Journal of Multivariate Analysis 102 (2011) 506–520

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Local likelihood estimation for nonstationary random fields

Ethan B. Anderes [a,*], Michael L. Stein [b]

[a] Statistics Department, University of California at Davis, One Shields Avenue, Davis, CA 95616, United States
[b] Department of Statistics, University of Chicago, 5734 S. University Avenue, Chicago, IL 60637, United States

ARTICLE INFO

Article history:
Received 7 January 2010
Available online 26 October 2010

AMS 2000 subject classifications:
62M30
62M40
62G05

Keywords:
Local likelihood
Random fields
Nonstationarity
Local parameters

ABSTRACT

We develop a weighted local likelihood estimate for the parameters that govern the local spatial dependency of a locally stationary random field. The advantage of this local likelihood estimate is that it smoothly downweights the influence of faraway observations, works for irregular sampling locations, and when designed appropriately, can trade bias and variance for reducing estimation error. This paper starts with an exposition of our technique on the problem of estimating an unknown positive function when multiplied by a stationary random field. This example gives concrete evidence of the benefits of our local likelihood as compared to unweighted local likelihoods. We then discuss the difficult problem of estimating a bandwidth parameter that controls the amount of influence from distant observations. Finally we present a simulation experiment for estimating the local smoothness of a local Matérn random field when observing the field at random sampling locations in $[0, 1]^2$. The local Matérn is a fully nonstationary random field, has a closed form covariance, can attain any degree of differentiability or Hölder smoothness and behaves locally like a stationary Matérn. We include an appendix that proves the positive definiteness of this covariance function.

1. Introduction

Stationary random fields play a fundamental role in both theoretical and applied spatial statistics. Unfortunately, stationarity is often violated when working with real data. The issue is more pressing with the recent data deluge and high resolution sensing for which nonstationarity can be clearly visible. This presents a challenge for the spatial statistician who is interested in estimating and modeling dependency structure in random fields.

Even though stationarity is often criticized as being too simplistic, we believe it is still important for the understanding and development of nonstationary models. The reason is that any type of statistical estimation requires some sort of replication to "average over". In spatial statistics the data often comprise of one realization of a random field. In this case, the assumption of stationarity provides a type of replication that makes statistical estimation possible. For nonstationary random fields, however, the lack of any assumption leads to a breakdown in statistical estimation due to the absence of replication. This problem can be mitigated by adding assumptions on the nonstationary random field such as local stationarity, for example. The idea is that on small enough spatial scales, one hopes that the local dependency of the random field near a point is well approximated by some stationary random field. In this paper, we do not attempt a precise definition of local stationarity (see [5] for a definition in the time series literature). Instead, we take it for granted that such random fields exist (however one defines it) and enter into a discussion of how one might estimate the parameters of a local

* Corresponding author.
E-mail addresses: anderes@stat.ucdavis.edu, anderes@wald.ucdavis.edu (E.B. Anderes), stein@galton.uchicago.edu (M.L. Stein).

0047-259X/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2010.10.010

stationary approximation when observing a single realization of the nonstationary random field at dense (possibly uneven) observation locations. To accomplish these goals we develop a local likelihood approach for estimating local parameters.

There has been a significant amount of research on modeling and estimating nonstationarity in spatial statistics (see [8,13,14,10,6,3,12]). These papers present a variety of techniques for estimating the nonstationarity parameters but none use local likelihood techniques beyond simply dividing the observation locations into neighborhoods and fitting a stationary random field on each neighborhood (which we call *hard thresholding local likelihood estimates*). Typically two problems arise with this hard thresholding approach. First, the range of validity of a stationary approximation can be too small to contain enough local data to estimate it. Second, it can produce non-smooth local parameter estimates, which is undesirable in many cases. There do exist alternative weighted local likelihood techniques (see [7], for example), but unfortunately they are not applicable to random fields. These techniques utilize the independence structure of the data to decompose the log-likelihood as a sum, the summands of which are downweighted as a function of some spatial covariate. In the random field case, however, there is no independence and no such decomposition of the log-likelihood. Notice that the dependence structure cannot be ignored since the parameters of interest are what govern the dependency.

In this paper we present an exposition, through computation, simulation and some theory, of our version of local likelihood estimation. A large portion of the paper is devoted to the discussion of different ways of constructing and estimating the weights used in our local likelihood that downweight the influence of distant observations of the random field. We start in Section 2 with our definition of a weighted local likelihood and then immediately apply it to the problem of variance modulation in Section 3. This example is convenient since the local likelihood estimate has a closed form and one can derive the expected risk under a polynomial prior for the local variance. In Section 4 we study the problem of estimating the bandwidth parameter from data. Finally, in Section 5 we apply these techniques to the estimation of the local fractional index of a local Matérn random field when observing one realization of the field at uneven observation locations in $\mathbb{R}^2$.

2. Weighted local likelihood

Before we present our notion of local likelihood it will be advantageous to set some notation. We write a random field as $\{Z(t)\colon t \in \mathbb{R}^d\}$ or just Z when the domain of definition is clear from context. We distinguish two types of random field models: *global* and *local*. The only real distinction is nonstationary versus stationary, but the nomenclature is useful since we regard global models as the true nonstationary sampling distribution and local models as the local stationary approximations. We consider global models that are nonstationary random fields indexed by some function $\theta(t)$ that takes spatial arguments $t \in \mathbb{R}^d$ and returns values in some m-dimensional parameter space $\Theta \subset \mathbb{R}^m$. We call the function $\theta(\cdot)$ the *local parameter function* and denote the resulting global model by $G_{\theta(\cdot)}$. For each fixed $t_0 \in \mathbb{R}^d$, the parameter vector $\theta(t_0)$ determines a *local* random field model, denoted by $L_{\theta(t_0)}$, which is generally stationary and models the stochastic behavior of Z near some point t_0. Informally this means that $\mathcal{L}_{L_{\theta(t_0)}}\{Z(t_0+h)\colon |h| < \epsilon\} \approx \mathcal{L}_{G_{\theta(\cdot)}}\{Z(t_0+h)\colon |h| < \epsilon\}$ when ϵ is small ($\mathcal{L}_{L_{\theta(t_0)}}$ denotes the law of the finite dimensional distributions under model $L_{\theta(t_0)}$). For the remainder of the paper we will typically use the notation θ_0, as a shortened version of $\theta(t_0)$, to denote a parameter vector (as opposed to a local parameter function) that specifies a local stationary approximation. The local parameter function, on the other hand, will be denoted by $\theta(\cdot)$, $\theta(t)$ or just θ. We also use the convention that t_0 typically denotes a fixed spatial location at which we wish to estimate $\theta(t_0)$. This partly explains why we use θ_0 to denote a particular value of the local parameter function θ evaluated at t_0: we are estimating the parameter vector θ_0 at t_0.

For a concrete example, consider a fixed but unknown function $\sigma(t)\colon \mathbb{R}^d \to \mathbb{R}^+$ and a stationary Gaussian random field W with known autocovariance function and define $Z(t) = \sigma(t)W(t)$. In this case, $\sigma(t)$ is the local parameter function for Z and $G_{\sigma(\cdot)}$ denotes the true distribution of Z. If $\sigma(t_0 + h) \approx \sigma(t_0)$ when $|h|$ is small then the local model for Z at t_0, denoted by L_{σ_0}, is just the law of $\sigma_0 W(t)$. Presumably if one has enough data in a neighborhood of t_0 one could successfully estimate $\sigma(t_0)$ by fitting the model L_{σ_0} to the data. Notice there is an inherent bias-variance tradeoff when fitting L_{σ_0} locally to data: increasing the size of the local neighborhood reduces the variability of the estimate but increases the bias due to the inaccuracy of the stationary approximation. It is with local likelihoods that we seek to balance these two competing terms by smoothly downweighting the dependence of more distant observations.

To define the local likelihood estimation of $\theta(\cdot)$ at some fixed spatial location $t_0 \in \mathbb{R}^d$, suppose we have n observations $(t_1, z_1), \ldots, (t_n, z_n)$ of a single realization of a random field Z (the spatial locations are $t_j \in \mathbb{R}^d$ and the responses are $z_j = Z(t_j)$). For convenience we suppose the indices of $(t_1, z_1), \ldots, (t_n, z_n)$ are ordered by their distance to t_0. Therefore t_1 is the closest observation location to t_0 and t_n is the farthest. Finally, let $\mathcal{N}_{t_0,k}$ denote the set of k observations nearest to t_0 (so that $\mathcal{N}_{t_0,3} = \{z_1, z_2, z_3\}$, for example).

Now given weights $w_1, \ldots, w_n$ (possibly depending on $t_0, t_1, \ldots, t_n$ and a bandwidth parameter λ) we define the following weighted local likelihood:

$$\mathcal{W}_\lambda(\theta_0, t_0|\text{data}) \triangleq \sum_{k=1}^{n} w_k\big[\ell(\mathcal{N}_{t_0,k}|L_{\theta_0}) - \ell(\mathcal{N}_{t_0,k-1}|L_{\theta_0})\big] \tag{1}$$

where $\ell(\mathcal{N}_{t_0,k}|L_{\theta_0})$ is the log likelihood of the data in $\mathcal{N}_{t_0,k}$ under the model L_{θ_0} (note: we define $\ell(\mathcal{N}_{t_0,0}|L_{\theta_0})$ to be 0). Notice that $\mathcal{W}_\lambda$ orders and subsequently downweights the incremental changes in the stationary likelihood when adding

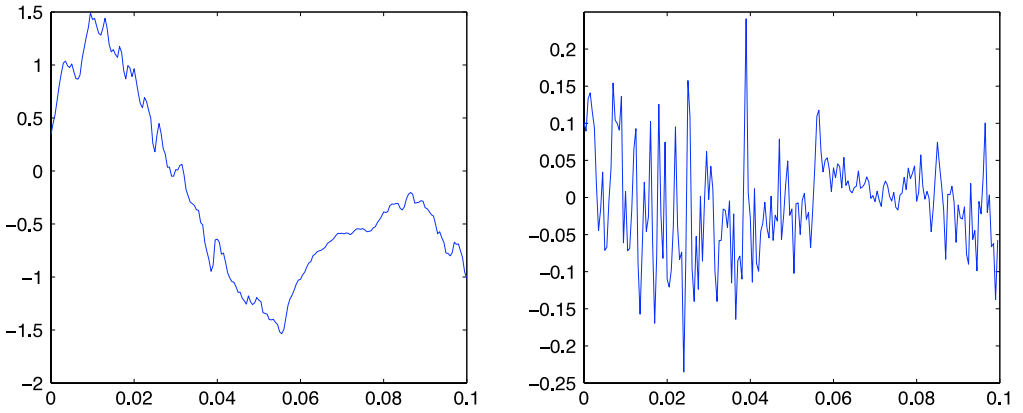

Fig. 1. *Left*: An example of a realization of the random field $\sigma(t)W(t)$, where W is a mean zero Gaussian random field with Matérn autocovariance function with parameters $(\tilde{\sigma}^2, \nu, \rho) = (1, 0.8, 0.2)$ and $\sigma(t) = 2\sin(t/0.015) + 2.8$. *Right*: The increments of $\sigma(t)W(t)$ on the 200 evenly spaced observation locations in the interval $[0, 0.1]$. See Fig. 2 for the estimates of $\sigma^2(t)$.

the observations one by one in order of their distance to t_0. Now our local likelihood estimate of $\theta(t_0)$ is defined as

$$\hat{\theta}_\lambda(t_0) \triangleq \arg\max_{\theta_0 \in \mathbb{R}^m} \mathcal{W}_\lambda(\theta_0, t_0|\text{data}).$$

An important feature of the estimate $\hat{\theta}_\lambda$ is that, at least for Gaussian random fields, the computational cost of $\mathcal{W}_\lambda(\theta_0, t_0|\text{data})$ is comparable to that of $\ell(\theta|\text{data})$ by either *updating* or *downdating* a Cholesky decomposition. Moreover, the estimate $\hat{\theta}_\lambda(t)$ will typically be a "smooth" function of t (when the weights are smooth) even when there is no natural additive or multiplicative structure on the parameter space $\Theta \subset \mathbb{R}^m$. Finally, we remark that $\hat{\theta}_\lambda$ depends on the bandwidth parameter λ through the weights w_k, which typically have the form $w_k \triangleq K((t_0 - t_k)/\lambda)$ for some smoothing kernel K.

Remark. If the weights $w_k = 1$ for all k, then the telescoping sum in (1) collapses and one recovers the full likelihood for the local stationary model. If weights w_k are 1 up to a fixed index k_0 and zero thereafter, the resulting estimate is simply the MLE under the stationary approximation fitted on the k_0-neighborhood of t_0. We call this the *hard thresholding local likelihood estimate* to emphasize the truncation nature of the weights.

3. Variance modulation

We start with a particularly simple example, already mentioned in the previous section: estimating the local variance of a Gaussian random field. Consider the estimation of the function $\sigma(t)\colon \mathbb{R} \to \mathbb{R}^+$ when observing $\sigma(t)W(t)$, where W is a mean zero Gaussian process on $\mathbb{R}$ with known Matérn parameters. This example is useful since the local maximum likelihood estimate has a closed form solution. We take advantage of this solution by deriving the expected risk under polynomial priors. The expected risk is used to give concrete evidence of a bias and variance trade-off for reducing estimation error. We also use the expected risk to explore the relationship between higher order kernels, bias of the resulting estimates and optimal bandwidth.

To derive the closed form solution of the local likelihood estimate we first need some notation. Let Σ_k denote the covariance matrix of the observations in $\mathcal{N}_{t_0,k}$, and let $\boldsymbol{z}_k \triangleq (z_1, \ldots, z_k)^T$ denote the responses of the data in $\mathcal{N}_{t_0,k}$. Notice that

$$\ell(\mathcal{N}_{t_0,k}|L_{\sigma_0}) = -\frac{\boldsymbol{z}_k^T \Sigma_k^{-1} \boldsymbol{z}_k}{2\sigma_0^2} - \frac{\log|\sigma_0^2 \Sigma_k|}{2} - \frac{k}{2}\log(2\pi)$$

and therefore

$$\mathcal{W}_\lambda(\sigma_0, t_0|\text{data}) = -\log\sigma_0 \sum_{k=1}^n w_k + \frac{1}{2\sigma_0^2}\sum_{k=1}^n w_k\left[\boldsymbol{z}_{k-1}^T \Sigma_{k-1}^{-1}\boldsymbol{z}_{k-1} - \boldsymbol{z}_k^T \Sigma_k^{-1}\boldsymbol{z}_k\right] + c.$$

The maximum occurs at

$$\hat{\sigma}_\lambda^2(t_0) \triangleq \frac{\sum_{k=1}^n w_k\left[\boldsymbol{z}_k^T \Sigma_k^{-1}\boldsymbol{z}_k - \boldsymbol{z}_{k-1}^T \Sigma_{k-1}^{-1}\boldsymbol{z}_{k-1}\right]}{\sum_{k=1}^n w_k} \tag{2}$$

which is the local maximum weighted likelihood estimate of $\sigma^2(t)$ at t_0.

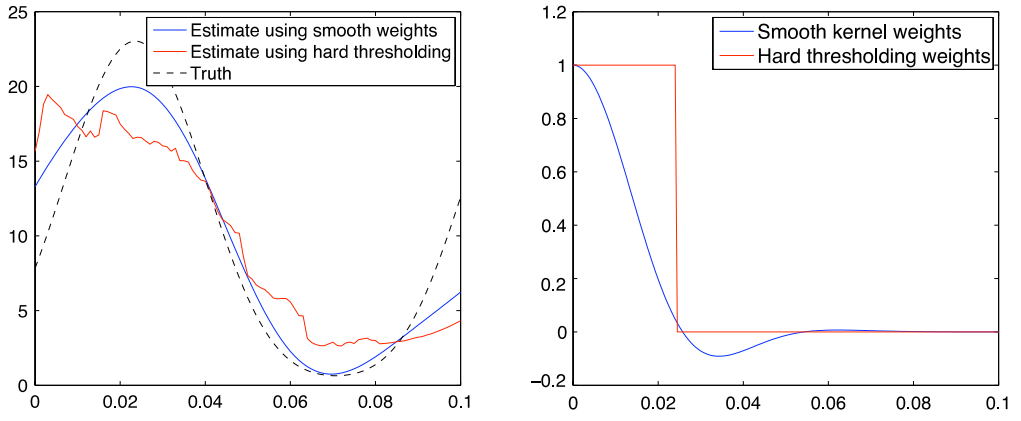

Fig. 2. *Left*: Plot of the true local parameter function $\sigma^2(t)$ (dashed) along with two estimates of $\sigma^2(t)$ (from the realization shown in Fig. 1), one using smooth weights (blue) and the other using hard thresholding weights (red). *Right*: The plot of the smooth weights (blue) and the hard thresholding weights (red) as a function of lag $|t_0 - t_k|$ used in the two estimates of $\sigma^2(t)$ shown in the left panel. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 1 shows a simulation of the random field $\sigma(t)W(t)$ at 200 evenly spaced observation locations in $[0, 0.1]$, where W is a mean zero Gaussian random field with known Matérn autocovariance function with parameters $(\tilde{\sigma}^2, \nu, \rho) = (1, 0.8, 0.2)$ (using the parameterization given on page 50 of [18], where we use $\tilde{\sigma}^2$ to denote the overall variance to avoid confusion with the local parameter function $\sigma^2(t)$) and unknown $\sigma(t) = 2\sin(t/0.015) + 2.8$. Notice that even though the local variance changes significantly throughout the observation region, it is difficult to see since the range parameter $\rho = 0.2$ is relatively large compared with the observation region. However, if one looks at increments of $\sigma(t)W(t)$, shown in the right-hand diagram, the change in local variance becomes clearly visible. In Fig. 2 we plot the local likelihood estimate $\hat{\sigma}^2_\lambda(t_0)$ as a function of t_0 (blue line in the left-hand diagram) as compared with the hard thresholding local likelihood estimate (red line in the left-hand diagram) and the true local parameter function $\sigma^2(t)$ (dashed line in the left-hand diagram). The smooth weights take the form $w_k = K(|t_0 - t_k|/\lambda)$, where K is defined as $K(t) \triangleq e^{-t^2/2}(15 - 10t^2 + t^4)/8$, which is kernel K_6 presented in [21] (and also used in Section 3.1) and is plotted as a function of lag $|t_0 - t_k|$ in the right-hand diagram of Fig. 2 (blue) along with the hard thresholding weights (red). The bandwidth used for both estimators is obtained by minimizing the Kullback–Leibler divergence of the estimated global model $G_{\hat{\sigma}^2_{\lambda(\cdot)}}$ to the truth at the observation locations. Notice that using smooth weights, the local likelihood estimate not only can reduce estimation error but can also yield smooth estimates of the local parameter function σ^2.

3.1. Expected risk for random nusance parameters

Now we study the behavior of $\hat{\sigma}^2_\lambda(t_0)$ at some fixed spatial location t_0. The main goal of this section is to derive an analytic expression for the expected risk and expected (over the prior) bias squared for our local likelihood estimates under a prior model on nuisance parameters that govern the local deviation of the true model from stationarity. The analytic expressions are given in formula (4). We finish the section with a numerical example that explores the relationship between bias, risk, bandwidth and parameters of the stationary random field.

To give a detailed analysis of the estimate $\hat{\sigma}^2_\lambda(t_0)$ we suppose the true local parameter function $\sigma(t)$, near t_0, can be expressed with a finite Taylor expansion

$$\sigma(t) = c_0 + \sum_{p=1}^{N} c_p(t - t_0)^p. \tag{3}$$

The zero order coefficient, $c_0 = \sigma(t_0)$, is the parameter of interest and the higher order coefficients $c_1, \ldots, c_N$ are nuisance parameters. The main source of bias in a hard thresholding local likelihood estimate comes from the parameters $\boldsymbol{c} \triangleq (c_1, \ldots, c_N)$. This section shows concrete evidence of the stated goal of the weighted local likelihood – to balance bias and variance – by using higher order kernels for reducing estimation bias due to $\boldsymbol{c}$. In particular, we suppose the vector $\boldsymbol{c}$ is chosen randomly from some prior density $\pi(\boldsymbol{c})$. Under this assumption we establish a decomposition of the expected risk (averaging over the model π on $\boldsymbol{c}$ and resampling of the data) in terms of bias and variance. Some components of the bias terms are shown to converge to zero under fixed domain asymptotics.

For simplicity of exposition we suppose the prior density π factors as $\pi(\boldsymbol{c}) = \pi_1(c_1)\cdots\pi_N(c_N)$, $E_{\pi_j} c_j = 0$ and $E_{\pi_j} c_j^3 = 0$ for all $1 \le j \le N$. In Appendix C we derive the following decomposition for the expected risk of $\hat{\sigma}^2_\lambda(t_0)$ (averaging over the model π on $\boldsymbol{c}$):

$$E\left\{\hat{\sigma}^2_\lambda(t_0) - \sigma^2(t_0)\right\}^2 = \underbrace{\sum_{p_1,p_2,p_3,p_4=0}^{N} E_\pi\left[c_{p_1}c_{p_2}c_{p_3}c_{p_4}\right]B^{p_1,p_2,p_3,p_4}}_{\text{variance}} + \underbrace{E_\pi\left\{\sum_{p_1,p_2=1}^{N} c_{p_1}c_{p_2}B^{p_1,p_2}\right\}^2 + 4\,c_0^2\,E_\pi\left\{\sum_{p=1}^{N} c_p B^{0,p}\right\}^2}_{\text{bias}^2\text{ terms (averaging over the prior }\pi)} \tag{4}$$

where each of the three terms is positive and

$$B^{0,p} = \frac{1}{\sum_{j=1}^{n} w_j}\sum_{k=1}^{n} w_k(t_k - t_0)^p$$

(the terms B^{p_1,p_2,p_3,p_4} and B^{p_1,p_2} depend on the weights w_k and the covariances Σ_k. The full definition is given in Appendix C). The significance of this decomposition is to notice that if we design the weights w_k so that $w_k \triangleq K((t_0-t_k)/\lambda)$, $\int_{\mathbb{R}} K(t)\mathrm{d}t = 1$ and $\int_{\mathbb{R}} K(t)t^p\mathrm{d}t = 0$ for some $p > 1$, then under mild conditions on the sampling locations t_k and K

$$B^{0,p} = \frac{\sum_{k=1}^{n} K\left[\frac{t_0-t_k}{\lambda}\right](t_k - t_0)^p}{\sum_{j=1}^{n} K\left[\frac{t_0-t_j}{\lambda}\right]} \longrightarrow 0$$

as $n \to \infty$, $\lambda \to 0$ and the t_k's get more dense in a bounded region near t_0. In particular, if one uses higher order kernels, the last term in (4) can be made arbitrarily small under fixed domain asymptotics. Note: it is only necessary to consider $\lambda \to 0$ when our observation locations t_k stay bounded and K has infinite support (which is the case we will consider in this section).

Numerical results. For exposition we consider a class of Gaussian-based higher order kernels K_{2r} defined in [21] as

$$K_{2r}(t) \triangleq Q_{2r-2}(t)\phi(t) \tag{5}$$

where $\phi(t) = \mathrm{e}^{-t^2/2}/\sqrt{2\pi}$ is the Gaussian kernel, $Q_{2r-2} = \{2^{r-1}(r-1)!\}^{-1}H_{2r-1}(t)/t$, and H_j denotes the jth normalized Hermite polynomial defined by $H_j(t) = (-1)^j\phi^{(j)}(t)/\phi(t)$. These kernels have the required property that $\int_{\mathbb{R}} K(t)t^p\mathrm{d}t = 0$ for all $0 < p < 2r$. These are not the only kernels with this property but we use them since the resulting estimates $\hat{\sigma}^2_\lambda(t)$ will be very smooth in t. We compare all our results to the hard thresholding weights defined as $w_k \triangleq \mathbf{1}_{B_\lambda(t_0)}(t_k)$, where $\mathbf{1}_A$ is the indicator of the set $A \in \mathbb{R}$ and $B_\lambda(t_0)$ is a ball of radius λ centered at t_0.

To investigate the performance of the local likelihood estimates, we computed the expected risk and bias using Eq. (4) under the assumption that $t_0 = 1/2$, $\sigma(t_0) = 2$, and $c_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,4)$ for $j = 1,\ldots,4$ (so that $N = 4$ in (3)). The stationary random field W is assumed to be a mean zero, unit variance Matérn Gaussian random field and the observation locations of $\sigma(t)W(t)$ are 100 evenly spaced points in $[0, 1]$. For each Matérn parameter value ν and ρ (on an even grid in $(0, 1]$ with spacing $1/10$ starting at 0.1) we computed the percent improvement over hard thresholding when using kernel K_6 defined in (5) to generate the weights w_k. The bandwidth for both K_6 and hard thresholding were chosen using the Kullback–Leibler oracle criterion. We found a minimum of 17% improvement in expected risk over hard thresholding uniformly over the computed ρ and ν parameters (mean improvement of 17.3%). Moreover, the improvement in expected bias squared can reach 67% (mean improvement of 58.2%).

In Fig. 3 we plot the expected risk and expected bias squared as a function of bandwidth for the estimate $\hat{\sigma}^2_\lambda(1/2)$ using the kernels K_2, K_4, K_6, K_8 and hard thresholding. The estimate is based on the Matérn random field with parameters $\nu = \rho = 0.8$ with 150 even observation locations in $[0, 1]$. We use the same polynomial prior as in the last paragraph. Notice the significant improvement in risk and bias squared using kernels K_4, K_6 and K_8. Also notice that the improvement in risk and bias occurs at larger bandwidths for higher order kernels. Finally, we remark that the discrete steps seen in the risk and bias plots of the hard thresholding estimates are due to the fact that as bandwidth increases there are discrete jumps in the number of local points used in the hard thresholding estimates.

4. Bandwidth selection

The local likelihood estimates $\hat{\theta}_\lambda$ can vary dramatically depending on the choice of smoothing parameter λ. Since the theoretical derivation of risk in anything but the simplest setting is extremely difficult it is necessary to develop estimates of λ from data. Classical methods such as cross-validation may fail when the data is comprised of a single realization of a nonstationary random field. The data are very highly correlated so that a "leave out" prediction is problematic since the data

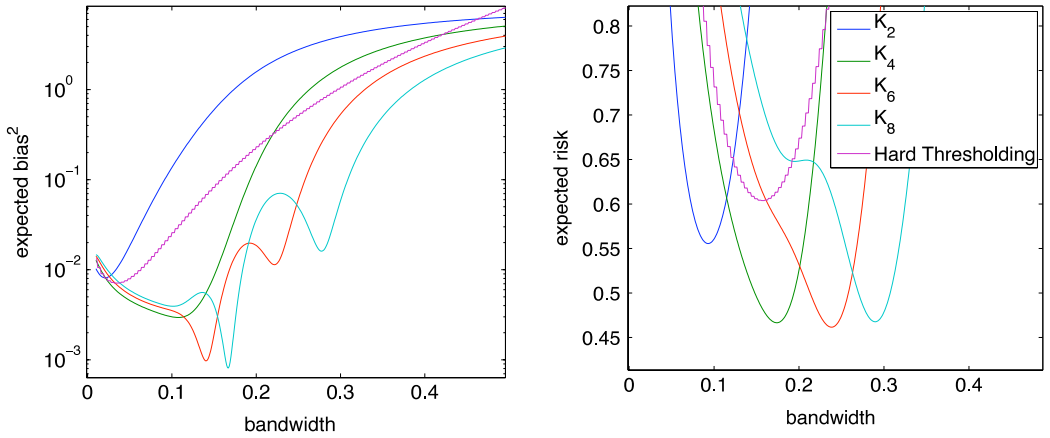

Fig. 3. Expected risk (right) and expected bias squared (left) of $\hat{\sigma}^2_\lambda(1/2)$ plotted against bandwidth for 5 different kernels: K_2, K_4, K_6, K_8 and hard thresholding. The estimate $\hat{\sigma}^2_\lambda(1/2)$ is based on 150 even observation locations in [0, 1] and the Matérn autocovariance function with $\nu = \rho = 0.8$.

"left out" is highly correlated with the data "left in". In this section we give a heuristic for constructing a reasonable estimate $\hat{\lambda}$. We then present our interpretation of this heuristic in terms of two different estimates of λ. At the end of this section we present some numerical simulations to illustrate their behavior. We make no claim of optimality or any theoretical justification other than heuristics.

Our heuristic for the bandwidth estimator says that one should choose the bandwidth to maximize the spatial variability in $\hat{\theta}_\lambda$ beyond what is expected from the realization of the random field itself. The idea is that there are two sources of spatial variability in $\hat{\theta}_\lambda$. The first is the spatial variability from the true local parameter function θ which, of course, $\hat{\theta}_\lambda$ is trying to estimate. The second is the spatial variation coming from the particular realization of the random field itself. Notice that the smaller the bandwidth λ the more the spatial variation in $\hat{\theta}_\lambda$ is due to the random field realization, and less to the variability of θ. To describe our mathematical interpretation of this heuristic, we need some notation. Let $\mathcal{P}(\hat{\theta}_\lambda)$ be a measure of spatial variation in $\hat{\theta}_\lambda$. For example, when the local parameter θ is univariate so that $\theta(t)$ maps $\mathbb{R}^d$ to $\mathbb{R}$, a natural choice might be $\mathcal{P}(\hat{\theta}_\lambda) \triangleq \int_{\mathbb{R}^d} |\nabla\hat{\theta}_\lambda(t)|^2 dt$. Let $\overline{\theta}$ be the MLE of the local parameter function obtained by assuming the global model is stationary (i.e. by assuming $\theta(t)$ is a constant function of t). Now consider simulating an independent realization of the data under the stationary random field model using the estimated parameter $\overline{\theta}$. Let $E_{\overline{\theta}}\mathcal{P}(\hat{\theta}_\lambda)$ denote the expected value of $\mathcal{P}(\hat{\theta}_\lambda)$, where $\hat{\theta}_\lambda$ is applied to the new realization of the stationary random field. The idea is that $E_{\overline{\theta}}\mathcal{P}(\hat{\theta}_\lambda)$ quantifies the spatial variation in $\hat{\theta}_\lambda$ that is exclusively due to the random field itself (since the true local parameter function is constant). Now we estimate λ as follows:

$$\hat{\lambda}_1 \triangleq \arg\max_\lambda \frac{\mathcal{P}(\hat{\theta}_\lambda) - E_{\overline{\theta}}\mathcal{P}(\hat{\theta}_\lambda)}{\mathrm{sd}_{\overline{\theta}}\mathcal{P}(\hat{\theta}_\lambda)}. \tag{6}$$

In Eq. (6) we divided by $\mathrm{sd}_{\overline{\theta}}\mathcal{P}(\hat{\theta}_\lambda)$, which denotes the standard deviation of $\mathcal{P}(\hat{\theta}_\lambda)$ under the stationary model given by $\overline{\theta}$, to improve comparison across λ.

Our second bandwidth selector is similar to the first, the main difference being that the measure of spatial variation $\mathcal{P}(\hat{\theta}_\lambda)$ is replaced with $\sum_t(\mathcal{W}_\lambda(\hat{\theta}_\lambda(t), t) - \mathcal{W}_\lambda(\overline{\theta}, t))$. This is recognized as a type of local likelihood ratio test statistic (summed over the spatial variable t) that compares the estimate $\hat{\theta}_\lambda$ to the stationary fit $\overline{\theta}$. Then, after adjusting by the expected behavior of this quantity under the stationary fit $\overline{\theta}$, we get the following estimate of bandwidth:

$$\hat{\lambda}_2 = \arg\max_\lambda \frac{\sum_t(\mathcal{W}_\lambda(\hat{\theta}_\lambda(t), t) - \mathcal{W}_\lambda(\overline{\theta}, t)) - E_{\overline{\theta}}\left[\sum_t(\mathcal{W}_\lambda(\hat{\theta}_\lambda(t), t) - \mathcal{W}_\lambda(\overline{\theta}, t))\right]}{\mathrm{sd}_{\overline{\theta}}\left[\sum_t(\mathcal{W}_\lambda(\hat{\theta}_\lambda(t), t) - \mathcal{W}_\lambda(\overline{\theta}, t))\right]}. \tag{7}$$

In our implementation of both (6) and (7) we use simulations to estimate the expected value and standard deviation under the stationary fit $\overline{\theta}$. Unfortunately these simulations are rather slow, and faster methods will be needed for data sets larger than a few thousand.

To illustrate $\hat{\lambda}_1$ and $\hat{\lambda}_2$ we provide some simulations in the case of variance modulation (see Section 3). Our first simulation is a single realization of $\sigma(t)W(t)$ at 1000 evenly spaced sampling locations in the interval $[0, 0.1]$, where $\sigma(t) = 2\sin(t/0.015) + 2.8$ is unknown and W is a mean zero stationary Gaussian random field with known Matérn autocovariance parameters $(\tilde{\sigma}^2, \nu, \rho) = (1, 0.5, 0.5)$ (again, we use $\tilde{\sigma}^2$ to denote the overall variance to avoid confusion

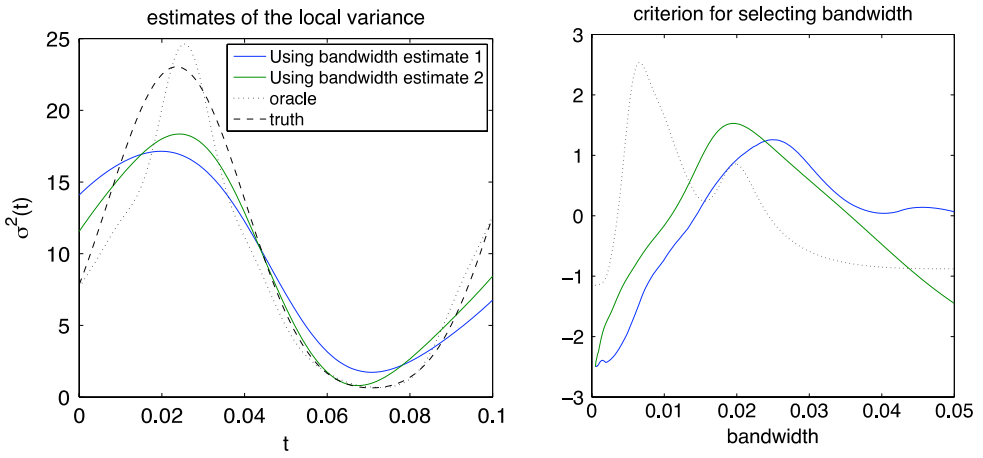

Fig. 4. *Left*: σ^2 (dashed), $\hat{\sigma}^2_{\hat{\lambda}_1}$ (blue), $\hat{\sigma}^2_{\hat{\lambda}_2}$ (green) and $\hat{\sigma}^2_{\lambda_{\text{orc}}}$ (dotted) when observing a single realization of $\sigma(t)W(t)$, where $\sigma(t) = 2\sin(t/0.015) + 2.8$ is unknown and W is a stationary Gaussian random field with known Matérn parameters $(\tilde{\sigma}^2, \nu, \rho) = (1, 0.5, 0.5)$, at 1000 even sampling locations in $[0, 0.1]$. *Right*: Plots of the standardized criterion profiles which, when maximized, give λ_{orc} (dotted), $\hat{\lambda}_1$ (blue) and $\hat{\lambda}_2$ (green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

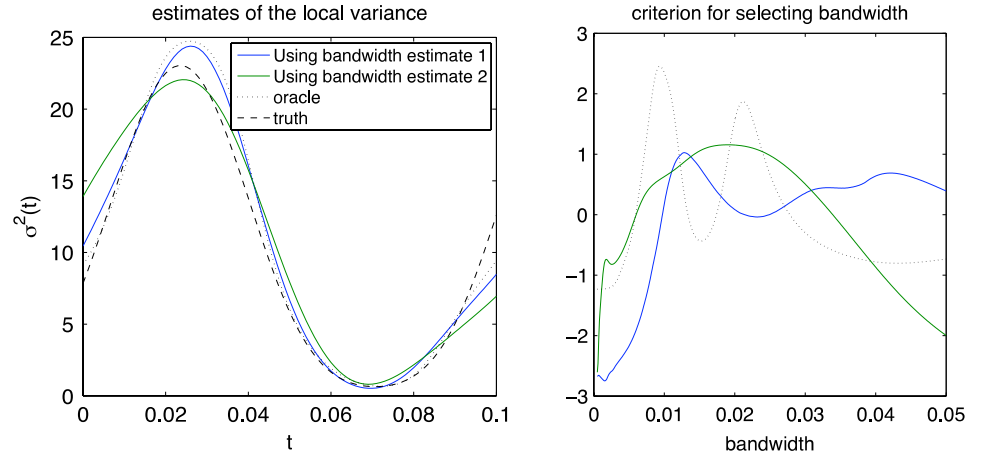

Fig. 5. Estimates $\hat{\sigma}^2_{\hat{\lambda}_1}$, $\hat{\sigma}^2_{\hat{\lambda}_2}$ and $\hat{\sigma}^2_{\lambda_{\text{orc}}}$ of σ^2 (left) along with the criterion profiles for estimating λ (right). The simulation parameters are the same as in Fig. 4 with the exception that $\nu = 1$ instead of $\nu = 0.5$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with the local parameter function $\sigma^2(t)$). The left plot of Fig. 4 shows the true local parameter function $\sigma^2(t)$ (dashed line) along with three different estimates $\hat{\sigma}^2_{\hat{\lambda}_1}$ (blue), $\hat{\sigma}^2_{\hat{\lambda}_2}$ (green) and $\hat{\sigma}^2_{\lambda_{\text{orc}}}$ (dotted line). The estimate $\hat{\sigma}^2_{\lambda_{\text{orc}}}$ is the oracle estimate, which estimates the smoothness parameter by

$$\lambda_{\text{orc}} \triangleq \arg\max_{\lambda} \left\{ D(G_{\hat{\sigma}^2_{\lambda}} \parallel G_{\sigma^2})^{-1} \right\}, \tag{8}$$

where G_{σ^2} denotes the true global model for $\sigma(t)W(t)$ at the observation locations and $D(\cdot \parallel \cdot)$ is the Kullback–Leibler divergence. The right plot of Fig. 4 shows the criterion profiles that are maximized for estimating the bandwidths $\hat{\lambda}_1$ (blue), $\hat{\lambda}_2$ (green) and λ_{orc} (dotted). The profiles in this diagram are standardized by their average and standard deviation so they can be compared on the same scale. The parameters of our second simulation are exactly the same with the exception that now $\nu = 1$ instead of $\nu = 0.5$. The results of this simulation are shown in Fig. 5, where again, the left plot shows σ^2 (dashed), $\hat{\sigma}^2_{\hat{\lambda}_1}$ (blue), $\hat{\sigma}^2_{\hat{\lambda}_2}$ (green) and $\hat{\sigma}^2_{\lambda_{\text{orc}}}$ (dotted). The right plot, again shows the standardized profiles for estimating λ.

For both simulations, $\hat{\lambda}_1$ and $\hat{\lambda}_2$ do a good job of estimating an appropriate bandwidth. The modes of the data-driven criterion profiles seem to come reasonably close to at least one of the modes in the oracle profiles. It is unclear to us, at this moment, why the oracle profiles in both simulations have two modes. Regardless, we think it is a good sign that the data-driven profile modes are close to at least one oracle mode. Moreover the resulting estimates $\hat{\sigma}^2_{\hat{\lambda}_1}$ and $\hat{\sigma}^2_{\hat{\lambda}_2}$ give good agreement with the true σ^2, although some under-fitting can be seen in Fig. 4.

5. Estimating a local fractional index in 2 dimensions

In this section we take full advantage of our local likelihood technique and estimate a spatially varying smoothness parameter when observing one realization of the field at random observation locations in $[0, 1]^2$. In our simulation we suppose that there is *a priori* information that the observed random field is locally isotropic with a known local variance and local range that do not vary spatially but with an unknown spatially varying smoothness parameter, denoted by ν_t (we will use the notation ν_t instead of $\nu(t)$ to make the following formulas more readable). There are many nonstationary random fields that have a spatially varying smoothness parameter (see [15,20,4,17] for example). For our simulation we use a nonstationary local Matérn, presented below, that has a closed form covariance, can attain any degree of differentiability or Hölder smoothness and behaves locally like a stationary Matérn.

We start by deriving our nonstationary covariance structure with spatially varying local parameters. Let $\mathcal{M}_\nu(\cdot) \triangleq |\cdot|^\nu \mathcal{K}_\nu(\cdot)$, where $\mathcal{K}_\nu$ is the modified Bessel function of the second kind, and let ν_t, σ_t and α_t denote spatially varying Matérn parameters. The local parameter function $\nu_t\colon \mathbb{R}^d \to \mathbb{R}^+$ determines the local smoothness, $\sigma_t^2\colon \mathbb{R}^d \to \mathbb{R}^+$ determines a local variance and $\alpha_t\colon \mathbb{R}^d \to PD_d(\mathbb{R})$ determines a local geometric anisotropy that maps $\mathbb{R}^d$ into the set of $d \times d$ positive definite matrices with real entries. Now define the following covariance function

$$R(s,t) = \sigma_t \sigma_s \det\left(\alpha_{st}^{-1/2}\right) \mathcal{M}_{\nu_{st}}\left(\left|\alpha_{st}^{-1/2}(t-s)\right|\right) \tag{9}$$

where $\alpha_{st} \triangleq (\alpha_t + \alpha_s)/2$, $\nu_{st} \triangleq (\nu_t + \nu_s)/2$. In Appendix A we show that the nonstationary covariance function R is positive definite (which originally appeared in the technical report [19]).

One problematic feature concerning R is that it is difficult to separate the interpretation of the local scale $\det(\alpha_t^{-1/2})$ and the smoothness parameter ν_t. This is because both $\det(\alpha_t^{-1/2})$ and ν_t affect the local lag for which the correlation becomes close to zero. Therefore it is desirable to re-parameterize (9) to give distinct interpretations of the three local parameter functions: variance, range and smoothness. To simplify the exposition we suppose there is no geometric anisotropy, so that α_t maps $t \in \mathbb{R}^d$ into the set of positively scaled $d \times d$ identity matrices. In particular let $\rho_t\colon \mathbb{R}^d \to \mathbb{R}^+$ and define

$$K(s,t) = \sigma_s \sigma_t \left[\frac{(\rho_s^2/4\nu_s)^{d/2}}{\Gamma(\nu_s)2^{\nu_s-1}}\right]^{1/2} \left[\frac{(\rho_t^2/4\nu_t)^{d/2}}{\Gamma(\nu_t)2^{\nu_t-1}}\right]^{1/2} \left[\frac{\rho_s^2}{8\nu_s} + \frac{\rho_t^2}{8\nu_t}\right]^{-d/2} \mathcal{M}_{\nu_{st}}\left[\left(\frac{\rho_s^2}{8\nu_s} + \frac{\rho_t^2}{8\nu_t}\right)^{-1/2} |s-t|\right].$$

The advantage of this covariance function (which is a re-parameterization and simplification of (9)) is that when both s, t are near some fixed t_0, and the local parameter functions ν_t, ρ_t and σ_t^2 are sufficiently smooth, we have that

$$K(s,t) \approx \frac{\sigma_{t_0}^2}{\Gamma(\nu_{t_0})2^{\nu_{t_0}-1}} \mathcal{M}_{\nu_{t_0}}(2\sqrt{\nu_{t_0}}|s-t|/\rho_{t_0}).$$

This is recognized as an isotropic Matérn autocovariance with the parameterization found on page 50 of [18]. Therefore σ_t^2 has the interpretation as the local variance, ν_t as the local smoothness, and ρ_t has the interpretation of the local range. For the simulations presented later in this section the local range and variance are set to be constant functions of $t \in \mathbb{R}^d$, so that $\rho_t \equiv \rho$, $\sigma_t \equiv \sigma$ and K simplifies to

$$K(s,t) = \sigma^2 \frac{\nu_s^{d/4}\nu_t^{d/4}}{\nu_{st}^{d/2}\left\{\Gamma(\nu_s)2^{\nu_s-1}\right\}^{1/2}\left\{\Gamma(\nu_t)2^{\nu_t-1}\right\}^{1/2}} \mathcal{M}_{\nu_{st}}\left[\frac{2\sqrt{\nu_s\nu_t}}{\rho\sqrt{\nu_{st}}}|s-t|\right]. \tag{10}$$

When estimating a local parameter function in two dimensions, bias becomes prominent near the boundary of the observation region. In our simulation we attempted to design the local likelihood weights to automatically mitigate this bias near the boundary. We can loosely motivate our weights by the theory of estimating equations. Consider the scenario where one wants to estimate a univariate local parameter function θ (so that it maps $\mathbb{R}^d$ into $\mathbb{R}$) at a fixed spatial location t_0. Now one can rewrite $\frac{\mathrm{d}}{\mathrm{d}\theta_0}\mathcal{W}_\lambda(\theta_0, t_0|\text{data})$ as $\sum_{k=1}^n w_k S_k(\theta_0)$, where $S_k(\theta_0) = \frac{\mathrm{d}}{\mathrm{d}\theta_0}\log f_{\theta_0}(z_k|\mathcal{N}_{t_0,k-1})$ is the score of the conditional log likelihood of the kth nearest observation to t_0 conditional on $k-1$ nearer observations. Therefore the local likelihood estimate $\hat{\theta}(t_0)$ is implicitly defined as solving

$$\sum_{k=1}^n w_k S_k(\hat{\theta}(t_0)) = 0,$$

where both S_k and w_k depend on t_0. This suggests designing the weights w_k to satisfy the unbiasedness constraint $\sum_{k=1}^n w_k E_\theta\{S_k(\theta(t_0))\} = 0$ for all local parameter functions θ. Notice that E_θ denotes the expected value with respect to the true nonstationary model $G_{\theta(\cdot)}$. To approximate the unbiasedness requirement $\sum_{k=1}^n w_k E_\theta\{S_k(\theta(t_0))\} = 0$ we notice that $E_\theta\{S_k(\theta(t_0))\}$ is a function of $t_0, t_1, \ldots, t_k$ and converges to zero as $t_k \to t_0$. Therefore, the first order Taylor expansion as the points $(t_1, \ldots, t_k)$ shrink towards t_0 is $E_\theta\{S_k(\theta(t_0))\} \approx \sum_{j=1}^k c_{j,k} \cdot (t_0 - t_j)$, where $c_{j,k} \triangleq \lim_{(t_1,\ldots,t_k)\to(t,\ldots,t)} \nabla_{t_i} E_\theta\{S_k(\theta(t_0))\}$. We make a further approximation by assuming $c_{j,k} = 0$ when $j \neq k$ and $c_{k,k}$ is a constant function of k. It is unclear whether this approximation holds in general; however, we give an informal derivation that the assumption is true in the

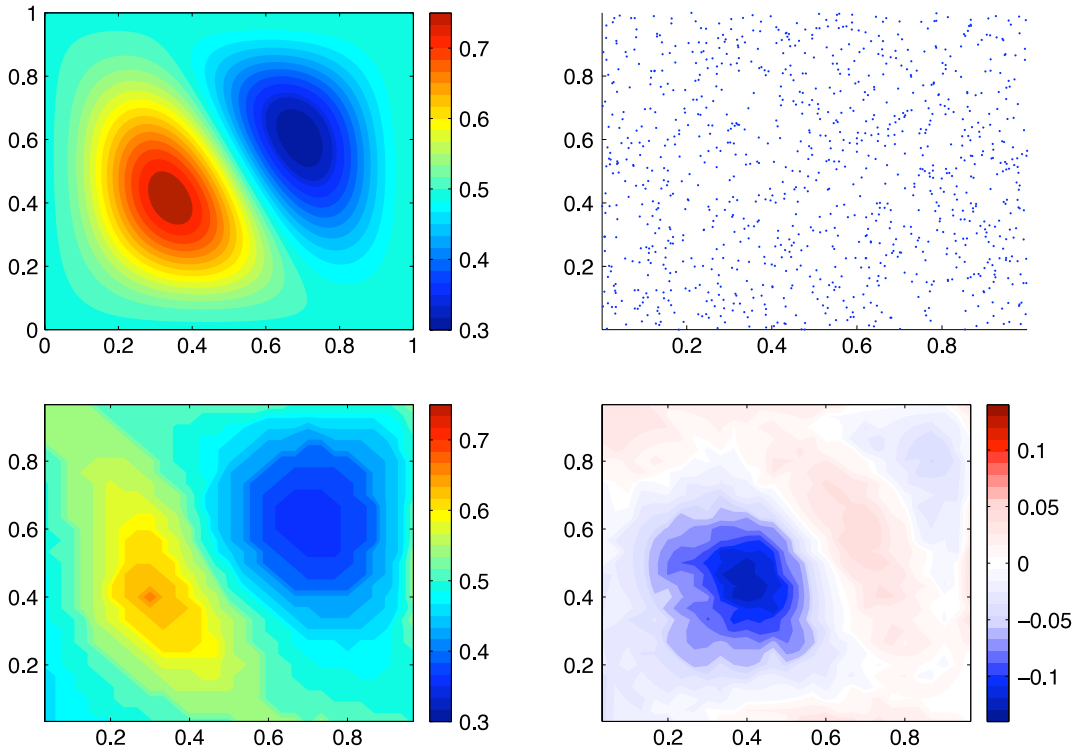

Fig. 6. *Top left*: The true local parameter function ν_t used in the simulation. *Top right*: The 1000 observation locations. *Bottom left*: The estimate of the local parameter function ν_t on the grid of estimation locations. *Bottom right*: The error $\hat{\nu}_t - \nu_t$.

case of variance modulation in Appendix B. These approximations allow one to approximate the unbiasedness condition $\sum_{k=1}^n w_k E_\theta\{S_k(\theta(t_0))\} = 0$ by $\sum_{k=1}^n w_k(t_0 - t_k) = (0,0)^T$ (when $t_k \in \mathbb{R}^2$). In addition, we believe that faraway observations (from t_0) are increasingly uninformative for estimating $\theta(t_0)$. Therefore, the weights w_k should approach 0 as $|t_k - t_0| \to \infty$. This is our motivation for the following variational characterization of the local likelihood weights for estimating $\theta(t_0)$:

$$\text{minimize} \sum_{k=1}^n w_k^2 \exp[|t_0 - t_k|^2/(2\lambda^2)] \quad \text{subject to} \begin{cases} \sum_{k=1}^n w_k = 1, \\ \sum_{k=1}^n w_k(t_0 - t_k) = (0,0)^T. \end{cases} \tag{11}$$

Notice that this particular variational characterization is essentially the same as in the nonparametric regression setting (see [11]).

Remark. It is interesting to note that the solution of (11) is in the form

$$w_k = [a + b \cdot (t_0 - t_k)] \exp[-|t_0 - t_k|/(2\lambda^2)]$$

where a, b are essentially the Lagrange multipliers of the variational problem. Therefore the solution of w_k in (11) can be computed quickly by inversion of a 2×2 matrix for each λ.

To test our estimate of ν_t we simulated one realization of a mean zero Gaussian random field at 1000 random observation locations in $[0,1]^2$ using the autocovariance function (10), where $\rho = 10$ and $\sigma = 1$ are assumed to be known. We used Cholesky downdating to compute the log-likelihoods $\ell(\mathcal{N}_{t_0,k}|\nu)$ for $k = 1, \ldots, 150$ (150 was chosen so the computations could be done in reasonable time). A plot of the local parameter function ν_t and the sampling locations are shown on the top row of diagrams in Fig. 6. To adjust for the fact that there is ∼75% less data at the corners, and ∼50% less near the edges, we scaled the bandwidth by 2 near the corner and by $\sqrt{2}$ near the edges (by linear interpolation). The bottom left plot of Fig. 6 shows the estimate $\hat{\nu}_t$ on a square grid of estimation locations in $[0,1]^2$, using the bandwidth estimator $\hat{\lambda}_1$ with smoothness measure $\mathcal{P}(\hat{\nu}) \triangleq \int_{\mathbb{R}^2} |\Delta \log \hat{\nu}|^2$. The bottom right plot shows the error: $\hat{\nu}_t - \nu_t$. One can see that the estimate $\hat{\nu}_t$ does a good job of recovering the main features of the true local parameter function ν_t, with some modest underestimation near the peak and overestimation near the trough. In Fig. 7, we compare our estimate with and without bias correcting weights. The right plot shows the estimate of ν_t without the bias correction (11) and instead uses weights $w_k = \exp[-|t - t_k|/(2\lambda^2)]$

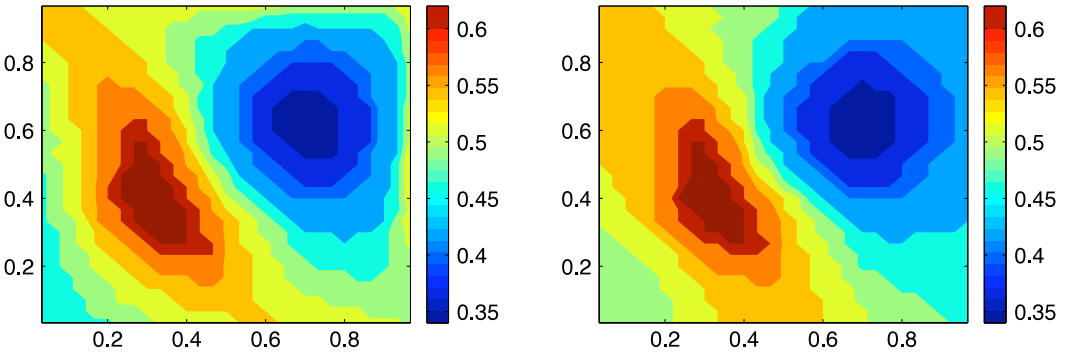

Fig. 7. *Left*: The estimate $\hat{\nu}_t$ using bias correcting weights w_k defined by Eq. (11). *Right*: The estimate $\hat{\nu}_t$ without bias correcting weights.

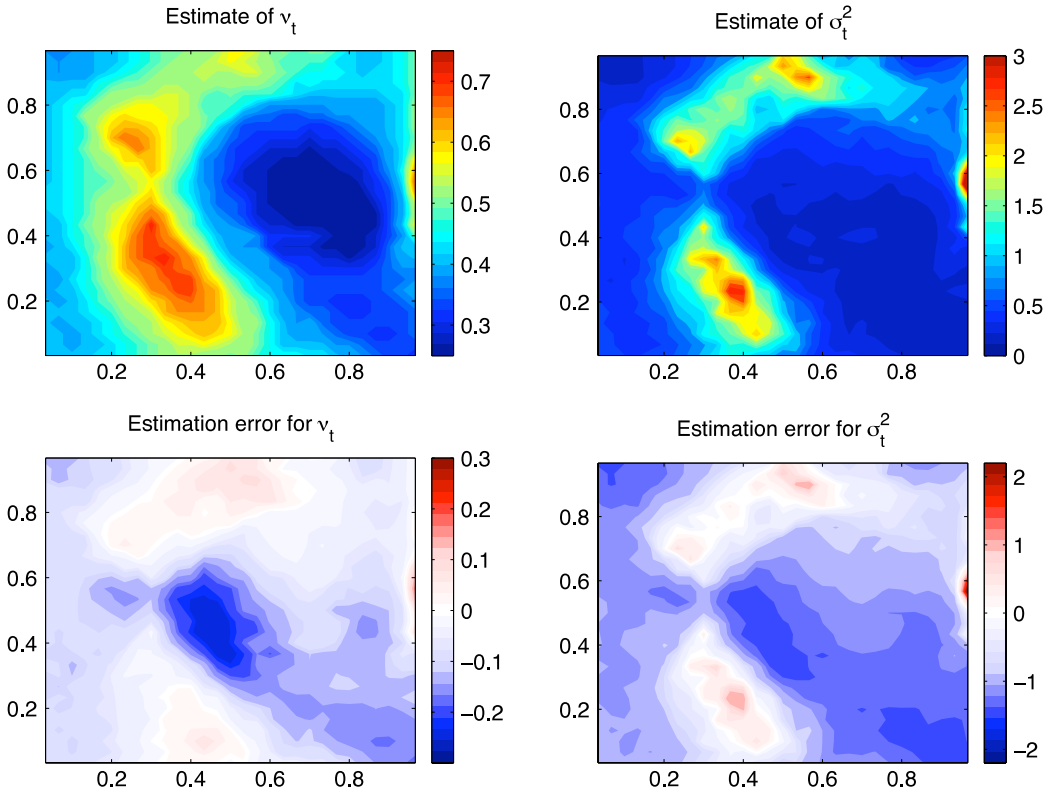

Fig. 8. *Top Left*: The estimate $\hat{\nu}_t$. *Top Right*: The estimate $\hat{\sigma}_t^2$. *Bottom Left*: The estimation error $\hat{\nu}_t - \nu_t$. *Top Right*: The estimation error $\hat{\sigma}_t^2 - \sigma_t^2$.

(using the same bandwidth estimated with the bias correction). It is clear from Fig. 7 that the weights designed by (11) can successfully mitigate bias near the boundary of the observation region (specifically in the upper right corner and left edge).

For our last example we use the same simulation as in the previous paragraph ($\rho_t \equiv 10$, $\sigma_t^2 \equiv 1$ and ν_t is shown in the top left plot of Fig. 6) but now we only assume the local range ρ_t is known and proceed to estimate σ_t^2 and ν_t. There is some justification for assuming ρ_t is fixed at a known constant since, in the two dimensional stationary Matérn covariance model, it is impossible to consistently distinguish ρ from σ^2 (see [22,2]). The results are shown in Fig. 8. The top two diagrams show the estimates of ν_t and σ_t^2, respectively. The bottom two diagrams show the errors $\hat{\nu}_t - \nu_t$ and $\hat{\sigma}_t^2 - \sigma_t^2$. One can clearly see an increase in estimation error for estimating ν_t when fitting a model that has an unnecessary spatially varying variance parameter. However, we note that the errors for estimating ν_t are compensated by complementary errors in the estimates $\hat{\sigma}_t^2$. Indeed, under-estimates of ν_t seem to correspond to under-estimates of σ_t^2, and vice versa. This makes some sense in that small values of ν_t model *greater* local spatial variability whereas small values of σ_t^2 model *smaller* local spatial variability (albeit in a different manner), hence the compensatory nature of the errors (this behavior is expected in light of the results in Section 6.7 of [18]).

Acknowledgment

The second author was supported by the US Department of Energy through grant DE-SC0002557.

Appendix A

In this appendix we show that the nonstationary covariance function (9) is positive definite. Originally presented in [19], the proof combines the results found in [13,14] for generating a spatially varying geometric anisotropy with those found in [16] for generating a local isotropic Matérn with spatially varying smoothness parameter. We start by stating the following lemma, which is proved using a convolution argument found in [14].

Lemma 1. *Let α_t map $t \in \mathbb{R}^d$ to the set of real $d \times d$ positive definite matrices (denoted $PD_d(\mathbb{R})$) and let $\phi(u) \triangleq \exp(-|u|^2/2)$ Then*

$$\det(\alpha_{st}^{-1/2})\phi\big[\alpha_{st}^{-1/2}(s-t)/\sigma\big] = c_t c_s \int_{\mathbb{R}^d} \phi\big[\alpha_s^{-1/2}(u-s)/\sigma\big]\phi\big[\alpha_t^{-1/2}(u-t)/\sigma\big]\mathrm{d}u \tag{12}$$

where $\alpha_{st} \triangleq (\alpha_s+\alpha_t)/2$ and $c_s \triangleq (2\pi)^{-d/4}\sigma^{-d/2}\det(\alpha_t^{-1/2})$.

The importance of this lemma is that the right hand side of (12) is positive definite with locally varying geometric anisotropy $\alpha_t^{-1/2}$. This follows since the right hand side of Eq. (12) equals $\mathrm{cov}(Z_\sigma(s), Z_\sigma(t))$ where $Z_\sigma(t) = c_t \int_{\mathbb{R}^d} \phi\big[\alpha_t^{-1/2}(u-t)/\sigma\big]\mathrm{d}W(u)$ and $\mathrm{d}W$ is Gaussian white noise. Now let $K_\sigma(s,t) \triangleq \det(\alpha_{st}^{-1/2})\phi\big[\alpha_{st}^{-1/2}(s-t)/\sigma\big]$ and notice that for any function $g_\sigma(\cdot)$ the function $g_\sigma(s)g_\sigma(t)K_\sigma(s,t)$ is positive definite. Since convex combinations and limits of positive definite functions are again positive definite we have that

$$\int_0^\infty g_\sigma(s)g_\sigma(t)K_\sigma(s,t)\mathrm{d}\mu(\sigma) \tag{13}$$

is positive definite for any positive finite measure μ such that $\int_{\mathbb{R}^d} g_\sigma^2(t)K_\sigma(t,t)\mathrm{d}\mu(\sigma) < \infty$ for all t.

Claim 1. *Let $\sigma_t\colon \mathbb{R}^d \to \mathbb{R}^+$, $\nu_t\colon \mathbb{R}^d \to \mathbb{R}^+$ and $\alpha_t\colon \mathbb{R}^d \to PD_d(\mathbb{R})$ and define $\alpha_{st} \triangleq (\alpha_t+\alpha_s)/2$, $\nu_{st} \triangleq (\nu_t+\nu_s)/2$. Then*

$$R(s,t) = \sigma_t\sigma_s \det\big(\alpha_{st}^{-1/2}\big)\mathcal{M}_{\nu_{st}}\left(\left|\alpha_{st}^{-1/2}(s-t)\right|\right)$$

is positive definite on $\mathbb{R}^d$ where $\mathcal{M}_\nu(x) = x^\nu\mathcal{K}_\nu(x)$ and $\mathcal{K}_\nu$ is the modified Bessel function of the second kind of order $\nu > 0$ (see [1]).

Proof. Let $g_\sigma(t) = \left(\frac{\sigma^2}{2}\right)^{\nu_t/2-1}$ and μ have density $\frac{\sigma^2}{2}\mathrm{e}^{-\sigma^2/2}$ with respect to Lebesque measure. Then (13) becomes

$$\begin{aligned}
\int_0^\infty g_\sigma(s)g_\sigma(t)K_\sigma(s,t)\mathrm{d}\mu(\sigma) &= \det(\alpha_{st}^{-1/2})\int_0^\infty \left(\frac{\sigma^2}{2}\right)^{\nu_{st}-1}\exp\left[-\frac{|\alpha_{st}^{-1/2}(s-t)|^2}{4(\sigma^2/2)} - \frac{\sigma^2}{2}\right]\mathrm{d}\sigma \\
&\overset{x=\sigma^2/2}{=} \det(\alpha_{st}^{-1/2})\int_0^\infty x^{\nu_{st}-1}\exp\left[-\frac{|\alpha_{st}^{-1/2}(s-t)|^2}{4x} - x\right]\mathrm{d}x \\
&= 2\det(\alpha_{st}^{-1/2})2^{-\nu_s/2}2^{-\nu_t/2}|\alpha_{st}^{-1/2}(s-t)|^{\nu_{st}}\mathcal{K}_{\nu_{st}}(|\alpha_{st}^{-1/2}(s-t)|)
\end{aligned}$$

where the last line is from (3.472.9) of Gradshteyn and Ryzhik [9]. Therefore $\det(\alpha_{st}^{-1/2})\mathcal{M}_{\nu_{st}}\big(|\alpha_{st}^{-1/2}(s-t)|\big)$ is positive definite. □

Appendix B

In this appendix we present evidence that the first order Taylor expansion of $E_\theta S_k(\theta(t_0))$ as the spatial points $t_1,\ldots,t_n$ shrink toward t_0 gives $c\cdot(t_0-t_k)$ for the variance modulation example presented in Section 3. In this example one observes $\sigma(t)W(t)$ at spatial locations $t_1,\ldots,t_n$ (let the observations be denoted $z_1,\ldots,z_n$), where the covariance structure of W is completely known. The goal is to then estimate $\sigma(t_0)$ at some fixed spatial location $t_0 \in \mathbb{R}^d$. To make the following equations more readable we use σ_0 to denote the value $\sigma(t_0)$. We additionally suppose the spatial locations $t_1,\ldots,t_n$ are ordered by their distance to t_0 and let $\boldsymbol{z}_k \triangleq (z_1,\ldots,z_k)$ be the set of k nearest observations to t_0. In what follows we give an informal derivation that

$$E_\sigma S_k(\sigma_0) = -\frac{2[\nabla\sigma(t_0)]\cdot(t_k-t_0)}{\sigma_0^2} + o(|t_k-t_0|) \tag{14}$$

where $S_k(\sigma_0) = \frac{\partial}{\partial\sigma_0}\log f_{\sigma_0}(z_k|\mathcal{N}_{k-1,t_0})$. In the above equation E_σ denotes expectation with respect to the fully nonstationary model and E_{σ_0} denotes expectation with respect to the local stationary model determined by $\sigma(t_0) = \sigma_0$.

Notice that the density, $f_\sigma(\boldsymbol{z}_k)$, of the observations in $\boldsymbol{z}_k$ can be regarded as depending on k parameter values $\sigma(t_1), \ldots, \sigma(t_k)$. Let $f_{\sigma_0}(\boldsymbol{z}_k)$ denote the density of $\boldsymbol{z}_k$ obtained by replacing all the values of $\sigma(t_1), \ldots, \sigma(t_k)$ with σ_0 (this is the stationary model given by σ_0). Proceeding formally we then get

$$\begin{aligned} f_\sigma(\boldsymbol{z}_k) &= f_{\sigma_0}(\boldsymbol{z}_k) + \sum_{j=1}^{k} F^{(j)}_{\sigma_0}(\boldsymbol{z}_k)(\sigma(t_j) - \sigma(t_0)) + o(|\sigma(t_j) - \sigma(t_0)|) \\ &= f_{\sigma_0}(\boldsymbol{z}_k) + \nabla\sigma(t_0) \cdot \sum_{j=1}^{k} F^{(j)}_{\sigma_0}(\boldsymbol{z}_k)(t_j - t_0) + o(|t_k - t_0|) \end{aligned} \tag{15}$$

where $F^{(j)}_{\sigma_0}(\boldsymbol{z}_k)$ is defined to be $\left[\frac{\partial}{\partial\sigma(t_j)} f_\sigma(\boldsymbol{z}_k)\right]_{\sigma=\sigma_0}$, which denotes the result of replacing each occurrence of $\sigma(t_1), \ldots, \sigma(t_k)$ in $\frac{\partial}{\partial\sigma(t_j)} f_\sigma(\boldsymbol{z}_k)$ by the constant σ_0. Notice that the error in the last equality is $o(|t - t_k|)$ since the points are ordered so that t_k is the kth furthest point from t_0. Now by letting $\Delta_\sigma \triangleq \operatorname{diag}(\sigma(t_1), \ldots, \sigma(t_n))$

$$\begin{aligned} \frac{F^{(j)}_{\sigma_0}(\boldsymbol{z}_k)}{f_{\sigma_0}(\boldsymbol{z}_k)} &= \left[\frac{\partial}{\partial\sigma(t_j)} \log f_\sigma(\boldsymbol{z}_k)\right]_{\sigma=\sigma_0} \\ &= \left[\frac{\partial}{\partial\sigma(t_j)} \left(-\frac{\boldsymbol{z}_k^T \Delta_\sigma^{-1} \Sigma_k^{-1} \Delta_\sigma^{-1} \boldsymbol{z}_k}{2} - \frac{\log\det \Delta_\sigma^2}{2}\right)\right]_{\sigma=\sigma_0} \\ &= \left[\sigma^{-2}(t_j) \frac{\boldsymbol{z}_k^T \xi_j \Sigma_k^{-1} \Delta_\sigma^{-1} \boldsymbol{z}_k}{2} + \sigma^{-2}(t_j) \frac{\boldsymbol{z}_k^T \Delta_\sigma^{-1} \Sigma_k^{-1} \xi_j \boldsymbol{z}_k}{2} - \frac{1}{2\sigma(t_j)}\right]_{\sigma=\sigma_0} \\ &= \frac{\boldsymbol{z}_k^T \xi_j \Sigma_k^{-1} \boldsymbol{z}_k}{2\sigma_0^3} + \frac{\boldsymbol{z}_k^T \Sigma_k^{-1} \xi_j \boldsymbol{z}_k}{2\sigma_0^3} - \frac{1}{2\sigma_0} \end{aligned}$$

where ξ_j is the $k \times k$ matrix with 1 in the (j, j) entry and zeros everywhere else. Notice that $\boldsymbol{z}_k^T \xi_j \Sigma_k^{-1} \boldsymbol{z}_k = \boldsymbol{z}_k^T \Sigma_k^{-1} \xi_j \boldsymbol{z}_k$ since Σ_k^{-1} is symmetric. Therefore

$$\frac{F^{(j)}_{\sigma_0}(\boldsymbol{z}_k)}{f_{\sigma_0}(\boldsymbol{z}_k)} = \frac{\boldsymbol{z}_k^T \xi_j \Sigma_k^{-1} \boldsymbol{z}_k}{\sigma_0^3} - \frac{1}{2\sigma_0}.$$

Now from (15) we get that

$$\begin{aligned} E_\sigma S_k(\sigma_0) &= \int S_k(\sigma_0) f_\sigma(\boldsymbol{z}_k) \mathrm{d}\boldsymbol{z}_k \\ &= E_{\sigma_0} S_k(\sigma_0) + \nabla\sigma(t_0) \cdot \sum_{j=1}^{k} (t_j - t_0) \int S_k(\sigma_0) F^{(j)}_{\sigma_0}(\boldsymbol{z}_k) \mathrm{d}\boldsymbol{z}_k + o(|t_0 - t_k|) \\ &= E_{\sigma_0} S_k(\sigma_0) + \nabla\sigma(t_0) \cdot \sum_{j=1}^{k} (t_j - t_0) E_{\sigma_0}\left[S_k(\sigma_0) \frac{F^{(j)}_{\sigma_0}(\boldsymbol{z}_k)}{f_{\sigma_0}(\boldsymbol{z}_k)}\right] + o(|t_0 - t_k|). \end{aligned}$$

Finally notice that

$$\begin{aligned} E_{\sigma_0}\left[S_k(\sigma_0) \frac{F^{(j)}_{\sigma_0}(\boldsymbol{z}_k)}{f_{\sigma_0}(\boldsymbol{z}_k)}\right] &= E_{\sigma_0}\left[S_k(\sigma_0)\left\{\frac{\boldsymbol{z}_k^T \xi_j \Sigma_k^{-1} \boldsymbol{z}_k}{\sigma_0^3} - \frac{1}{2\sigma_0}\right\}\right] \\ &= \operatorname{cov}\left(\frac{\boldsymbol{z}_{k-1}^T \Sigma_{k-1}^{-1} \boldsymbol{z}_{k-1} - \boldsymbol{z}_k^T \Sigma_k^{-1} \boldsymbol{z}_k}{\sigma_0^3}, \frac{\boldsymbol{z}_k^T \xi_j \Sigma_k^{-1} \boldsymbol{z}_k}{\sigma_0^3}\right) \\ &= \frac{2\operatorname{tr}(\Sigma_{k-1\to k}^{-1} \Sigma_k \xi_j)}{\sigma_0^2} - \frac{2\operatorname{tr}(\xi_j)}{\sigma_0^2}. \end{aligned}$$

The term $\operatorname{tr}(\Sigma_{k-1\to k}^{-1} \Sigma_k \xi_j)$ is 1 when $j < k$ and is 0 when $j = k$. Therefore

$$E_{\sigma_0}\left[S_k(\sigma_0) \frac{F^{(j)}_{\sigma_0}(\boldsymbol{z}_k)}{f_{\sigma_0}(\boldsymbol{z}_k)}\right] = \begin{cases} -2\sigma_0^{-2}, & \text{when } j = k \\ 0, & \text{otherwise.} \end{cases}$$

This establishes Eq. (14), which was to be shown.

Appendix C

In this appendix we derive Eq. (4), which decomposes $E\left\{\hat{\sigma}^2(t_0) - \sigma^2(t_0)\right\}^2$ into bias and variance terms under the assumption that $\sigma(t) = c_0 + \sum_{p=1}^{N} c_p (t - t_0)^p$, where the expectation is taken over resampling of the data and a prior π for the vector of nuisance parameters $\boldsymbol{c} \triangleq (c_1, \ldots, c_N)$. Before we begin we recall some notation and assumptions from Section 3.1. The observation locations are denoted $(t_1, \ldots, t_n)$ with corresponding random field observations $\boldsymbol{z}_n = (z_1, \ldots, z_n)^T$ so that $z_j = \sigma(t_j) W(t_j)$. The observations are ordered by their distance to t_0, so that the observation location t_1 is closest to t_0, for example. The matrix Σ_k denotes the covariance matrix of $(W(t_1), \ldots, W(t_n))$. This implies that the k-vector $\boldsymbol{z}_k = (z_1, \ldots, z_k)^T$ has covariance matrix $\Delta_\sigma \Sigma_k \Delta_\sigma$, where

$$\Delta_\sigma \triangleq \begin{pmatrix} \sigma(t_1) & & 0 \\ & \ddots & \\ 0 & & \sigma(t_k) \end{pmatrix} = \sum_{p=0}^{N} c_p \begin{pmatrix} (t_1 - t_0)^p & & 0 \\ & \ddots & \\ 0 & & (t_k - t_0)^p \end{pmatrix} \triangleq \sum_{p=0}^{N} c_p \Delta_k^p \tag{16}$$

and $\Delta_k \triangleq \operatorname{diag}[(t_1 - t_0), \ldots, (t_k - t_0)]$. We make the assumption that $\pi(\boldsymbol{c}) = \pi_1(c_1) \cdots \pi_N(c_N)$, $E_{\pi_j} c_j = 0$ and $E_{\pi_j} c_j^3 = 0$ for all $1 \le j \le N$. In the remainder of the Appendix E_π will denote expectation with respect to the prior on the nuisance parameters.

The expected risk decomposes as follows

$$\begin{aligned} E\left\{\hat{\sigma}^2(t_0) - \sigma^2(t_0)\right\}^2 &= E_\pi E\left\{\left[\hat{\sigma}^2(t_0) - \sigma^2(t_0)\right]^2 \middle| \boldsymbol{c}\right\} \\ &= \underbrace{E_\pi \left\{E[\hat{\sigma}^2(t_0)|\boldsymbol{c}] - \sigma^2(t_0)\right\}^2}_{\text{bias}^2 \text{ terms}} + \underbrace{E_\pi \operatorname{var}\left[\hat{\sigma}^2_\lambda(t_0)|\boldsymbol{c}\right]}_{\text{variance}}. \end{aligned} \tag{17}$$

To derive expressions for the above two terms we show the following lemma.

Lemma 2.

$$E\left[\hat{\sigma}^2(t_0)|\boldsymbol{c}\right] = \sum_{p_1, p_2 = 0}^{N} c_{p_1} c_{p_2} B^{p_1, p_2} \tag{18}$$

$$\operatorname{var}\left[\hat{\sigma}^2(t_0)|\boldsymbol{c}\right] = \sum_{p_1, \ldots, p_4 = 0}^{N} c_{p_1} \cdots c_{p_4} B^{p_1, \ldots, p_4} \tag{19}$$

where

$$B^{p_1, p_2} \triangleq \sum_{k=1}^{n} \tilde{w}_k \operatorname{tr}\left[\Sigma_k^{-1} \Delta_k^{p_1} \Sigma_k \Delta_k^{p_2}\right]$$

$$B^{p_1, \ldots, p_4} \triangleq 2 \sum_{j,k=1}^{n} \tilde{w}_k \tilde{w}_j \operatorname{tr}\left[\Sigma_k^{-1} \Delta_k^{p_1} \Sigma_k \Delta_k^{p_2} \Sigma_{j \to k}^{-1} \Delta_k^{p_3} \Sigma_k \Delta_k^{p_4}\right]$$

$$\tilde{w}_k \triangleq \begin{cases} (w_k - w_{k+1}) / \sum_{j=1}^{n} w_j, & \text{if } k < n; \\ w_n / \sum_{j=1}^{n} w_j, & \text{if } k = n \end{cases}$$

and $\Sigma_{j \to k}^{-1} \triangleq \begin{pmatrix} \Sigma_j^{-1} & 0 \\ 0 & 0 \end{pmatrix}$ *is the matrix* Σ_j^{-1} *padded with zeros so that it has the same size as* Σ_k^{-1}.

Proof. First notice that the expected value of $\boldsymbol{z}_k^T \Sigma_k^{-1} \boldsymbol{z}_k$, conditional on $\boldsymbol{c} = (c_1, \ldots, c_N)$, is

$$E\left[\boldsymbol{z}_k^T \Sigma_k^{-1} \boldsymbol{z}_k | \boldsymbol{c}\right] = E \operatorname{tr}\left[\boldsymbol{z}_k \boldsymbol{z}_k^T \Sigma_k^{-1} | \boldsymbol{c}\right] = \operatorname{tr}[\Delta_\sigma \Sigma_k \Delta_\sigma \Sigma_k^{-1}] = \sum_{p_1, p_2 = 0}^{N} c_{p_1} c_{p_2} \operatorname{tr}\left[\Sigma_k^{-1} \Delta_k^{p_1} \Sigma_k \Delta_k^{p_2}\right].$$

When $j \le k$ we have that

$$\begin{aligned} \operatorname{cov}\left[\boldsymbol{z}_k^T \Sigma_k^{-1} \boldsymbol{z}_k, \boldsymbol{z}_j^T \Sigma_j^{-1} \boldsymbol{z}_j | \boldsymbol{c}\right] &= \operatorname{cov}\left[\boldsymbol{z}_k^T \Sigma_k^{-1} \boldsymbol{z}_k, \boldsymbol{z}_k^T \Sigma_{j \to k}^{-1} \boldsymbol{z}_k | \boldsymbol{c}\right] = 2 \operatorname{tr}[\Sigma_k^{-1} \Delta_\sigma \Sigma_k \Delta_\sigma \Sigma_{j \to k}^{-1} \Delta_\sigma \Sigma_k \Delta_\sigma] \\ &= 2 \sum_{p_1, \ldots, p_4 = 0}^{N} c_{p_1} \cdots c_{p_4} \operatorname{tr}[\Sigma_k^{-1} \Delta_k^{p_1} \Sigma_k \Delta_k^{p_2} \Sigma_{j \to k}^{-1} \Delta_k^{p_3} \Sigma_k \Delta_k^{p_4}]. \end{aligned}$$

At this point it becomes convenient to rewrite $\hat{\sigma}^2(t_0)$ as $\sum_{k=1}^n \tilde{w}_k \, \boldsymbol{z}_k^T \Sigma_k^{-1} \boldsymbol{z}_k$. The conditional expected value of $\hat{\sigma}^2_\lambda(t_0)$ can now be written

$$E\big[\hat{\sigma}^2(t_0)|\boldsymbol{c}\big] = \sum_{k=1}^n \tilde{w}_k E\big[\boldsymbol{z}_k^T \Sigma_k^{-1} \boldsymbol{z}_k|\boldsymbol{c}\big] = \sum_{k=1}^n \sum_{p_1,p_2=0}^N \tilde{w}_k c_{p_1} c_{p_2} \mathrm{tr}\big[\Sigma_k^{-1} \Delta_k^{p_1} \Sigma_k \Delta_k^{p_2}\big] = \sum_{p_1,p_2=0}^N c_{p_1} c_{p_2} B^{p_1,p_2}$$

and the conditional variance is

$$\begin{aligned}
\mathrm{var}\big[\hat{\sigma}^2(t_0)|\boldsymbol{c}\big] &= \mathrm{var}\left[\sum_{k=1}^n \tilde{w}_k \, \boldsymbol{z}_k^T \Sigma_k^{-1} \boldsymbol{z}_k|\boldsymbol{c}\right] = \sum_{j,k=1}^n \tilde{w}_k \tilde{w}_j \, \mathrm{cov}[\boldsymbol{z}_k^T \Sigma_k^{-1} \boldsymbol{z}_k, \boldsymbol{z}_j^T \Sigma_j^{-1} \boldsymbol{z}_j] \\
&= 2 \sum_{j,k=1}^n \sum_{p_1,\ldots,p_4=0}^N \tilde{w}_k \tilde{w}_j c_{p_1} \cdots c_{p_4} \mathrm{tr}[\Sigma_k^{-1} \Delta_k^{p_1} \Sigma_k \Delta_k^{p_2} \Sigma_{j\wedge k \to j \vee k}^{-1} \Delta_k^{p_3} \Sigma_k \Delta_k^{p_4}] \\
&= \sum_{p_1,\ldots,p_4=0}^N c_{p_1} \cdots c_{p_4} B^{p_1,\ldots,p_4}
\end{aligned}$$

where $B^{p_1,\ldots,p_4} \triangleq 2 \sum_{j,k=1}^n \tilde{w}_k \tilde{w}_j \mathrm{tr}\Big[\Sigma_k^{-1} \Delta_k^{p_1} \Sigma_k \Delta_k^{p_2} \Sigma_{j\wedge k \to j \vee k}^{-1} \Delta_k^{p_3} \Sigma_k \Delta_k^{p_4}\Big]$. This completes the proof of Lemma 2. □

Now to prove the decomposition (4) we use Lemma 2 to show the expected squared bias can be decomposed into two quadratic terms

$$\begin{aligned}
E_\pi \left\{ E\big[\hat{\sigma}^2(t_0)|\boldsymbol{c}\big] - \sigma^2(t_0) \right\}^2 &= E_\pi \left\{ \sum_{p_1,p_2=0}^N c_{p_1} c_{p_2} B^{p_1,p_2} - c_0^2 \right\}^2, \quad \text{by Lemma 2} \\
&= E_\pi \left\{ \sum_{p_1,p_2=1}^N c_{p_1} c_{p_2} B^{p_1,p_2} + 2 c_0 \sum_{p=1}^N c_p B^{0,p} \right\}^2 \\
&= E_\pi \left\{ \sum_{p_1,p_2=1}^N c_{p_1} c_{p_2} B^{p_1,p_2} \right\}^2 + 4 \, c_0^2 \, E_\pi \left\{ \sum_{p=1}^N c_p B^{0,p} \right\}^2 \qquad (20)
\end{aligned}$$

since $B^{0,0} = \sum_{k=1}^n k \tilde{w}_k = 1$ and $B^{0,p} = B^{p,0}$. Note that the cross term in (20) is zero since each c_j is independent and mean zero under the prior π and by the assumption $E_{\pi_j} c_j^3 = 0$ for all $j \geq 1$. Therefore the expected risk is

$$\begin{aligned}
E_\pi E\left\{ \left[\hat{\sigma}^2(t_0) - \sigma^2(t_0)\right]^2 |\boldsymbol{c} \right\} &= \sum_{p_1,p_2,p_3,p_4=0}^N E_\pi \big[c_{p_1} c_{p_2} c_{p_3} c_{p_4}\big] B^{p_1,p_2,p_3,p_4} \\
&\quad + E_\pi \left\{ \sum_{p_1,p_2=1}^N c_{p_1} c_{p_2} B^{p_1,p_2} \right\}^2 + 4 \, c_0^2 \, E_\pi \left\{ \sum_{p=1}^N c_p B^{0,p} \right\}^2.
\end{aligned}$$

Finally, we mention the relationship between higher order weights and the bias terms in the expected risk. Notice that

$$\begin{aligned}
B^{0,p} &= \sum_{k=1}^n \tilde{w}_k \mathrm{tr}\big[\Delta_k^p\big] = \sum_{k=1}^n \tilde{w}_k \mathrm{tr}\big[\mathrm{diag}((t_1 - t_0)^p, \ldots, (t_k - t_0)^p)\big] \\
&= \sum_{k=1}^n \tilde{w}_k [(t_1 - t_0)^p + \cdots + (t_k - t_0)^p] \\
&= \frac{1}{\sum_{j=1}^n w_j} \sum_{k=1}^n w_k (t_k - t_0)^p,
\end{aligned}$$

which implies, as is mentioned in Section 3.1, that by using higher order weights one can enforce $B^{0,p} \to 0$ as $n \to \infty$ and the bandwidth of the weights $\lambda \to 0$ and the t_k's get more dense in a bounded region near t_0.

References

[1] M. Abramowitz, I. Stegun, Handbook of Mathematical Functions, ninth ed., Dover, New York, 1965.
[2] E. Anderes, On the consistent separation of scale and variance for Gaussian random fields, Ann. Statist. 38 (2010) 870–893.
[3] E. Anderes, M. Stein, Estimating deformations of isotropic Gaussian random fields on the plane, Ann. Statist. 36 (2008) 719–741.

[4] A. Ayache, S. Cohen, J.L. Vehel, The covariance structure of multifractional Brownian motion, with application to long range dependence, in: ICASSP'00: Proceedings of the Acoustics, Speech, and Signal Processing, 2000, on IEEE International Conference, IEEE Computer Society, Washington, DC, USA, 2000, pp. 3810–3813.
[5] R. Dahlhaus, A likelihood approximation for locally stationary processes, Ann. Statist. 28 (2000) 1762–1794.
[6] D. Damian, P. Sampson, P. Guttorp, Bayesian estimation of semi-parametric non-stationary spatial covariance structures, Environmetrics 12 (2001) 161–178.
[7] J. Fan, M. Farmen, I. Gijbels, Local maximum likelihood estiamtion and inference, J. R. Stat. Soc. Ser. B (1998) 591–608.
[8] M. Fuentes, A high frequency kriging approach for non-stationary environmental processes, Environmetrics 12 (2001) 469–483.
[9] I.S. Gradshteyn, I.M. Ryzhik, Table of Integrals, Series, and Products, Academic Press, San Diego, 2000.
[10] D. Higdon, J. Swall, J. Kern, Non-stationary spatial modeling, in: Bayesian Statistics, vol. 6, Oxford University Press, 1999, pp. 761–768.
[11] H.G. Müller, Weighted local regression and kernel methods for nonparametric curve fitting, J. Amer. Statist. Assoc. 82 (1987) 231–238.
[12] D. Nychka, C. Wikle, A. Royle, Multiresolution models for nonstationary spatial covariance functions, Stat. Model. 2 (2002) 315–331.
[13] C. Paciorek, M. Schervish, Nonstationary covariance functions for Gaussian process regression, Adv. Neural Inf. Process. Syst. 16 (2004) 273–280.
[14] C. Paciorek, M. Schervish, Spatial modelling using a new class of nonstationary covariance functions, Environmetrics 17 (2007) 483–506.
[15] R. Peltier, J.L. Véhel, Multifractional Brownian motion: definition and preliminary results, Tech. Rep., Res. Rept. 2645, INRIA, 2005.
[16] A. Pintore, C. Holmes, Spatially adaptive non-stationary covariance functions via spatially adaptive spectra, 2004. Available at: www.stats.ox.ac.uk/~cholmes/reports/spectral_tempering.pdf.
[17] A. Sly, Integrated fractional white noise as an alternative to multifractional Brownian motion, J. Appl. Probab. 44 (2007) 393–408.
[18] M.L. Stein, Interpolation of Spatial Data: Some Theory for Kriging, Springer, New York, 1999.
[19] M.L. Stein, Nonstationary spatial covariance fucntions, Tech. Rep., University of Chicago, Department of Statistics, 2005. Available at: galton.uchicago.edu/cises/research/cises-tr21.pdf.
[20] S. Stoev, M. Taqqu, How rich is the class of multifractional Brownian motions? Stochastic Process. Appl. 116 (2006) 200–221.
[21] M.P. Wand, W.R. Schucany, Gaussian-based kernels, Canad. J. Statist. 18 (3) (1990) 197–204.
[22] H. Zhang, Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics, J. Amer. Statist. Assoc. 99 (2004) 250–261.