbol{t})\colon \boldsymbol{t}\in\Omega\} \overset{\mathcal{D}}{=} \{\sigma Z(\alpha M\boldsymbol{t})\colon \boldsymbol{t}\in\Omega\}. \tag{5}$$

The matrix M is called a geometric anisotropy and is used to model a directional sheer of Z. The assumption that $\det M = 1$ removes identifiability problems with the overall scale parameter α. In Section 2, we construct estimates of $\sigma^2\alpha^{2\nu}$, M and α. We show that the estimates of $\sigma^2\alpha^{2\nu}$ and M are strongly consistent in any dimension and the estimate of α is strongly consistent when $d > 4$.

There is a fair amount of literature on estimating $\sigma^2\alpha^{2\nu}$ for the Matérn autocovariance. In 1991, Ying [32] established strong consistency and the asymptotic distribution of the maximum likelihood estimate of $\sigma^2\alpha^{2\nu}$ for the Ornstein-Uhlenbeck process when $d = 1$ (which has a Matérn autocovariance for $\nu = 1/2$). In 2004, Zhang [33] established that the maximum likelihood estimate of $\sigma^2\alpha^{2\nu}$ (obtained by fixing α and ν) is strongly consistent when $d \le 3$. In related work, Wei-Liem Loh [23] shows that maximum likelihood estimates of scale and variance parameters in a non-isotropic multiplicative Matérn model are consistent when $\nu = 3/2$ (similar results for the Gaussian autocovariance model can be found in [24]). In Section 6.7 of [28], Stein derives asymptotic properties of the maximum likelihood estimates of α, σ and ν for a periodic version of the Matérn random field. For this periodic random field all the parameters are consistently estimable when $d \ge 4$. Our results confirm these findings for α and σ with the non-periodic Matérn when $d > 4$. The case $d = 4$ is still open.

Recent work by Kaufman et al. [8] and Du et al. [13] studies maximum likelihood estimates of $\sigma^2\alpha^{2\nu}$ using a tapered Matérn autocovariance when $d \le 3$. The advantage gained by tapering is a reduction of the computational load for computing the likelihood and for computing kriging estimates. We will see that our estimates of the same quantity, $\sigma^2\alpha^{2\nu}$, yield strongly consistent estimates in any dimension which are “root n” consistent and are easily computed with no maximization required. However, our estimates depend on the grid format of the observations whereas the maximum likelihood estimates are not confined to such restrictions. We also expect some loss of efficiency in our estimates as compared to the MLE. We hope that there is potential to combine the two estimation methods using a one-Newton-step tapered likelihood adjustment to the increment based estimate. Since our results can be easily extended, by a Lindeberg-Feller argument, to obtain the asymptotic normality of $\widehat{\sigma^2\alpha^{2\nu}}$ when $d \le 3$, we believe this has the potential to mitigate any loss of efficiency and reduce the computational load for the maximum likelihood estimate.

Finally we mention the long tradition of using squared increments to estimate properties of random fields, beginning with the quadratic variation theorem of Lévy in 1940 ([22]). For example, increments have been used in [20] and [6] for identification of a local fractional index and in [11] to identify the singularity function of a fractional process. In [4] they are used to estimate a deformation of an isotropic random field. For more results on the convergence of quadratic variations see, [5], [15], [14], [21], [7], [30],[2], [6], [16], [10], [20].

2. The geometric anisotropic Matérn class. In this section we construct estimates of $\sigma^2\alpha^{2\nu}$, M and α using increments of Y observed on a dense grid within Ω. Using fixed domain asymptotics, we establish consistency of our estimates under assumptions (4) and (5) and provide bounds on the rate of variance decay as it depends on the number of increments used, the dimension of Ω and the smoothness of Y measured by ν. These results will hold in any dimension. However, when the dimension is large enough ($d > 4$), the second term in (3) is influential enough so that α can be estimated consistently.

If the observation region Ω is an open subset of $\mathbb{R}^d$ and the random field Y is modeled by (4) and (5), then Y is said to be a d-dimensional geometric anisotropic Matérn random field with parameters (σ, α, ν, M). In this case, the covariance structure of Y is $\mathrm{cov}(Y(\boldsymbol{s}), Y(\boldsymbol{t})) = K(|M\boldsymbol{s} - M\boldsymbol{t}|)$, where K is defined as

$$K(t) \triangleq \frac{\sigma^2(\alpha t)^\nu}{\Gamma(\nu)2^{\nu-1}}\mathcal{K}_\nu(\alpha t) \tag{6}$$

for $t > 0$ and $K(0) \triangleq \lim_{t\downarrow 0} K(t) = \sigma^2$. The function $\mathcal{K}_\nu$ is the modified Bessel function of the second kind of order $\nu > 0$. Since $|M\boldsymbol{s} - M\boldsymbol{t}| = |OM\boldsymbol{s} - OM\boldsymbol{t}|$ for any orthogonal matrix O, one can only identify M up to left multiplication by an orthogonal matrix. To remove this identifiability problem we suppose that $M \in SL(d,\mathbb{R})/SO(d,\mathbb{R})$ where $SO(n,\mathbb{R})$ denotes the orthogonal matrices in $SL(d,\mathbb{R})$. In the theorems below, we write $M_1 =_{SL/SO} M_2$ to mean that there exists a $O \in SO(n,\mathbb{R})$ such that $M_1 = OM_2$, and similarly for $M_1 \neq_{SL/SO} M_2$. Operationally, however, we estimate a representer of the cosets in $SL(d,\mathbb{R})/SO(d,\mathbb{R})$ given by the upper triangular matrices which have positive diagonal elements and determinant 1 (that this is a representer follows from the QR factorization, see [18]).

As discussed in the introduction, the principle irregular term is important in determining the sample path properties of the random field Y. The principle irregular term for the Matérn covariance function is

$$G_\nu(t) \triangleq \begin{cases} \dfrac{(-1)^{\nu+1}}{2^{2\nu-1}\Gamma(\nu)\Gamma(\nu+1)} t^{2\nu}\log t, & \text{if } \nu \in \mathbb{Z}; \\ \dfrac{-\pi}{2^{2\nu}\sin(\nu\pi)\Gamma(\nu)\Gamma(\nu+1)} t^{2\nu}, & \text{otherwise.} \end{cases}$$

where $G_\nu(0)$ is defined to be 0. Moreover,

$$\mathrm{cov}(Y(\boldsymbol{t}+\boldsymbol{h}), Y(\boldsymbol{t})) = \sigma^2 G_\nu(|\alpha M\boldsymbol{h}|) - \nu\sigma^2 G_{\nu+1}(|\alpha M\boldsymbol{h}|) + \epsilon(|\alpha M\boldsymbol{h}|) \tag{7}$$

where $\epsilon(h) = \sigma^2 p(|h|) + o(G_{\nu+1}(|h|))$ as $|h| \to 0$ and p is an even polynomial. Notice that when M is the identity matrix and $\nu \notin \mathbb{Z}$, this gives the expansion (3) so that $\sigma^2 G_\nu(|\alpha\boldsymbol{h}|)$ is the first principle irregular term and $-\nu\sigma^2 G_{\nu+1}(|\alpha\boldsymbol{h}|)$ is the second term.

2.1. *Estimating $\sigma^2\alpha^{2\nu}$ and M in any dimension.* Let Ω be a bounded, open subset of $\mathbb{R}^d$ and let $\Omega_n \triangleq \Omega \cap \{\mathbb{Z}^d/n\}$. The idea is that we will be observing Y on a region, just a bit larger than Ω_n, so that we can form the m^{th} order increments of Y on Ω_n. These will then be used to estimate M and $\sigma^2\alpha^{2\nu}$ in any dimension and additionally α, in dimension $d > 4$.

For a fixed nonzero vector $\boldsymbol{h} \in \mathbb{R}^d$ define the increment in the direction $\boldsymbol{h}$ by $\Delta_{\boldsymbol{h}} Y(\boldsymbol{t}) \triangleq Y(\boldsymbol{t}+\boldsymbol{h}) - Y(\boldsymbol{t})$ and the m^{th} iterated directional increment $\Delta^m_{\boldsymbol{h}} Y(\boldsymbol{t}) \triangleq \Delta_{\boldsymbol{h}}\Delta^{m-1}_{\boldsymbol{h}} Y(\boldsymbol{t})$. The following lemma establishes the relationship between the variance of these increments and the terms in (7) when the number of increments is sufficiently large.

LEMMA 1. *Let Y be a mean zero, geometric anisotropic d-dimensional Matérn Gaussian random field with parameters (σ,α,ν,M). If m is a positive integer such that $m > \nu + 1$ and $\boldsymbol{h} \in \mathbb{R}^d$ is a non-zero vector, then*

$$\mathsf{E}(\Delta^m_{\boldsymbol{h}/n} Y(\boldsymbol{t}))^2 = \frac{a^m_\nu}{n^{2\nu}} + \frac{b^m_\nu}{n^{2\nu+2}} + o(n^{-2\nu-2}) \tag{8}$$

as $n \to \infty$ where

$$a^m_\nu \triangleq \sigma^2\alpha^{2\nu}|M\boldsymbol{h}|^{2\nu} \sum_{i,j=0}^{m} (-1)^{i+j} \binom{m}{i}\binom{m}{j} G_\nu(|i-j|) \tag{9}$$

$$b^m_\nu \triangleq \sigma^2\alpha^{2\nu+2}|M\boldsymbol{h}|^{2\nu+2} \sum_{i,j=0}^{m} (-1)^{i+j} \binom{m}{i}\binom{m}{j} (-\nu) G_{\nu+1}(|i-j|). \tag{10}$$

Now we are in a position to estimate the coefficient a^m_ν. Let $\#\Omega_n$ denote the cardinality of the finite set $\Omega_n \triangleq \Omega \cap \{\mathbb{Z}^d/n\}$ and define

$$Q^m_n \triangleq \frac{1}{\#\Omega_n} \sum_{\boldsymbol{j}\in\Omega_n} n^{2\nu} (\Delta^m_{\boldsymbol{h}/n} Y(\boldsymbol{j}))^2 \tag{11}$$

Notice that by equation (8), $\mathsf{E}Q^m_n \to a^m_\nu$ as $n \to \infty$. In addition, since Q^m_n is itself an average, one might hope that Q^m_n converges to a^m_ν. The following theorem shows that, indeed, this is the case. In addition, the theorem quantifies the decay of the variance of Q^m_n as a function of the number of increments, the smoothness of the random field Y and the dimension of the domain. The heuristic is that when the number of increments m is large enough, there is sufficient decorrelation of the summands of Q^m_n to guarantee convergence as $n \to \infty$. Generally, more increments leads to more spatial decorrelation and hence a reduction in variance. However, this only holds up to a point, after which taking more increments no longer effects the rate of variance decay. Finally, the higher the dimension, the more increments one needs to take to get the best rate.

THEOREM 1. *Let Y be a mean zero, geometric anisotropic d-dimensional Matérn Gaussian random field with parameters (σ, α, ν, M) and let Ω be a bounded, open subset of $\mathbb{R}^d$. If $m > \nu$ then*

$$Q_n^m \to a_\nu^m, \quad w.p.1 \tag{12}$$

as $n \to \infty$. Moreover, there exists a constant $c > 0$ such that

$$var\, Q_n^m \leq \begin{cases} c\, n^{4(\nu-m)}, & \text{if } 4(\nu - m) > -d; \\ c\, n^{-d} \log n, & \text{if } 4(\nu - m) = -d; \\ c\, n^{-d}, & \text{if } 4(\nu - m) < -d \end{cases}$$

for all sufficiently large n.

The above theorem establishes that Q_n^m consistently estimates a_ν^m (which depends on $\boldsymbol{h}$). Now we show how these estimates can be used to recover M and $\sigma^2\alpha^{2\nu}$. As was mentioned above, we suppose M is upper triangular with determinant one and positive diagonal elements. After renormalizing by known constants, the values of a_ν^m allow us to consistently estimate $|\tilde{M}\boldsymbol{h}|^2$ where $\tilde{M} \triangleq \sigma^{1/\nu}\alpha M$ for finitely many directions $\boldsymbol{h}$. We show by induction that these values are sufficient to recover each column of $\tilde{M}$. Once this is established, the requirement $\det M = 1$ gives $M = (\det \tilde{M})^{-1/d}\tilde{M}$ and $\sigma^2\alpha^{2\nu} = (\det \tilde{M})^{2\nu/d}$.

Let $\tilde{M}_{i,j}$ denote the i, j^{th} element of $\tilde{M}$ and let $\tilde{M}_{:,i}$ denote the i^{th} column of $\tilde{M}$. Also let $\tilde{M}_{1:k,1:k}$ be the submatrix with elements $\tilde{M}_{i,j}$ for $i, j = 1, \ldots, k$. For the first column of $\tilde{M}$, notice that $|\tilde{M}e_1| = \tilde{M}_{1,1}$ where $e_1, \ldots, e_d$ denote the standard basis of $\mathbb{R}^d$. This follows since $\tilde{M}$ is upper triangular with positive diagonal. For the inductive step suppose the first k columns $\tilde{M}_{:,1} \ldots, \tilde{M}_{:,k}$ are known. Taking $\boldsymbol{h} = e_{k+1}$ and $\boldsymbol{h} = e_{k+1} - e_i$ allows us to recover $|\tilde{M}_{:,k+1}|^2$ and $|\tilde{M}_{:,k+1} - \tilde{M}_{:,i}|^2$ for $i = 1, \ldots, k$. By adding and subtracting appropriate terms we can then recover: $\langle \tilde{M}_{:,k+1}, \tilde{M}_{:,i}\rangle$, for all $i = 1, \ldots, k+1$. Therefore $\tilde{M}_{:,k+1} = \Big(v, \sqrt{|\tilde{M}_{:,k+1}|^2 - |v|^2}, 0, \ldots, 0\Big)^T$ where $v \triangleq \tilde{M}^{-1}_{1:k,1:k}\big(\langle \tilde{M}_{:,k+1}, \tilde{M}_{:,i}\rangle\big)_{i=1}^k$. This establishes the inductive step and therefore $\tilde{M}$ can be identified from observing $|\tilde{M}\boldsymbol{h}|^2$ at $d(d+1)/2$ different vectors $\boldsymbol{h}$ (let them be denoted by $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_{d(d+1)/2}$).

Notice that as $\tilde{M}$ ranges over the set of upper triangular matrices with positive diagonal, the transformation $\{|\tilde{M}\boldsymbol{h}| \colon \boldsymbol{h} = \boldsymbol{h}_1 \ldots, \boldsymbol{h}_{d(d+1)/2}\} \xrightarrow{f_1} \tilde{M} \xrightarrow{f_2} (M, \sigma^2\alpha^{2\nu})$ sends an open subset of $\mathbb{R}^{d(d+1)/2}$ to $SL(d, \mathbb{R}) \times \mathbb{R}^+$. Since $f_2 \circ f_1$ is a continuous map,

$$(\widehat{\sigma^2\alpha^{2\nu}}, \widehat{M}) \to (\sigma^2\alpha^{2\nu}, M), \quad w.p.1$$

as $n \to \infty$.

2.2. *Estimating α, when $d > 4$.* In this section we construct an estimate of $\sigma^2\alpha^{2\nu+2}|M\boldsymbol{h}|^{2\nu+2}$ when $d > 4$, which, in combination with M and $\sigma^2\alpha^{2\nu}$, allows us to consistently estimate α. We start by noticing that by Lemma 1, for any $p, q > \nu + 1$

$$\mathsf{E}n^2\left[Q_n^p - \frac{a_\nu^p}{a_\nu^q}Q_n^q\right] \to \left[b_\nu^p - \frac{a_\nu^p}{a_\nu^q}b_\nu^q\right]$$

as $n \to \infty$. The term $b_\nu^p - \frac{a_\nu^p}{a_\nu^q}b_\nu^q$ is significant because, for any positive integer p, q

$$b_\nu^p - \frac{a_\nu^p}{a_\nu^q}b_\nu^q = c\,\sigma^2\alpha^{2\nu+2}|M\boldsymbol{h}|^{2\nu+2}$$

where $0 \leq c \leq \infty$ is a known constant depending on p and q. In addition, Lemma 2 in the Appendix establishes that $c \neq 0$ and $c \neq \infty$ for at least one $p, q > \nu + 1$. Moreover, a_ν^p/a_ν^q doesn't depend on the unknown parameters σ^2, α and M and therefore one can construct $n^2\left[Q_n^p - \frac{a_\nu^p}{a_\nu^q}Q_n^q\right]$ from the observed values of the random field Y. The following theorem quantifies how large p and q need to be for the almost sure convergence of $n^2\left[Q_n^p - \frac{a_\nu^p}{a_\nu^q}Q_n^q\right]$ to $b_\nu^p - \frac{a_\nu^p}{a_\nu^q}b_\nu^q$.

THEOREM 2. *Let Y be a mean zero, geometric anisotropic d-dimensional Matérn Gaussian random field with parameters (σ, α, ν, M) and let Ω be a bounded, open subset of $\mathbb{R}^d$. Suppose $p \neq q$ are positive integers such that $p, q > \nu + 1$ and both are large enough so that $4 < \min\{2p - 2\nu, d\}$ and $4 < \min\{2q - 2\nu, d\}$. Then*

$$n^2\left[Q_n^p - \frac{a_\nu^p}{a_\nu^q}Q_n^q\right] \to \left[b_\nu^p - \frac{a_\nu^p}{a_\nu^q}b_\nu^q\right], \quad w.p.1$$

as $n \to \infty$.

Theorems 1 and 2 show that there exists strongly consistent estimates of $\sigma^2\alpha^{2\nu}$, M and $\sigma^2\alpha^{2\nu+2}|M\boldsymbol{h}|^{2\nu+2}$. This, in turn, gives consistent estimates of α, σ and M. Notice that when $d \leq 3$ this is impossible due to the mutual absolute continuity of Matérn Gaussian random fields with different scale and variance parameters (see Zhang [33]). Since Gaussian measures are either mutually absolutely continuous or orthogonal, the fact that we have strongly consistent estimates of α, σ and M gives the following corollary.

COROLLARY 3. *Let Y_1 and Y_2 be two, mean zero, geometric anisotropic d-dimensional Matérn Gaussian random fields defined a bounded open set $\Omega \subset \mathbb{R}^d$ with parameters $(\sigma_1, \alpha_1, \nu, M_1)$ and $(\sigma_2, \alpha_2, \nu, M_2)$ where $d > 4$. If $(\sigma_1, \alpha_1) \neq (\sigma_2, \alpha_2)$ or $M_1 \neq_{SL/SO} M_2$ then the Gaussian measures induced by the random fields Y_1 and Y_2 are orthogonal.*

Remark: The strong consistency results for our estimates of $\sigma^2\alpha^{2\nu}$, α and M all depend on knowledge of the true value of ν. However, our results can be extended when using an estimate $\hat{\nu}$ so long as the error $\epsilon_n \triangleq \hat{\nu} - \nu$ satisfies $\epsilon_n \log n \to 0$ with probability one as $n \to \infty$. This follows since the ratio of the quadratic variation, Q_n^m, using the true ν, to the quadratic variation using the estimated $\hat{\nu}$, is $n^{-\epsilon_n}$ which converges to 1 if $\epsilon_n \log n \to 0$.

3. Beyond the Matérn. The previous section dealt exclusively with the Matérn autocovariance. Now we show how these results can be extended to other autocovariance functions. We choose two examples to illustrate how the methodology can be easily extended beyond the Matérn autocovariance function. The key components for showing extensions are establishing versions of Lemmas 1 and 4. Lemma 1 quantifies the expected value of the squared increments $(\Delta^p_{\boldsymbol{h}/n} Y(\boldsymbol{t}))^2$ in terms of n. Lemma 4 establishes that, in effect, derivatives of the covariance away from the origin are dominated by the derivatives of the principle irregular term. Once the analogs of these Lemmas are established all the subsequent arguments for versions of Theorems 1 and 2 follow almost immediately.

For our first example we consider the case when Y is a mean zero Gaussian random field on $\mathbb{R}^d$ with generalized autocovariance function $c_1|t|^{\delta_1} + c_2|t|^{\delta_2}$ where δ_1 and δ_2 are known but c_1 and c_2 are unknown (it is tacitly assumed that the values of c_1 and c_2 give a conditionally positive definite function of order $\lfloor \delta_2/2 \rfloor$ in $\mathbb{R}^d$, see [9]). In what follows we suppose $\delta_2 > \delta_1 > 0$ and neither are even integers. The appropriate version of Lemma 1 says that when $p > \delta_2/2$

$$\mathsf{E}(\Delta^p_{\boldsymbol{h}/n} Y(\boldsymbol{t}))^2 = \frac{c_1 C_{p,\delta_1}}{n^{\delta_1}} + \frac{c_2 C_{p,\delta_2}}{n^{\delta_2}} \tag{13}$$

where $C_{p,\delta} \triangleq |\boldsymbol{h}|^\delta \sum_{i,j=0}^p (-1)^{i+j}\binom{p}{i}\binom{p}{j}|i-j|^\delta$. Now Q_n^p is defined as in (11) with δ_1 in place of 2ν. In this case, $\mathsf{E}Q_n^p = c_1 C_{p,\delta_1} + c_2 C_{p,\delta_2} n^{\delta_1-\delta_2}$ and therefore we set $\hat{c}_1 \triangleq Q_n^p / C_{p,\delta_1}$. Also, for an integer $q > p$ we have $\mathsf{E} n^{\delta_2-\delta_1}\big[Q_n^p - \frac{C_{p,\delta_1}}{C_{q,\delta_1}} Q_n^q\big] = c_2\big[C_{p,\delta_1} - \frac{C_{p,\delta_1}}{C_{q,\delta_1}} C_{q,\delta_2}\big]$ and after a renormalization one gets the estimate $\hat{c}_2$. The analog to Lemma 4 says that when $p > \delta_2/2$ and Ω is a bounded open subset of $\mathbb{R}^d$ there exists a constant $c > 0$ such that

$$\big|\partial^{(p,p)}_{\boldsymbol{h}} \operatorname{cov}(Y(\boldsymbol{s}), Y(\boldsymbol{t}))\big| \leq c|\boldsymbol{s} - \boldsymbol{t}|^{\delta_1 - 2p} \tag{14}$$

for all $\boldsymbol{s}, \boldsymbol{t} \in \Omega$ such that $\boldsymbol{s} \neq \boldsymbol{t}$. Once (13) and (14) are established, versions of Lemma 6, Lemma 7, Lemma 8 and Theorem 1 following by replacing 2ν with δ_1. To establish Theorem 2, replace the n^2 term with $n^{\delta_2-\delta_1}$ in equation (47) and continue in an similar manner to establish the following theorem.

THEOREM 4. *Suppose Y is a mean zero Gaussian random field on $\mathbb{R}^d$ with generalized autocovariance function $c_1|t|^{\delta_1} + c_2|t|^{\delta_2}$ observed on $\Omega \cap \{\mathbb{Z}^d/n\}$ where Ω is a bounded open subset of $\mathbb{R}^d$ and $0 < \delta_1 < \delta_2$ are known and not even integers. If $0 < 2(\delta_2 - \delta_1) < d$ then* *there exists integers $q > p > 0$ such that $\hat{c}_1$ and $\hat{c}_2$ (defined above) converge with probability one to c_1 and c_2 (respectively) as $n \to \infty$.*

There are different conditions on p to guarantee convergence of c_1 versus c_2. Generally, one only needs $p > \delta_1/2$ for consistent estimation of c_1, which will hold in any dimension. However, in our case, we need the additional requirement that $p > \delta_2/2$ since we are working with a conditionally positive definite function of order $\lfloor \delta_2/2 \rfloor$. To get consistent estimation of c_2 we need the additional inequality $2(\delta_2 - \delta_1) < \min\{2p - \delta_1, d\}$. To relate this to our Matérn results in Section 2 set $\delta_1 = 2\nu$ and $\delta_2 = 2\nu + 2$ so that the inequality becomes $4 < \min\{2p - 2\nu, d\}$ which appears in Theorem 2. Finally the analog to Lemma 2 guarantees there exits a $q > p$ such that $\left[C_{p,\delta_1} - \frac{C_{p,\delta_1}}{C_{q,\delta_1}} C_{q,\delta_2}\right]$ is non-zero which allows us to define $\hat{c}_2$.

Before we continue, we mention a comment in Wahba's book ([31], page 44) which argues in favor of using the generalized autocovariance $|t|^{2m-1}$ over the model $|t|^{2m-1} + c_1|t|^{2m+1} + \cdots + c_k|t|^{2m+2k-1}$ when $d = 1, 2, 3$. The reasoning is that the two models yield mutually absolutely continuous Gaussian measures, and therefore can not be consistently distinguished. We can see, however, that the dimension requirement $d = 1, 2, 3$ is an integral component of this argument. When the dimension gets above 4, this reasoning no longer holds since the two models are orthogonal by the above theorem (setting $\delta_1 = 2m - 1$ and $\delta_2 = 2m + 1$).

For our second extension we show that the variance σ^2 and scale α can be separately estimated in the exponential autocovariance model $\sigma^2 e^{-|\alpha t|^\delta}$ when the dimension $d > 2\delta$ and $\delta \neq 1$. In this case, the appropriate version of Lemma 1 becomes

$$\mathsf{E}(\Delta^p_{\boldsymbol{h}/n} Y(\boldsymbol{t}))^2 = -\frac{\sigma^2 \alpha^\delta C_{p,\delta}}{n^\delta} + \frac{\sigma^2 \alpha^{2\delta} C_{p,2\delta}}{2n^{2\delta}} + O(n^{-3\delta}) \tag{15}$$

as $n \to \infty$ when $p > \delta/2$. From (15) one can now easily construct estimates of $\sigma^2\alpha^\delta$ and $\sigma^2\alpha^{2\delta}$. When a geometric anisotropy M is present, the techniques of Section 2 are also sufficient to also construct $\widehat{M}$. Notice that by direct differentiation, equation (14) holds when δ_1 is replaced by δ. Using similar arguments for the previous theorem and extending to a geometric anisotropy the following theorem is obtained.

THEOREM 5. *Let Y be a mean zero, Gaussian process on $\mathbb{R}^d$ with autocovariance function $\sigma^2 e^{-|\alpha M \boldsymbol{t}|^\delta}$ observed on $\Omega \cap \{\mathbb{Z}^d/n\}$ where Ω is a bounded open subset of $\mathbb{R}^d$. Suppose $\delta \in (0, 2)$ is known, σ and α are positive and M is upper triangular with positive diagonal and determinant 1. If $p \geq 1$ then $\widehat{\sigma^2\alpha^\delta} \to \sigma^2\alpha^\delta$ and $\widehat{M} \to M$ with probability one as $n \to \infty$. Moreover, if $2\delta < d$ and $\delta \neq 1$ then for any $p > 3\delta/2$ there exists $q > p$ such that $\hat{\sigma} \to \sigma$ and $\hat{\alpha} \to \alpha$ with probability one as $n \to \infty$.*

Many other extensions are possible, including more general non-stationary random fields. In this case, both a^m_ν and b^m_ν depend on $\boldsymbol{t} \in \Omega$ and Q^p_n will convergence to $\int_\Omega a^m_\nu d\boldsymbol{t}$ and similarly for $\int_\Omega \left[b_\nu^p - \frac{a_\nu^p}{a_\nu^q} b_\nu^q\right] d\boldsymbol{t}$. If one also needs pointwise convergence to a_ν^m or $b_\nu^p - \frac{a_\nu^p}{a_\nu^q} b_\nu^q$ one can consider weighted local averaging of the terms in Q_n^p. This was the technique used in [4] when observing a deformed isotropic Gaussian random field that locally behaved like a fractional Brownian field. However, obtaining extensions in these cases are more difficult since one needs to consider rates of decay for a bandwidth parameter. That being said, this work leaves open the possibility of constructing consistent estimates of the two deformations f_1, f_2 when observing $Y_1 \circ f_1 + Y_2 \circ f_2$ where Y_1 and Y_2 have generalized autocovariance functions $|t|^{\delta_1}$ and $|t|^{\delta_2}$ respectively. Finally we mention that since Q_n^p is constructed from increments, one can extend our results to random fields Y with a polynomial drift of known order.

4. Simulations. We finish with two simulations that illustrate (and hopefully compliment) our theoretical results. The first simulation shows how one can use directional increments to estimate $\sigma^2\alpha^{2\nu}$ and a geometric anisotropy M using finitely many directions. The second simulation shows how to estimate the coefficient on the 'second principle irregular term' (c_2 in equation (2)) and how it can be used to construct an unbiased estimate of the coefficient on the 'first principle irregular term' (c_1 in equation (2)).

In our first example, we simulated 500 independent realizations of a Matérn random field with parameters $\sigma = 1.5$, $\alpha = 0.8$, $\nu = 1.75$, $M(1,1) = 1.2$, $M(1,2) = 0.5$, $M(2,1) = 0$ and $M(2,2) = 1/1.2$ observed on a square grid in $[0,1]^2$ with spacing $1/55$. On each realization we estimated $\sigma^2\alpha^{2\nu}$ and M using 2, 3 and 4 horizontal, vertical and diagonal increments. Notice that since $1 < \nu < 2$, this random field is once, but not twice, mean square differentiable. Intuitively, we therefore need at least two increments for sufficient de-correlation of the terms in the quadratic variation sum (2). Table 1 displays the root mean squared error (RMSE) for estimating $\sigma^2\alpha^{2\nu}$, the true value is approximately 1.03, and the elements of M. Figure 2 plots histograms of the estimates for 2 and 3 increments. It is immediately clear that there is a large reduction in RMSE when using 3 increments as compared to 2 increments (and an additional bias reduction when estimating $\sigma^2\alpha^{2\nu}$). Indeed, by Theorem 1, more increments leads to more spatial decorrelation and hence a reduction in variance. In this case, $\nu < 2 < \nu + 1$ so that the estimate based on 2 increments is guaranteed to be consistent but the variance decays at a sub-optimal rate. Since $3 > (4\nu + d)/4 = 2.25$, the variance of the estimate based on 3 increments decays at the optimal rate. However, Theorem 1 also says that this variance reduction only holds up to a point, after which taking more increments no longer effects the rate of variance decay. Indeed, it is seen in Table 1 that taking 4 increments do not improve the RMSE nearly as much.

Our second simulation uses the results of Section 3 to estimate c_1 and c_2 when observing $\sqrt{c_1}\, Y_1 + \sqrt{c_2}\, Y_2$ on $[0, 1/\sqrt{2})^2$ at 1000×1000 pixel locations where $c_1 = 100$, $c_2 = 36$ and Y_1 is independent of Y_2. The random field Y_1 has autocovariance $\frac{9}{10} - |t|^{0.2} + \frac{1}{10}|t|^2$ and Y_2 has autocovariance $\frac{8}{10} - |t|^{0.4} + \frac{2}{10}|t|^2$ which is positive definite on $[0, 1/\sqrt{2})^2$ (see [29] for a

TABLE 1
RMSE for estimating $\sigma^2\alpha^{2\nu}$ and M using 2, 3 and 4 increments.

	2 increments	3 increments	4 increments
$\sigma^2\alpha^{2\nu}$	0.1664	0.0300	0.0289
$M(1,1)$	0.0360	0.0114	0.0113
$M(1,2)$	0.0475	0.0147	0.0147
$M(2,2)$	0.0248	0.0079	0.0079

proof). Our estimates of c_1 and c_2 are defined by

$$\hat{c}_1 \triangleq Q_n^p / C_{p,\delta_1} \tag{16}$$

$$\hat{c}_2 \triangleq n^{\delta_2-\delta_1} \frac{Q_n^p - \frac{C_{p,\delta_1}}{C_{q,\delta_1}} Q_n^q}{C_{p,\delta_2} - \frac{C_{p,\delta_1}}{C_{q,\delta_1}} C_{q,\delta_2}} \tag{17}$$

where $\delta_1 = 0.2$, $\delta_2 = 0.4$, $p = 2$, $q = 3$ and $C_{p,\delta} \triangleq -|\boldsymbol{h}|^\delta \sum_{i,j=0}^p (-1)^{i+j} \binom{p}{i}\binom{p}{j} |i-j|^\delta$. This example was chosen to illustrate the duality when estimating c_1 and c_2: the smaller $|\delta_1 - \delta_2|$ (in relation to the dimension d) the smaller the variance of $\hat{c}_1$ and $\hat{c}_2$ but the larger the bias of $\hat{c}_1$. In fact, as the dimension grows, the variance $\hat{c}_1$ decreases at a faster rate (proportional to n^{-d} when using enough increments) but the bias decreases at the same asymptotic rate for any d (proportional to $n^{\delta_1-\delta_2}$). In our example, since $p = 2$ (so the quadratic term $\frac{1}{10}|t|^2$ vanishes), we can explicitly compute the bias using equation (13) so that $\mathsf{E}\hat{c}_1 = c_1 + c_2 \frac{C_{p,\delta_2}}{C_{p,\delta_1}} n^{\delta_1-\delta_2}$. Notice that using our estimate of c_2 we can now correct the bias in $\hat{c}_1$. The left plot of Figure 3 shows two histograms of the estimate $\hat{c}_1$ and the bias corrected estimate $\hat{c}_1 - \hat{c}_2 \frac{C_{p,\delta_2}}{C_{p,\delta_1}} n^{\delta_1-\delta_2}$ on the 500 simulated realizations. The right plot of Figure 3 shows the histogram of the estimate $\hat{c}_2$. We can see that not only is it possible to get an estimate of c_2, but using it to correct the bias in $\hat{c}_1$ reduces the RMSE for estimating c_1 (from 7.84 down to 2.29).

APPENDIX A: PROOFS

We start with some notation. For a function of two variables $F(\boldsymbol{s},\boldsymbol{t})$ let $\Delta_{\boldsymbol{h}}^{(m,n)} F(\boldsymbol{s},\boldsymbol{t}) \triangleq \Delta_{\boldsymbol{h}}^m \Delta_{\boldsymbol{h}}^n F(\boldsymbol{s},\boldsymbol{t})$ where $\Delta_{\boldsymbol{h}}^m$ acts on the variable $\boldsymbol{s}$ and $\Delta_{\boldsymbol{h}}^n$ acts on the variable $\boldsymbol{t}$. Define $\partial_{\boldsymbol{h}} \triangleq \boldsymbol{h}\cdot\nabla$ to be the directional derivative in the direction $\boldsymbol{h}$ and $\partial_{\boldsymbol{h}}^{(m,n)} F(\boldsymbol{s},\boldsymbol{t}) \triangleq \partial_{\boldsymbol{h}}^m \partial_{\boldsymbol{h}}^n F(\boldsymbol{s},\boldsymbol{t})$ where $\partial_{\boldsymbol{h}}^m$ acts on the variable $\boldsymbol{s}$ and $\partial_{\boldsymbol{h}}^n$ acts on $\boldsymbol{t}$.

Let $f(\xi), g(\xi)$ be real valued functions defined on some set Ξ and let $\Xi' \subset \Xi$. We write $f(\xi) \lesssim g(\xi)$ for all $\xi \in \Xi'$ if there there exists a positive constant $c > 0$ such that $|f(\xi)| \leq c\, g(\xi)$ for all $\xi \in \Xi'$. Notice that this definition also works for a sequence of functions f_n, g_n by considering the variable n as an argument and replacing Ξ by $\Xi \times \mathbb{N}$.

Proof of Lemma 1. We suppose $\sigma = \alpha = 1$ and M is the identity matrix, then rescale for the general case. First note two immediate facts about the m^{th} directional increment

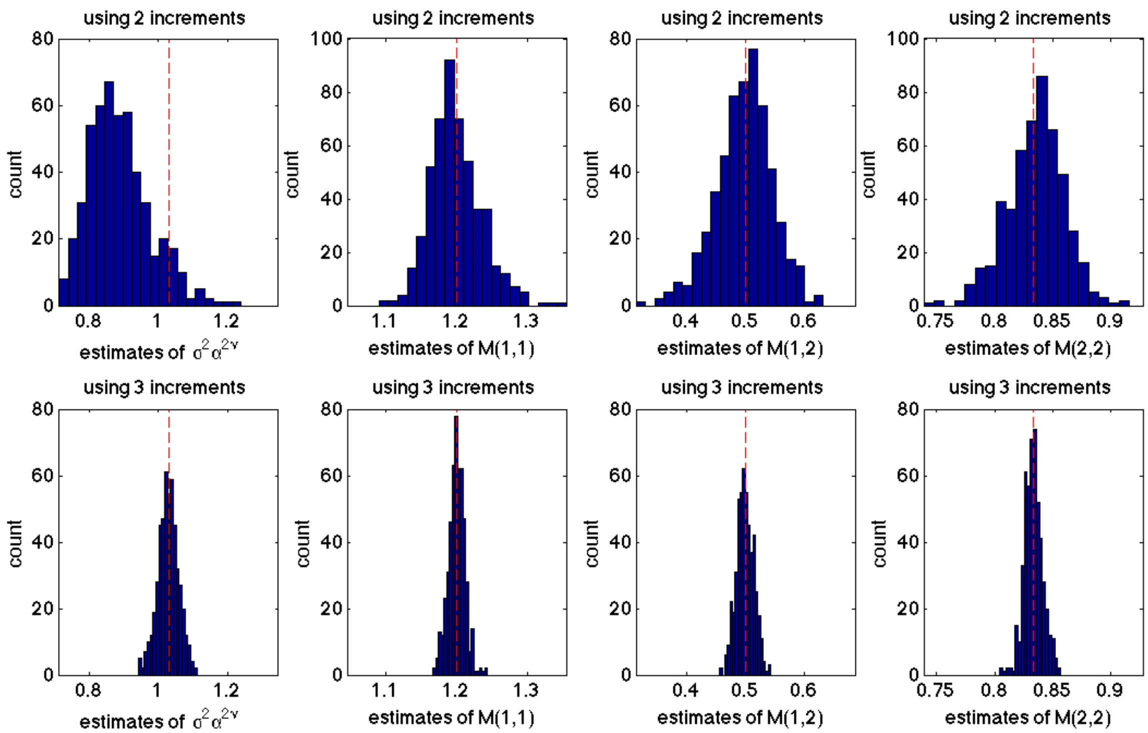

Fig 2: 500 independent simulations of a Matérn random field with $\sigma = 1.5$, $\alpha = 0.8$, $\nu = 1.75$, $M(1,1) = 1.2$, $M(1,2) = 0.5$, $M(2,1) = 0$ and $M(2,2) = 1/1.2$ observed on a square grid in $[0,1]^2$ with spacing 1/55. The top row of figures shows the histograms of the estimates of $(\sigma^2\alpha^{2\nu}, M(1,1), M(1,2), M(2,2))$ using the techniques derived in Section 2.1 based on increments of order 2. The bottom row shows the histograms of the estimates using increments of order 3.

operator $\Delta^m_{\boldsymbol{h}/n}$: for any function $f\colon \mathbb{R}^d \to \mathbb{R}$ the m^{th}-increment of f can be computed $\Delta^m_{\boldsymbol{h}/n} f(\boldsymbol{t}) = \sum_{i=0}^m d_i f(\boldsymbol{t} + i\boldsymbol{h}/n)$ where $d_i = (-1)^{m+i}\binom{m}{i}$; The m^{th}-increment $\Delta^m_{\boldsymbol{h}/n}$ annihilates monomials of degree less than m so that $\Delta^{(m,m)}_{\boldsymbol{h}/n} |\boldsymbol{t} - \boldsymbol{s}|^{2k} = 0$ for all $k = 0, \ldots, m-1$. Therefore, by the expansions given on page 375 of [1] we have

$$\Delta^{(m,m)}_{\boldsymbol{h}/n} K(|\boldsymbol{s} - \boldsymbol{t}|) = \Delta^{(m,m)}_{\boldsymbol{h}/n} \Big\{ G_\nu(|\boldsymbol{s} - \boldsymbol{t}|) - \nu G_{\nu+1}(|\boldsymbol{s} - \boldsymbol{t}|) + r(|\boldsymbol{s} - \boldsymbol{t}|) \Big\}$$

where $r(\epsilon) = o(\epsilon^{2\nu+2})$ as $\epsilon \to 0$. Now for a fixed $\boldsymbol{t}_0 \in \mathbb{R}^d$

$$\mathsf{E}(\Delta^m_{\boldsymbol{h}/n} Y(\boldsymbol{t}_0))^2 = \Delta^{(m,m)}_{\boldsymbol{h}/n} \Big\{ K(|\boldsymbol{s} - \boldsymbol{t}|) \Big\} \Big|_{\boldsymbol{s},\boldsymbol{t}=\boldsymbol{t}_0} = \mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3$$

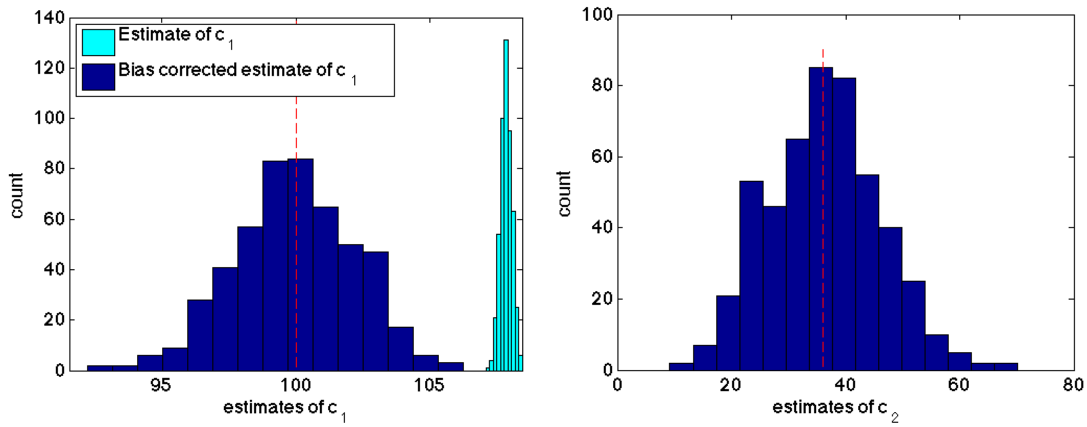

Fig 3: Histograms of the estimates of c_1 and c_2 for 500 independent realizations of $\sqrt{c_1}\,Y_1 + \sqrt{c_2}\,Y_2$ where $c_1 = 100$, $c_2 = 36$ and Y_1 is independent of Y_2. The random field Y_1 has principle irregular term $-|t|^{0.2}$ and Y_2 has principle irregular term $-|t|^{0.4}$. Each realization is on $[0, 1/\sqrt{2})^2$ measured at 1000×1000 pixel locations.

where

$$\mathcal{I}_1 \triangleq \Delta^{(m,m)}_{\boldsymbol{h}/n}\Big\{G_\nu(|\boldsymbol{s}-\boldsymbol{t}|)\Big\}\Big|_{\boldsymbol{s},\boldsymbol{t}=\boldsymbol{t}_0} = \sum_{ij} d_i d_j G_\nu(|(i-j)\boldsymbol{h}/n|) \tag{18}$$

$$\mathcal{I}_2 \triangleq \Delta^{(m,m)}_{\boldsymbol{h}/n}\Big\{(-\nu)G_{\nu+1}(|\boldsymbol{s}-\boldsymbol{t}|)\Big\}\Big|_{\boldsymbol{s},\boldsymbol{t}=\boldsymbol{t}_0} = \sum_{ij} d_i d_j (-\nu) G_{\nu+1}(|(i-j)\boldsymbol{h}/n|) \tag{19}$$

$$\mathcal{I}_3 \triangleq \Delta^{(m,m)}_{\boldsymbol{h}/n}\Big\{r(|\boldsymbol{s}-\boldsymbol{t}|)\Big\}\Big|_{\boldsymbol{s},\boldsymbol{t}=\boldsymbol{t}_0} = \sum_{ij} d_i d_j r(|(i-j)\boldsymbol{h}/n|) \tag{20}$$

Notice that $\sum_{ij} d_i d_j G_\nu(|(i-j)\boldsymbol{h}/n|) = |\boldsymbol{h}/n|^{2\nu} \sum_{ij} d_i d_j G_\nu(|i-j|)$. This is obviously true with $\nu \notin \mathbb{Z}$. It also holds when $\nu \in \mathbb{Z}$ since

$$G_\nu(|(i-j)\boldsymbol{h}/n|) = |\boldsymbol{h}/n|^{2\nu}\big(G_\nu(|i-j|) + |i-j|^{2\nu}\log|\boldsymbol{h}/n|\big) \tag{21}$$

and $\sum_{ij} d_i d_j |i-j|^{2\nu} = 0$ (since $\nu \in \mathbb{Z}$ and $m > \nu$). Similar arguments can be applied to $G_{\nu+1}$ when $m > \nu+1$ which gives $\mathcal{I}_1 + \mathcal{I}_2 = \frac{a_\nu^m}{n^{2\nu}} + \frac{b_\nu^m}{n^{2\nu+2}}$. Finally, notice that $r(\epsilon) = o(\epsilon^{2\nu+2})$ implies that $\mathcal{I}_3 = o(n^{-2\nu-2})$. This establishes the claim when $\sigma = \alpha = 1$ and M is the identity matrix. The general result when $\sigma, \nu > 0$ and $M \in GL(d, \mathbb{R})$ is then established by an easy rescaling argument (using equation (21) when $\nu \in \mathbb{Z}$). □

Lemma 2. *For $\nu > 0$, let a_ν^m be defined by (9) and b_ν^m be defined by (10). If $m > \nu$ then $a_\nu^m \neq 0$. If $m > \nu + 1$ then $b_\nu^m \neq 0$. Finally, there exits $p, q > \nu + 1$ such that $b_\nu^p - \frac{a_\nu^p}{a_\nu^q} b_\nu^q \neq 0$.*

PROOF. Notice first that $a_\nu^m \propto \mathrm{var}(\Delta_1^m Z_\nu) > 0$ where Z_ν is an intrinsic random function on $\mathbb{R}$ observed on $\mathbb{Z}$ with generalized covariance G_ν (since Δ_1^m annihilates polynomials of order $m-1$, and $m > \nu$, see [28]). The same reasoning establishes that $-b_\nu^m \propto \mathrm{var}(\Delta_1^m Z_{\nu+1}) > 0$ when $m > \nu + 1$.

For the last part of the lemma we show that there exists $p, q > \nu + 1$ such that

$$\frac{\mathrm{var}(\Delta_1^p Z_\nu)}{\mathrm{var}(\Delta_1^q Z_\nu)} \neq \frac{\mathrm{var}(\Delta_1^p Z_{\nu+1})}{\mathrm{var}(\Delta_1^q Z_{\nu+1})}.$$

We will argue by contradiction and suppose that for all $k > 0$,

$$\frac{\mathrm{var}(\Delta_1^{q+k} Z_\nu)}{\mathrm{var}(\Delta_1^q Z_\nu)} = \frac{\mathrm{var}(\Delta_1^{q+k} Z_{\nu+1})}{\mathrm{var}(\Delta_1^q Z_{\nu+1})}. \tag{22}$$

By a spectral representation of G_ν (see [28] page 36) and an easy induction establishes that $\mathrm{var}(\Delta_1^{q+k} Z_\nu) = \int |e^{iw}-1|^{2q+2k}|w|^{-2\nu-1}\,dw$ and $\mathrm{var}(\Delta_1^{q+k} Z_{\nu+1}) = \int |e^{iw}-1|^{2q+2k}|w|^{-2\nu-3}\,dw$. Notice also that $|e^{iw}-1|^2 = 2 - 2\cos w$. Let F_ν and $F_{\nu+1}$ be two probability measures on $\mathbb{R}$ defined by

$$\begin{aligned} F_\nu(B) &\triangleq \frac{1}{\mathrm{var}(\Delta_1^q Z_\nu)} \int_B (2 - 2\cos w)^q |w|^{-2\nu-1}\,dw \\ F_{\nu+1}(B) &\triangleq \frac{1}{\mathrm{var}(\Delta_1^q Z_{\nu+1})} \int_B (2 - 2\cos w)^q |w|^{-2\nu-3}\,dw. \end{aligned}$$

Our assumption (22) then becomes

$$\int (2 - 2\cos w)^k dF_\nu(w) = \int (2 - 2\cos w)^k dF_{\nu+1}(w), \tag{23}$$

for all $k > 0$. Notice that the variances $\mathrm{var}(\Delta_1^q Z_\nu)$ and $\mathrm{var}(\Delta_1^q Z_{\nu+1})$ serve as the normalizing constants so that F_ν and $F_{\nu+1}$ have total mass one. In what follows we show that the normalizing constants satisfy both $\mathrm{var}(\Delta_1^q Z_\nu) > \mathrm{var}(\Delta_1^q Z_{\nu+1})$ and $\mathrm{var}(\Delta_1^q Z_\nu) < \mathrm{var}(\Delta_1^q Z_{\nu+1})$ to establish the desired contradiction.

By the equalities in (23), the random variables $2(1 - \cos W_\nu)$ and $2(1 - \cos W_{\nu+1})$ have the same moments when $W_\nu \sim F_\nu$ and $W_{\nu+1} \sim F_{\nu+1}$. In addition, $0 \leq 2(1 - \cos W_\nu) \leq 4$ and $0 \leq 2(1 - \cos W_{\nu+1}) \leq 4$ so that the moment generating functions are both finite in a non-empty radius of the origin. Therefore $2(1-\cos W_\nu) \overset{\mathcal{L}}{=} 2(1-\cos W_{\nu+1})$, where $\overset{\mathcal{L}}{=}$ denotes equality in law. This gives $\mathsf{P}(\cos W_\nu < 0) = \mathsf{P}(\cos W_{\nu+1} < 0)$, for example. However

$$\begin{aligned} \mathsf{P}(\cos W_\nu < 0) &= \frac{1}{\mathrm{var}(\Delta_1^q Z_\nu)} \int \mathbf{1}_{\{\cos w < 0\}} (2 - 2\cos w)^q |w|^{-2\nu-1} dw \\ &> \frac{1}{\mathrm{var}(\Delta_1^q Z_\nu)} \int \mathbf{1}_{\{\cos w < 0\}} (2 - 2\cos w)^q |w|^{-2\nu-3} dw, \end{aligned}$$

by the fact that $\cos w < 0 \Rightarrow |w| > \pi/2$. Therefore

$$\operatorname{var}(\Delta_1^q Z_{\nu+1}) < \operatorname{var}(\Delta_1^q Z_\nu). \tag{24}$$

To show the contradicting inequality let's start by computing the density of these two random variables. The idea is to show that the non-normalized (i.e. without the term $\operatorname{var}(\Delta_1^q Z_\nu)$) density of $2(1-\cos W_\nu)$ is strictly smaller than the non-normalized density of $2(1-\cos W_{\nu+1})$ in a positive neighborhood of 0. In particular, the density of $2(1-\cos W_\nu)$ can be written as $2\sum_{k=1}^\infty f_{W_\nu}(g_k(x))|g_k(x)'|$ where the g_k's are the different positive branches of the inverse $\cos^{-1}(1-x/2)$ and $f_{W_\nu}(w) \triangleq (2-2\cos w)^q |w|^{-2\nu-1}/\operatorname{var}(\Delta_1^q Z_\nu)$ is the density of W_ν. This simplifies to

$$\frac{2x^q}{\operatorname{var}(\Delta_1^q Z_\nu)} \sum_{k=1}^\infty \frac{|g_k(x)'|}{|g_k(x)|^{2\nu+1}} = \frac{2x^q}{\operatorname{var}(\Delta^q Z_\nu)\sqrt{x - x^2/4}} \sum_{k=1}^\infty |g_k(x)|^{-2\nu-1}$$

for $0 < x < 4$. Notice that $g_1(x) \sim \sqrt{x}$ as $x \to 0$ and $g_k(x) \sim 2\pi\lfloor k/2 \rfloor$ as $x \to 0$ for all $k > 1$. Therefore the term g_1 dominates the sum when x is small. In particular for all $x > 0$ sufficiently small we have

$$f_{2-2\cos W_\nu}(x) < \frac{2x^q}{\operatorname{var}(\Delta_1^q Z_\nu)\sqrt{x - x^2/4}} \sum_{k=1}^\infty |g_k(x)|^{-2\nu-3} \tag{25}$$

$$= \frac{\operatorname{var}(\Delta_1^q Z_{\nu+1})}{\operatorname{var}(\Delta_1^q Z_\nu)} f_{2-2\cos W_{\nu+1}}(x). \tag{26}$$

Since $f_{2-2\cos W_\nu}(x)$ and $f_{2-2\cos W_{\nu+1}}(x)$ have the same integrate integrals over Borel subsets of $(0,4)$, we must have $\operatorname{var}(\Delta_1^q Z_{\nu+1}) > \operatorname{var}(\Delta_1^q Z_\nu)$. This contradicts (24) and therefore establishes the lemma. □

LEMMA 3. *For any $\nu > 0$, $T > 0$,*

$$\left|\frac{d^p}{dt^p} t^{\nu/2}\mathcal{K}_\nu(\sqrt{t})\right| \lesssim \begin{cases} 1, & \textit{when } p < \nu; \\ |\log t|, & \textit{when } p = \nu; \\ t^{\nu-p}, & \textit{when } p > \nu; \end{cases}$$

as t ranges in the interval $(0,T)$ where $\mathcal{K}_\nu$ is the modified Bessel function of the second kind of order ν.

PROOF. Using the expansions for $\mathcal{K}_\nu$ found in [1] (page 375) we can write

$$t^{\nu/2}\mathcal{K}_\nu(\sqrt{t}) = \begin{cases} F_1(t) + t^\nu \log(t) F_3(t); & \text{when } \nu = 0, 1, 2, \ldots \\ F_4(t) + t^\nu F_5(t); & \text{otherwise} \end{cases} \tag{27}$$

where the $F_j(t)$'s are of the form $\sum_{k=0}^\infty c_k t^k$ where the c_k's decay fast enough so that the series converges absolutely for all $t \in (0,\infty)$ and all it's derivatives exist and are bounded on $(0,T)$. This immediately establishes that when $p<\nu$, $\left|\frac{d^p}{dt^p} t^{\nu/2}\mathcal{K}_\nu(\sqrt{t})\right| \lesssim 1$ for all $t\in(0,T)$ since both $\frac{d^p}{dt^p}(t^\nu)$ and $\frac{d^p}{dt^p}(t^\nu \log t)$ are continuous and bounded on $(0,T)$.

When $p>\nu$ and $\nu\notin\mathbb{Z}$ we have that $t^\nu \lesssim \frac{d}{dt}(t^\nu) \lesssim \cdots \lesssim \frac{d^p}{dt^p}(t^\nu) \lesssim t^{\nu-p}$ as t ranges in the bounded interval $(0,T)$. Similarly, when $p>\nu$ and $\nu\in\mathbb{Z}$ we have

$$t^\nu \log t \lesssim \frac{d}{dt}(t^\nu \log t) \lesssim \cdots \lesssim \frac{d^p}{dt^p}(t^\nu \log t) \lesssim t^{\nu-p}.$$

Finally, when $p=\nu$, $\frac{d^p}{dt^p} t^\nu \log t \propto \log t + c_p$. The lemma now follows by equation (27) and the fact that the derivative of a product satisfies $(fg)^{(p)} = \sum_{k=0}^p \binom{p}{k} f^{(p)} g^{(p-k)}$.

□

Lemma 4. *Suppose $K(t)$ is the isotropic Matérn autocovariance function defined in (6) for fixed parameters $\sigma,\alpha,\nu>0$. Then for any integer $m>\nu$, nonzero vector $\boldsymbol{h}\in\mathbb{R}^d$, matrix $M\in GL(d,\mathbb{R})$ and bounded set $\Omega\subset\mathbb{R}^d$*

$$\left|\partial_{\boldsymbol{h}}^{(m,m)}\left[K(|M\boldsymbol{s}-M\boldsymbol{t}|)\right]\right| \gtrsim |\boldsymbol{s}-\boldsymbol{t}|^{2\nu-2m} \tag{28}$$

for all $\boldsymbol{s},\boldsymbol{t}\in\Omega$ such that $\boldsymbol{s}\neq\boldsymbol{t}$.

Proof. First notice that it is sufficient to show the claim when M is the identity matrix and $\alpha=1$ (extending to general M and α follows by the chain rule for derivatives). Define $K_{sq}(t)\triangleq K(\sqrt{t})$ and $F(\boldsymbol{s},\boldsymbol{t})\triangleq|\boldsymbol{s}-\boldsymbol{t}|^2$ so that $\partial_{\boldsymbol{h}}^{(m,m)}[K(|\boldsymbol{s}-\boldsymbol{t}|)] = \partial_{\boldsymbol{h}}^{(m,m)}[K_{sq}(F(\boldsymbol{s},\boldsymbol{t}))]$. Also let $\partial_{\boldsymbol{h}}^*$ denote a generic directional derivative on either the variable $\boldsymbol{s}$ or $\boldsymbol{t}$. By generic I mean that $(\partial_{\boldsymbol{h}}^*)^k F$ denotes $\partial_{\boldsymbol{h}}^{(i,j)}F$ for some $i+j=k$ and $(\partial_h^* F)^k = \partial_{\boldsymbol{h}}^* F\cdots\partial_{\boldsymbol{h}}^* F$ where each $\partial_{\boldsymbol{h}}^*$ could be with respect to $\boldsymbol{s}$ or $\boldsymbol{t}$. Now by successive application of the directional derivatives $\partial_{\boldsymbol{h}}^*$ we get that

$$\partial_{\boldsymbol{h}}^{(m,m)}[K_{sq}(F(\boldsymbol{s},\boldsymbol{t}))] = \sum_{i=1}^{2m} \sum_{\substack{0\leq j\leq i\\ j+i\leq 2m}} K_{sq}^{(i)}(F(\boldsymbol{s},\boldsymbol{t}))(\partial_{\boldsymbol{h}}^* F(\boldsymbol{s},\boldsymbol{t}))^{i-j} B_{ij} \tag{29}$$

where each B_{ij} is uniformly bounded on Ω^2. The functions B_{ij} are uniformly bounded by the nice fact that $(\partial_{\boldsymbol{h}}^*)^k F(\boldsymbol{s},\boldsymbol{t}) \lesssim 1$ on Ω^2 when $k\geq 2$.

We will bound the terms of the sum (29) when $i<\nu$, $i>\nu$, and $i=\nu$ separately. Notice first that since $i\geq j$ we have that

$$\left|\partial_{\boldsymbol{h}}^* F(\boldsymbol{s},\boldsymbol{t})\right|^{i-j} \lesssim |\boldsymbol{s}-\boldsymbol{t}|^{i-j}, \quad \text{for all } \boldsymbol{s},\boldsymbol{t}\in\Omega. \tag{30}$$

This implies, by Lemma 3, that the terms in the sum (29), for which $i < \nu$, are bounded. When $i > \nu$

$$\begin{aligned} |K_{sq}^{(i)}(F(\boldsymbol{s},\boldsymbol{t}))(\partial_{\boldsymbol{h}}^{*}F(\boldsymbol{s},\boldsymbol{t}))^{i-j}B_{ij}| &\lesssim |F(\boldsymbol{s},\boldsymbol{t})|^{\nu-i}|\boldsymbol{s}-\boldsymbol{t}|^{i-j}, \quad \text{by (30) and Lemma 3} \\ &= |\boldsymbol{s}-\boldsymbol{t}|^{2\nu-(i+j)} \\ &\lesssim |\boldsymbol{s}-\boldsymbol{t}|^{2\nu-2m}, \quad \text{since } i+j \le 2m, \end{aligned}$$

where the inequality holds for all $\boldsymbol{s},\boldsymbol{t} \in \Omega$ such that $|\boldsymbol{s}-\boldsymbol{t}| > 0$ (note that we use the fact that Ω is bounded implies $|\boldsymbol{s}-\boldsymbol{t}| < T$ for some T). For the last case, $i = \nu$, a similar argument establishes

$$\begin{aligned} |K_{sq}^{(i)}(F(\boldsymbol{s},\boldsymbol{t}))(\partial_{\boldsymbol{h}}^{*}F(\boldsymbol{s},\boldsymbol{t}))^{i-j}B_{ij}| &\lesssim |\boldsymbol{s}-\boldsymbol{t}|^{i-j}|\log F(\boldsymbol{s},\boldsymbol{t})| \\ &\lesssim |\boldsymbol{s}-\boldsymbol{t}|^{2\nu-2m} \end{aligned}$$

for all $\boldsymbol{s},\boldsymbol{t} \in \Omega$ such that $|\boldsymbol{s}-\boldsymbol{t}| > 0$. Therefore $\partial_{\boldsymbol{h}}^{(m,m)}[K_{sq}(F(\boldsymbol{s},\boldsymbol{t}))] \lesssim |\boldsymbol{s}-\boldsymbol{t}|^{2\nu-2m}$ for all $\boldsymbol{s},\boldsymbol{t} \in \Omega$ such that $\boldsymbol{s} \neq \boldsymbol{t}$. □

LEMMA 5. *Let $\boldsymbol{h}$ be a nonzero vector in $\mathbb{R}^d$, $\nu > 0$ and H be the $d \times m$ matrix defined by*

$$H \triangleq \underbrace{(\boldsymbol{h}, \cdots, \boldsymbol{h})}_{m \text{ columns}}. \tag{31}$$

If m is a positive integer greater than ν then

$$\sup_{\boldsymbol{\xi},\boldsymbol{\eta}\in[0,1]^m} |\boldsymbol{i}-\boldsymbol{j}+H(\boldsymbol{\xi}-\boldsymbol{\eta})/n|^{2\nu-2m} \lesssim |\boldsymbol{i}-\boldsymbol{j}|^{2\nu-2m}$$

for all positive integers n and $\boldsymbol{i},\boldsymbol{j} \in \Omega_n$ such that $|\boldsymbol{i}-\boldsymbol{j}| > |(m+1)\boldsymbol{h}/n|$.

PROOF. First notice that $\sup_{\boldsymbol{\xi},\boldsymbol{\eta}\in[0,1]^m} |\boldsymbol{i}-\boldsymbol{j}+H(\boldsymbol{\xi}-\boldsymbol{\eta})/n|^{2\nu-2m} = \sup_{-1\le\tau\le1} |\boldsymbol{i}-\boldsymbol{j}+m\boldsymbol{h}\tau/n|^{2\nu-2m}$. Now for any $-1 \le \tau \le 1$, positive integer n and $\boldsymbol{i},\boldsymbol{j} \in \Omega_n$ such that $|\boldsymbol{i}-\boldsymbol{j}| > |(m+1)\boldsymbol{h}/n|$, we have

$$\begin{aligned} |\boldsymbol{i}-\boldsymbol{j}+m\boldsymbol{h}\tau/n| &\ge |\boldsymbol{i}-\boldsymbol{j}| - m|\tau||\boldsymbol{h}|/n \\ &\ge |\boldsymbol{i}-\boldsymbol{j}| - |\boldsymbol{i}-\boldsymbol{j}|\frac{m}{m+1}. \end{aligned}$$

The last line follows from the assumption that $|\boldsymbol{i}-\boldsymbol{j}| > (m+1)|\boldsymbol{h}|/n$ which implies $\frac{m}{m+1}|\boldsymbol{i}-\boldsymbol{j}| > m|\boldsymbol{h}|/n$. Therefore

$$\sup_{\boldsymbol{\xi},\boldsymbol{\eta}\in[0,1]^m} \left|\boldsymbol{i}-\boldsymbol{j}+H(\boldsymbol{\xi}-\boldsymbol{\eta})/n\right|^{2\nu-2m} \le \Big(1-\frac{m}{m+1}\Big)^{2\nu-2m}|\boldsymbol{i}-\boldsymbol{j}|^{2\nu-2m}.$$

□

Lemma 6. *Let Y be a mean zero, geometric anisotropic d-dimensional Matérn Gaussian random field with parameters (σ,α,ν,M) and let Ω be a bounded, open subset of $\mathbb{R}^d$. Fix a positive integer $m>\nu$ and a non-zero vector $\boldsymbol{h}\in\mathbb{R}^d$. Let Σ to be the covariance matrix of the increments $\Delta^m_{\boldsymbol{h}/n}Y(\boldsymbol{i})$ as $\boldsymbol{i}$ ranges in the set $\boldsymbol{i}\in\Omega_n$ so that*

$$\Sigma(\boldsymbol{i},\boldsymbol{j})\triangleq \mathsf{E}(\Delta^m_{\boldsymbol{h}/n}Y(\boldsymbol{i})\Delta^m_{\boldsymbol{h}/n}Y(\boldsymbol{j})) \tag{32}$$

for all $\boldsymbol{i},\boldsymbol{j}\in\Omega_n$. Then there exists an $N>0$ such that

$$|\Sigma(\boldsymbol{i},\boldsymbol{j})|\lesssim n^{-2m}|\boldsymbol{i}-\boldsymbol{j}|^{2\nu-2m} \tag{33}$$

for all $n>N$, and $\boldsymbol{i},\boldsymbol{j}\in\Omega_n$ such that $|\boldsymbol{i}-\boldsymbol{j}|>|(m+1)\boldsymbol{h}/n|$. Moreover,

$$|\Sigma(\boldsymbol{i},\boldsymbol{j})|\lesssim n^{-2\nu} \tag{34}$$

for all $n>N$ and $\boldsymbol{i},\boldsymbol{j}\in\Omega_n$ such that $|\boldsymbol{i}-\boldsymbol{j}|\le|(m+1)\boldsymbol{h}/n|$.

Proof. First notice that $\Sigma(\boldsymbol{i},\boldsymbol{j})=\mathsf{E}\Delta^m_{\boldsymbol{h}/n}Y(\boldsymbol{i})\Delta^m_{\boldsymbol{h}/n}Y(\boldsymbol{j})=\Delta^{(m,m)}_{\boldsymbol{h}/n}K(|M(\boldsymbol{i}-\boldsymbol{j})|)$ where K is the isotropic Matérn autocovariance function defined in (6). To simplify the notation let $F(\boldsymbol{i},\boldsymbol{j})\triangleq K(|M(\boldsymbol{i}-\boldsymbol{j})|)$ and H be the d by m matrix defined in (31). An induction argument on m establishes that when $|\boldsymbol{i}-\boldsymbol{j}|>(m+1)|\boldsymbol{h}|/n$ we can express directional increments as integrals of directional derivatives so that

$$\Delta^{(m,m)}_{\boldsymbol{h}/n}F(\boldsymbol{i},\boldsymbol{j})=\frac{1}{n^{2m}}\int_{\boldsymbol{\xi},\boldsymbol{\eta}\in[0,1]^m}(\partial^{(m,m)}_{\boldsymbol{h}}F)(\boldsymbol{i}+H\boldsymbol{\xi}/n,\boldsymbol{j}+H\boldsymbol{\eta}/n)d\boldsymbol{\xi}d\boldsymbol{\eta}.$$

Therefore

$$\begin{aligned}
|\Sigma(\boldsymbol{i},\boldsymbol{j})| &\lesssim \frac{1}{n^{2m}}\int_{\boldsymbol{\xi},\boldsymbol{\eta}\in[0,1]^m}|(\partial^{(m,m)}_{\boldsymbol{h}}F)(\boldsymbol{i}+H\boldsymbol{\xi}/n,\boldsymbol{j}+H\boldsymbol{\eta}/n)|d\boldsymbol{\xi}d\boldsymbol{\eta}\\
&\lesssim \frac{1}{n^{2m}}\int_{\boldsymbol{\xi},\boldsymbol{\eta}\in[0,1]^m}\left|\boldsymbol{i}-\boldsymbol{j}+H(\boldsymbol{\xi}-\boldsymbol{\eta})/n\right|^{2\nu-2m}d\boldsymbol{\xi}d\boldsymbol{\eta}, \quad \text{by Lemma 4}\\
&\lesssim \frac{1}{n^{2m}}\sup_{\boldsymbol{\xi},\boldsymbol{\eta}\in[0,1]^m}\left|\boldsymbol{i}-\boldsymbol{j}+H(\boldsymbol{\xi}-\boldsymbol{\eta})/n\right|^{2\nu-2m}\\
&\lesssim \frac{1}{n^{2m}}\left|\boldsymbol{i}-\boldsymbol{j}\right|^{2\nu-2m}, \quad \text{by Lemma 5}
\end{aligned}$$

for all $n>N$, $\boldsymbol{i},\boldsymbol{j}\in\Omega_n$ such that $|\boldsymbol{i}-\boldsymbol{j}|>|(m+1)\boldsymbol{h}/n|$. On the other hand when $|\boldsymbol{i}-\boldsymbol{j}|\le|(m+1)\boldsymbol{h}/n|$

$$|\Sigma(\boldsymbol{i},\boldsymbol{j})|\le\sqrt{\mathsf{E}(\Delta^m_{\boldsymbol{h}/n}Y(\boldsymbol{i}))^2}\sqrt{\mathsf{E}(\Delta^m_{\boldsymbol{h}/n}Y(\boldsymbol{j}))^2}\lesssim n^{-2\nu} \tag{35}$$

where the last inequality is by Lemma 1. Actually, a direct application of Lemma 1 only establishes (35) when $m>\nu+1$. However, a small adjustment of the proof of Lemma 1 establishes that $\mathsf{E}(\Delta^m_{\boldsymbol{h}/n}Y(\boldsymbol{t}))^2=\frac{a^m_\nu}{n^{2\nu}}+o(n^{-2\nu})$ as $n\to\infty$ when $m>\nu$. This is then is sufficient to establish (35). □

LEMMA 7. *Let Σ_{abs} be the component-wise absolute value of the covariance matrix Σ (defined in (32)). Then under the same assumptions as in Lemma 6, there exits an $N > 0$ such that*

$$\|\Sigma_{abs}\|_2 \lesssim n^{-2\nu} + c\, n^{d-2m} \int_{1/n}^{1} r^{2\nu-2m+d-1} dr.$$

for all $n > N$, where c is a constant and $\|\cdot\|_2$ is the spectral norm.

PROOF. First note that by symmetry, $\|\Sigma_{abs}\|_2 \leq \sqrt{\|\Sigma_{abs}\|_1 \|\Sigma_{abs}\|_\infty} = \|\Sigma_{abs}\|_\infty$, where $\|\Sigma_{abs}\|_\infty$ is the maximum of the ℓ_1 row norms and $\|\Sigma_{abs}\|_1$ is the maximum of the ℓ_1 column norms. To bound the ℓ_1 row norms, we bound the terms of the sum when $|\boldsymbol{i}-\boldsymbol{j}| > (m+1)|\boldsymbol{h}|/n$ and $|\boldsymbol{i}-\boldsymbol{j}| \leq (m+1)|\boldsymbol{h}|/n$ separately. For the off-diagonal terms we use Lemma 6 to ensure the existence of an $N > 0$ such that for all $n > N$

$$\max_{\boldsymbol{i}\in\Omega_n} \sum_{\substack{\boldsymbol{j}\in\Omega_n \\ |\boldsymbol{i}-\boldsymbol{j}|>(m+1)|\boldsymbol{h}|/n}} |\Sigma(\boldsymbol{i},\boldsymbol{j})| \lesssim \max_{\boldsymbol{i}\in\Omega_n} \sum_{\substack{\boldsymbol{j}\in\Omega_n \\ |\boldsymbol{i}-\boldsymbol{j}|>(m+1)|\boldsymbol{h}|/n}} n^{-2m}|\boldsymbol{i}-\boldsymbol{j}|^{2\nu-2m} \tag{36}$$

$$\lesssim n^{d-2m}\int_{1/n}^{1} r^{2\nu-2m+d-1} dr. \tag{37}$$

The last inequality, (37), follows by the fact that for any constant $a > 0$ and open set $\Theta \subset \mathbb{R}^d$ which is bounded and contains the origin, one has

$$\sum_{\substack{\boldsymbol{i}\in\Theta\cap\{\mathbb{Z}^d/n\} \\ |\boldsymbol{i}|>a/n}} n^{-d}|\boldsymbol{i}|^{\beta} \lesssim \int_{1/n}^{1} r^{\beta+d-1} dr \tag{38}$$

as $n \to \infty$ (for details see [3], Lemma 3, page 41). In addition, by Lemma 6

$$\max_{\boldsymbol{i}\in\Omega_n} \sum_{\substack{\boldsymbol{j}\in\Omega_n \\ |\boldsymbol{i}-\boldsymbol{j}|\leq(m+1)|\boldsymbol{h}|/n}} |\Sigma(\boldsymbol{i},\boldsymbol{j})| \lesssim n^{-2\nu} \tag{39}$$

for all $n > N$. This establishes the proof by noticing that the sum of the last terms in (37) and (39) bound $\|\Sigma_{abs}\|_\infty$. $\square$

LEMMA 8. *Under the same assumptions as in Lemma 6 there exits an $N > 0$ such that*

$$\|\Sigma\|_F^2 \lesssim n^{d-4\nu} + c\, n^{2d-4m} \int_{1/n}^{1} r^{4\nu-4m+d-1} dr$$

for all $n > N$ where c is a constant and $\|\cdot\|_F$ denotes the Frobenious matrix norm.

PROOF. First note that $\|\Sigma\|_F^2 = \sum_{\boldsymbol{i},\boldsymbol{j}\in\Omega_n} |\Sigma(\boldsymbol{i},\boldsymbol{j})|^2$. As in the proof of Lemma 8 we bound the near-diagonal terms of Σ separately from the off-diagonal terms. By Lemma 6 there exists an $N > 0$ such that

$$\sum_{\substack{\boldsymbol{i},\boldsymbol{j}\in\Omega_n \\ |\boldsymbol{i}-\boldsymbol{j}|>(m+1)|\boldsymbol{h}|/n}} |\Sigma(\boldsymbol{i},\boldsymbol{j})|^2 \lesssim n^{2d-4m} \sum_{\substack{\boldsymbol{i},\boldsymbol{j}\in\Omega_n \\ |\boldsymbol{i}-\boldsymbol{j}|>(m+1)|\boldsymbol{h}|/n}} n^{-2d}|\boldsymbol{i}-\boldsymbol{j}|^{4\nu-4m} \tag{40}$$

$$\lesssim n^{2d-4m}\int_{1/n}^{1} r^{4\nu-4m+d-1}dr \tag{41}$$

for all $n > N$. Notice that the last inequality is a slight variation on (38). For the near diagonal terms we also use Lemma 6 to get

$$\sum_{\substack{\boldsymbol{i},\boldsymbol{j}\in\Omega_n \\ |\boldsymbol{i}-\boldsymbol{j}|\leq(m+1)|\boldsymbol{h}|/n}} |\Sigma(\boldsymbol{i},\boldsymbol{j})|^2 \lesssim n^d n^{-4\nu}. \tag{42}$$

Adding (41) and (42) establishes the lemma. □

Proof of Thorem 1. Define the random vector ΔY to be the vector of m-increments, the components of which are indexed by Ω_n (in any order), so that

$$\Delta Y \triangleq \underbrace{(\ldots,\Delta^m_{\boldsymbol{h}/n}Y(\boldsymbol{j}),\ldots)}_{\text{terms are indexed by } \boldsymbol{j}\in\Omega_n} . \tag{43}$$

Now we can write $Q_n^m = \frac{n^{2\nu}}{\#\Omega_n}\Delta Y\Delta Y^T = \frac{n^{2\nu}}{\#\Omega_n}W\Sigma W^T$, where $W \sim \mathcal{N}(\mathbf{0}, I)$ (note that Σ is defined in (32)). Therefore $\operatorname{var} Q_n^m = 2\frac{n^{4\nu}}{(\#\Omega)^2}\|\Sigma\|_F^2$ and by Lemma 8

$$\begin{aligned}\frac{n^{4\nu}}{(\#\Omega)^2}\|\Sigma\|_F^2 &\lesssim n^{-d} + c\,n^{4\nu-4m}\int_{1/n}^{1} r^{4\nu-4m+d-1}dr \\ &\lesssim \begin{cases} n^{4(\nu-m)}, & \text{if } 4(\nu-m) > -d; \\ n^{-d}\log n, & \text{if } 4(\nu-m) = -d; \\ n^{-d}, & \text{if } 4(\nu-m) < -d \end{cases}\end{aligned}$$

for all sufficiently large n. This establishes the variance rates.

For the almost sure convergence result let $\tilde\Sigma \triangleq \frac{n^{2\nu}}{\#\Omega_n}\Sigma_{abs}$ where Σ_{abs} is the component-wise absolute value of Σ. The Hanson and Wright bound in [17] then gives

$$\mathsf{P}(|Q_n^m - EQ_n^m| \geq \epsilon) \leq 2\exp\left(-\frac{c_1\epsilon}{\|\tilde\Sigma\|_2} \wedge \frac{c_2\epsilon^2}{\|\tilde\Sigma\|_F^2}\right) \tag{44}$$

for all $\epsilon > 0$, where c_1, c_2 are positive constants not depending on n or $\tilde{\Sigma}$. First notice that by Lemma 7 we get

$$\|\tilde{\Sigma}\|_2 = \frac{n^{2\nu}}{\#\Omega_n}\|\Sigma_{abs}\|_2 \lesssim n^{-d} + c\, n^{2\nu-2m}\int_{1/n}^{1} r^{2\nu-2m+d-1}dr \tag{45}$$

$$\lesssim \begin{cases} n^{2(\nu-m)}, & \text{if } 2(\nu-m) > -d; \\ n^{-d}\log n, & \text{if } 2(\nu-m) = -d; \\ n^{-d}, & \text{if } 2(\nu-m) < -d. \end{cases} \tag{46}$$

for sufficiently large n. Also notice that this implies that $\|\tilde{\Sigma}\|_F^2 \lesssim \|\tilde{\Sigma}\|_2$ for sufficiently large n. Therefore for sufficiently small ϵ, $\mathsf{P}(|Q_n^m - EQ_n^m| \geq \epsilon) \leq 2\exp\left(-c_2\epsilon^2/\|\tilde{\Sigma}\|_2\right)$. Now the rates in (46) and the Borel-Cantelli Lemma are sufficient to establish that $Q_n^m - \mathsf{E}Q_n^m \to 0$, with probability one as $n \to \infty$. By Lemma 1, $\mathsf{E}Q_n^m \to a_\nu^m$ (a slight adjustment also proves the case when $m > \nu$ rather than $m > \nu + 1$) which establishes the theorem. □

Proof of Theorem 2. First notice that when $p, q > \nu + 1$

$$\mathsf{E}n^2\left[Q_n^p - \frac{a_\nu^p}{a_\nu^q}Q_n^q\right] \to \left[b_\nu^p - \frac{a_\nu^p}{a_\nu^q}b_\nu^q\right] \tag{47}$$

as $n \to \infty$ by Lemma 1. To get almost sure convergence notice

$$\mathsf{P}\Big(n^2\Big|[Q_n^p - \frac{a_\nu^p}{a_\nu^q}Q_n^q] - \mathsf{E}[Q_n^p - \frac{a_\nu^p}{a_\nu^q}Q_n^q]\Big| \geq \epsilon\Big) \tag{48}$$

$$\leq \mathsf{P}\big(|[Q_n^p - \mathsf{E}Q_n^p| \geq \epsilon/2n^2\big) + \mathsf{P}\big(|a_\nu^p||Q_n^q - \mathsf{E}Q_n^q| \geq |a_\nu^q|\epsilon/2n^2\big) \tag{49}$$

We can again use the Hanson and Wright bound ([17]) and the rates derived in Theorem 1 to get

$$\mathsf{P}(|Q_n^p - \mathsf{E}Q_n^p| \geq \epsilon/2n^2) \leq 2\exp\left(-c\, n^{-4}\epsilon^2/\|\tilde{\Sigma}\|_2\right) \tag{50}$$

for all sufficiently small $\epsilon > 0$ where c is a positive constant that doesn't depend on n or $\tilde{\Sigma}$. By inspection of the rates in (46) the Borel-Cantelli Lemma can be applied when $4 < \min\{2p - 2\nu, d\}$ so that $Q_n^p - EQ_n^p \to 0$ with probability one as $n \to \infty$. A similar result holds for the second term in (49) using the fact that both a_ν^p and a_ν^q are non-zero by Lemma 2. This, combined with convergence of the expectation in (47), completes the proof. □

REFERENCES

[1] M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions.* ninth ed. Dover, New York, 1965.

[2] R. J. Adler and R. Pyke. Uniform quadratic variation for Gaussian processes. *Stoch. Proc. Appl.*, 48:191–209, 1993.

[3] E. Anderes. *Estimating Deformations of Isotropic Gaussian Random Fields.* PhD thesis, University of Chicago, 2005.

[4] E. Anderes and S. Chatterjee. Consistent estimates of deformed isotropic gaussian random fields on the plane. *Ann. Stat. To appear.*

[5] G. Baxter. A strong limit theorem for Gaussian processes. *Proc. Amer. Math. Soc.*, 7:522–527, 1956.

[6] A. Benassi, S. Cohen, J. Istas, and S. Jaffard. Identification of filtered white noises. *Stoch. Proc. Appl.*, 75:31–49, 1998.

[7] S. M. Berman. A version of the Lévy-Baxter theorem for the increments of Brownian motion of several parameters. *Proc. Amer. Math. Soc.*, 18:1051–1055, 1967.

[8] D. Nychka C. Kaufman, M. Schervish. Covariance tapering for likelihood-based estimation in large spatial datasets. *J. Amer. Stat. Assoc. To appear.*

[9] J. Chilès and P. Delfiner. *Geostatistics: Modeling Spatial Uncertainty.* Wiley series in probability and statistics, 1999.

[10] S. Cohen, X. Guyon, O. Perrin, and M. Pontier. Identification of an isometric transformation of the standard brownian sheet. *J. of Stat. Planning and Inference.*, 136:1317–1330, 2006.

[11] S. Cohen, X. Guyon, O. Perrin, and M. Pontier. Singularity functions for fractional processes: application to the fractional brownian sheet. *Ann. Inst. Henri Poincaré*, PR 42:187–205, 2006.

[12] N. Cressie. *Statistics for Spatial Data, revised ed.* Wiley, New York, 1993.

[13] J. Du, H. Zhang, and V. Mandrekar. Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *Ann. Stat. To appear.*

[14] R. M. Dudley. Sample functions of the Gaussian process. *Ann. Probability*, 1:66–103, 1973.

[15] E. G. Gladyshev. A new limit theorem for stochastic processes with Gaussian increments. *Theor. Probability Appl.*, 6:52–61, 1961.

[16] X. Guyon and G. Leon. Convergence en loi des h-variations d'un processus gaussien stationnaire. *Ann. Inst. Henri Poincaré*, 25:265–282, 1989.

[17] D.L. Hanson and F.T. Wright. A bound on tail probabilities for quadratic form in independent random variables. *Ann. Math. Stat.*, 42:1079–1083, 1971.

[18] R. Horn and C. Johnson. *Matrix Analysis.* Cambridge University Press, 2007.

[19] I. A. Ibragimov and Y. A. Rozanov. *Gaussian Random Processes.* Springer, New York, 1978.

[20] J. Istas and G. Lang. Quadratic variations and estimation of the local Hölder index of a Gaussian process. *Ann. Inst. Henri Poincaré*, 33:407–436, 1997.

[21] R. Klein and E. Gine. On quadratic variation of processes with Gaussian increments. *Ann. Probability*, 3:716–721, 1975.

[22] P. Lévy. Le mouvement brownien plan. *Amer. J. Math.*, 62:487–550, 1940.

[23] W.-L. Loh. Fixed-domain asymptotics for a subclass of matérn-type gaussian random fields. *Ann. Stat.*, 33:2344–2394, 2005.

[24] W.-L. Loh and T.-K. Lam. Estimating structured correlation matrices in smooth gaussian random field models. *Ann. Stat.*, 28:880–904, 2000.

[25] M. L. Stein. Asymptotically efficient prediction of a random field with a misspecified covariance function. *Ann. Stat.*, 16:55–63, 1988.

[26] M. L. Stein. Uniform asymptotic optimality of linear predictions of a random field using an incorrect second-order structure. *Ann. Stat.*, 18:850–872, 1990.

[27] M. L. Stein. A simple condition for asymptotic optimality of linear predictions of random fields. *Statistics and Probability Letters*, 17:399–404, 1993.

[28] M. L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging.* Springer, New York, 1999.

[29] M. L. Stein. Fast and exact simulation of fractional Brownian surfaces. *J. of Comput. and Graph. Stat.*, 11:587–599, 2002.

[30] P. T. Strait. On Berman's version of the Lévy-Baxter theorem. *Proc. Amer. Math. Soc.*, 23:91–93, 1969.
[31] G. Wahba. *Spline Models for Observational Data.* SIAM, Philadelphia, 1990.
[32] Z. Ying. Asymptotic properties of a maximum likelihood estimator with data from a gaussian process. *J. of Multi. Analysis*, 36:280–296, 1991.
[33] H. Zhang. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Stat. Assoc.*, 99:250–261, 2004.

Department of Statistics
University of California at Davis
4214 Math Sci. Bldg.
One Shields Ave.
Davis, CA 95616