Probab. Theory Relat. Fields 99, 399–424 (1994)

Probability Theory and Related Fields

Properties of nonparametric estimators of autocovariance for stationary random fields

Peter Hall *, **Prakash Patil**

Centre for Mathematics and its Applications, Australian National University, Canberra ACT 2601, Australia

Received: 2 August 1993 / In revised form: 23 December 1993

Summary. We introduce nonparametric estimators of the autocovariance of a stationary random field. One of our estimators has the property that it is itself an autocovariance. This feature enables the estimator to be used as the basis of simulation studies such as those which are necessary when constructing bootstrap confidence intervals for unknown parameters. Unlike estimators proposed recently by other authors, our own do not require assumptions such as isotropy or monotonicity. Indeed, like nonparametric function estimators considered more widely in the context of curve estimation, our approach demands only smoothness and tail conditions on the underlying curve or surface (here, the autocovariance), and moment and mixing conditions on the random field. We show that by imposing the condition that the estimator be a covariance function we actually reduce the numerical value of integrated squared error.

Mathematics Subject Classification (1991): 62G05, 62G20

1 Introduction

The covariance function of a stationary random field in the plane may be regarded as a surface whose height above the plane at the point $t=(t^{(1)}, t^{(2)})$ equals the covariance between values of the process at points lagged by an amount t. This paper was motivated by the problem of constructing a "confidence sandwich" for the covariance function of a stationary random field. That is to say, we wished to construct two surfaces, one above the other, between which the true covariance function (over a specified region of the plane, for example a disc centred at the origin) lay with a

* CSIRO Division of Mathematics and Statistics, Lindfield, NSW, Australia

certain preassigned probability, such as 0.95. The confidence sandwich had to be constructed for the most part nonparametrically, at least in terms of the marginal distribution of the process and the strengths of long- and short-range association. The work described in the present paper develops theory behind one part of the overall algorithm for constructing a nonparametric confidence sandwich for a covariance function.

The full algorithm involves three steps, which we now describe. (i) Model the random process by a function, g say, of a Gaussian random field. The requisite function is estimated by the obvious empirical transformation – begin by estimating the first-order marginal distribution of the process, then apply this function to the entire process, and following that, apply the inverse of the standard normal distribution function. Only if the original process is really a function of a Gaussian process will such an approach correctly reproduce the marginal distributions in all dimensions. Nevertheless, from the point of view of data analysis this sequence of operations would capture a satisfactory proportion of the properties of both long- and short-range association, as well as the marginal distribution. (ii) Construct a nonparametric estimator of the covariance function of the derived Gaussian process. Critically, this estimate must itself be a covariance function, and that problem motivates this paper. (iii) Simulate from the Gaussian process estimated in step (ii), and take the simulated version of the original process to be g^{-1} of the simulated Gaussian process. (iv) Based on these simulations, use the percentile bootstrap method to construct the desired confidence sandwich.

We tackle the problem in a little more generality than the two-dimensional case discussed above. In particular, we assume that the random field is indexed in $\mathbb{R}^d$ for $d \geqq 1$, and we do not demand that the process be Gaussian. Indeed, we ask only that certain moment and mixing conditions be satisfied, and that the tails of the covariance function decay sufficiently fast.

In our analysis of the performance of covariance estimators we have employed the L^2 metric throughout, not least because of its technical tractability. We have not weighted this error, and could instead have considered mean squared error divided by the square of the target covariance function, or a similar relative measure of performance. However, while this approach is perhaps attractive towards the origin, where covariance can be estimated with ease, it is singularly unappealing in the tails, where it diverges without bound. Thus, a decision to use weighted L^2 error involves an awkward, *ad hoc* decision about truncation in the tails, which can drive the conclusion of the entire analysis. Alternative approaches, based for example on the L^1 metric, are not any more appealing for covariance estimation than L^2, unless one particularly wishes to downweight larger departures from the truth. (They do, however, have greater attraction in the context of density estimation, since a density is in the class of L^1 functions; see for example Devroye and Györfi (1985).)

We develop detailed theory describing the rate of convergence of integrated squared error (ISE) of our covariance function estimator. In particular, we prove that ISE is asymptotic to a small constant multiplied by the integral of the square of a certain fixed Gaussian process. The value of the constant converges to zero as sample size increases, and the convergence rate of the constant determines the overall convergence rate of the estimator.

We should stress that we make very few assumptions about the relative positions of the "design points" in $\mathbb{R}^d$, at which the process is observed. They should be sufficiently closely spaced to permit adequate smoothing, but this type of condition is necessary for nonparametric covariance estimation. In particular we do not ask that the design points be located on a grid, or lattice, in $\mathbb{R}^d$. Such assumptions are very common in related work on the case $d=1$; in fact, they are the rule in that context, and our non-lattice approach is a rare exception. However, lattice-valued designs are only very rarely encountered in the case of $d \geqq 2$ dimensions, and so it is important that we take the route which we have.

Our estimators are based on kernel methods, and in this sense they have similarities to those encountered in the context of nonparametric density and regression estimation; see for example Silverman (1986) and Härdle (1990). However, a more appropriate analogy is with distribution function estimation by integration of a kernel density estimator. In particular, the effect of the kernel and bandwidth vanishes entirely from first-order asymptotic theory, and the convergence rate is the analogue of root-n consistency in more classical statistical problems. The effects of kernel and bandwidth choice emerge only in second-order theory. All these are properties of kernel estimation of distribution functions, as is readily deduced after a little analysis.

Section 2 introduces our estimators and describes their basic properties. In particular, we show there that by insisting that the covariance estimator have the positive definiteness of a real covariance function we do reduce ISE. Section 3 explores properties in greater mathematical detail. There we define the limiting Gaussian process involved in our asymptotic formulae for ISE, and state the convergence results under explicit regularity conditions. All proofs are relegated to Sect. 4.

The practical problem of estimating a multivariate covariance function has been motivated by many authors; see for example Matheron (1971), Journel and Huijbregts (1978), Armstrong and Diamond (1984), Christakos (1984), Cressie (1991, Chap. 2), Shapiro and Botha (1991) and Sampson and Guttorp (1992). However, the estimators treated here are substantially different from those which have been considered earlier, in that the latter tend to rely either on parametric fitting (e.g. Cressie 1991) or repeated measurements (e.g. Sampson and Guttorp 1992), and do not always produce a covariance function in the continuum (e.g. Shapiro and Botha 1991). Even when nonparametric fitting has been contemplated by earlier authors it has often been under rather restrictive constraints, such as monotonicity or isotropy (e.g. Sampson and Guttorp 1992). The pointwise (as distinct from global, or ISE) properties of a one-dimensional version of our estimator have been considered before (Hall et al. 1992), but that work does not anticipate any of the mutivariate, global conclusions drawn in the present paper.

2 Estimator and its basic properties

Suppose the stationary random field $\{X(t), t \in \mathbb{R}^d\}$ is observed at points $t_1, \ldots, t_n$, which are not assumed to be on a grid. We wish to estimate the autocovariance, or covariance function,

$$\rho(t) = \operatorname{cov}\{X(0), X(t)\}.$$

A kernel estimator is defined by

$$\check{\rho}(t) = [\sum_i \sum_j K\{(t-t_{ij})/h\} X_{ij}] [\sum_i \sum_j K\{(t-t_{ij})/h\}]^{-1}, \tag{2.1}$$

where $t_{ij} = t_i - t_j$, $X_{ij} = \{X(t_i) - \bar{X}\}\{X(t_j) - \bar{X}\}$ and $\bar{X} = n^{-1} \sum X(t_i)$. Here, $h > 0$ is the bandwidth and K is a d-variate symmetric kernel function with $\int K = 1$.

It is not difficult to see that if, as n increases, the points t_i become increasingly dense in each bounded subset of $\mathbb{R}^d$, then the bandwidth h may be chosen so that $\check{\rho}(t) \to \rho(t)$ as $n \to \infty$, for each $t \in \mathbb{R}^d$. However, for a fixed sample $\{X(t_1), \ldots, X(t_n)\}$ the estimator $\check{\rho}(t)$ can become undefined as $|t|$ increases. (Here, $|t| = |(t^{(1)}, \ldots, t^{(d)})^T| = (\sum t^{(i)2})^{1/2}$ denotes the usual Euclidean distance.) For example, if K is compactly supported then for any $h > 0$ and all sufficiently large $|t|$, the ratio in (2.1) is of the indeterminate form 0/0. More problematically, the estimator $\check{\rho}(t)$ can be quite accurate for small to moderate values of $|t|$ but highly inaccurate, although well-defined, for large $|t|$'s. To illustrate this point, Fig. 2.1 depicts $\check{\rho}(t)$ and $\rho(t)$ when X is a Gaussian process with $d = 1$ and $\rho(t) = e^{-|t|^{1.98}}$, and when $n = 200$, $t_1, \ldots, t_n$ are uniformly distributed on (0, 40), $h = 0.4$, and $K(x) = (1 - |x|)\, I(|x| \leq 1)$ is the triangular kernel. It can be seen that $\check{\rho}(t)$ approximates $\rho(t)$ quite well for $0 < t \leq 4$, but quite poorly for $t > 4$.

We would usually only be interested in estimating ρ for processes that exhibit some degree of weak dependence, i.e. $\rho(t) \to 0$ as $|t| \to \infty$. The example illustrated in Fig. 2.1 indicates that in such cases, $\check{\rho}(t)$ may actually become large as $\rho(t)$ becomes small. Therefore, performance of $\check{\rho}$ may be improved by adjoining a suitable symmetric weight function $w = w_n$, which should enjoy the properties $\sup_{|t| \leq c} |w(t) - 1| \to 0$ as $n \to \infty$ for each $c > 0$, and

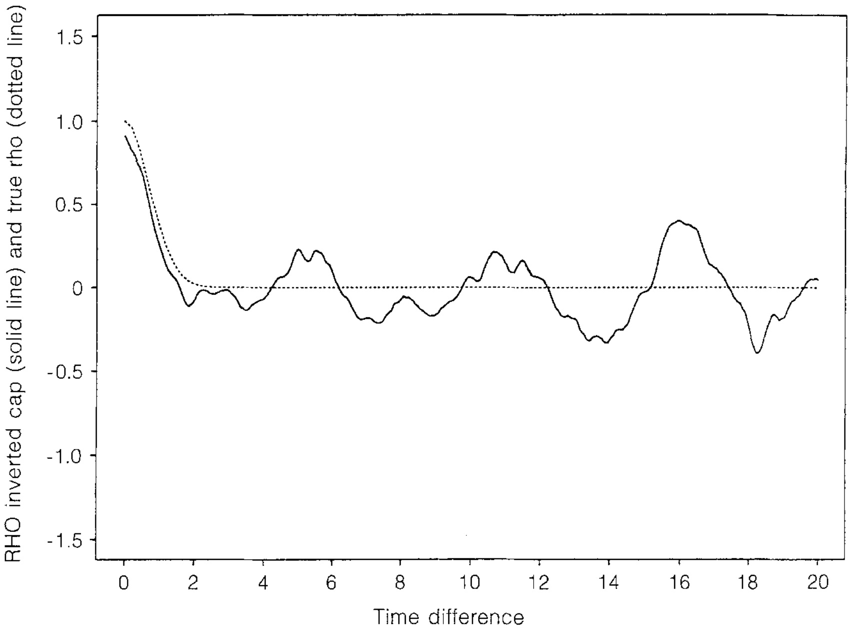

Fig. 2.1 Graphs of $\rho(t)=e^{-|t|^{1.98}}$, and $\check{\rho}(t)$, in the case where X is a Gaussian process with $d=1$, where $n=200$, $h=0.4$, K is the triangular kernel, and $t_1, \ldots, t_n$ are uniformly distributed on (0, 40)

$w(t)\to 0$ as $|t|\to\infty$ for each $n\geqq 1$. (For example, we might take $w(t)=I(|t|\leqq c_n)$ where $c_n\to\infty$ as $n\to\infty$.) The new estimator would be

$$\tilde{\rho}(t)=\check{\rho}(t)\,w(t).$$

Note that, in view of the symmetry of K and w, $\tilde{\rho}(-t)=\tilde{\rho}(t)$.

If w is chosen so that $\tilde{\rho}$ is square-integrable then the Fourier transform $\tilde{\psi}$ of $\tilde{\rho}$ is well-defined:

$$\tilde{\psi}(\theta)=\int e^{i\theta^T t}\,\tilde{\rho}(t)\,dt=\int\cos(\theta^T t)\,\tilde{\rho}(t)\,dt,\qquad \theta\in\mathbb{R}^d.$$

By Bochner's theorem, a necessary and sufficient condition for $\tilde{\rho}$ to be a covariance function is that $\tilde{\psi}\geqq 0$. The latter condition would never be satisfied in practice by the particular estimator that we have defined. This leads us to suggest an alternative estimator $\hat{\rho}$, which is defined as that function which is closest to $\tilde{\rho}$ in an L^2 sense subject to being a proper covariance function. This quantity is given by

$$\hat{\rho}(t)=(2\pi)^{-d}\int\cos(\theta^T t)\,\hat{\psi}(\theta)\,d\theta, \tag{2.2}$$

where $\hat{\psi}=\tilde{\psi}$ if $\tilde{\psi}>0$, and $\hat{\psi}=0$ otherwise.

Not only does $\hat{\rho}$ enjoy the covariance property, it is closer to ρ than is $\tilde{\rho}$. To appreciate this point, and the correctness of definition (2.2), note that if ρ_1, ρ_2 are two functions with respective Fourier transforms ψ_1, ψ_2, then by Parseval's identity,

$$\|\rho_1-\rho_2\|^2=\int(\rho_1-\rho_2)^2=(2\pi)^{-d}\int(\psi_1-\psi_2)^2=(2\pi)^{-d}\|\psi_1-\psi_2\|^2.$$

(The notation $\|\cdot\|$ denotes the L^2 norm for square-integrable functions on $\mathbb{R}^d$.) Taking $\rho_1=\tilde{\rho}$ (and $\psi_1=\tilde{\psi}$) we see that $\|\psi_1-\psi_2\|$ (and hence $\|\rho_1-\rho_2\|$) is minimised, subject to $\psi_2\geqq 0$, by choosing $\psi_2=\psi_1\, I(\psi_1>0)=\hat{\psi}$, i.e. $\rho_2=\hat{\rho}$. Furthermore, if ψ denotes the Fourier transform of ρ then, since $\psi\geqq 0$,

$$\begin{aligned}(2\pi)^d\,\|\hat{\rho}-\rho\|^2=\|\hat{\psi}-\psi\|^2=\|\tilde{\psi}\, I(\tilde{\psi}>0)-\psi\|^2\\ \leqq\|\tilde{\psi}-\psi\|^2=(2\pi)^d\,\|\tilde{\rho}-\rho\|^2,\end{aligned}$$

so that

$$\|\hat{\rho}-\rho\|\leqq\|\tilde{\rho}-\rho\|. \tag{2.3}$$

In practice this inequality will generally be strict, owing to the fact that $\tilde{\psi}<0$ in a place where $\psi\neq 0$.

If the points t_i are clustered sufficiently closely together then it is possible to estimate $\rho(t)$ with error $O_p(\lambda^{-d/2})$, where λ^d is proportional to the d-dimensional content of the "spread" of the set $\{t_i\}$. For example, the t_i's might be uniformly distributed inside a d-dimensional sphere of radius λ, with $n\gg\lambda^d$. In the coming sections we shall investigate circumstances where this is true, and show that the inequality in (2.3) is often very nearly an equality, in the sense that

$$\|\hat{\rho}-\rho\|=\{1+o_p(1)\}\,\|\tilde{\rho}-\rho\|$$

as n, $\lambda\to\infty$. We shall also develop limit theory for $\lambda^{d/2}\,\|\hat{\rho}-\rho\|$, proving that this quantity has a proper, nondegenerate limiting distribution.

3 Convergence of integrated squared errors of $\check{\rho}$ and $\tilde{\rho}$

3.1 Summary

Let $c>0$, and define

$$I_c=\int_{|t|\leqq c}\{\check{\rho}(t)-\rho(t)\}^2\,dt.$$

In subsection 3.2 we show that, for an appropriate sequence of constants a_n diverging to $+\infty$, $a_n I_c$ has a proper, nondegenerate limit distribution. This result is exceptional among related work on the integrated squared error of curve estimators, where it is typically the case that an appropriate scale multiple of integrated squared error converges to a constant, not to a nondegenerate weak limit; see for example Härdle and Marron (1986).

Note too that our result implies consistency of $\check{\rho}$ in a truncated L^2 metric; in fact, it yields a rate of convergence in that context.

Since I_c is not even well-defined for large c then the results discussed above cannot be extended to the case $c=\infty$. Nevertheless, if we modify $\check{\rho}$ to $\tilde{\rho}$ by adjoining a weight function, as discussed in Sect. 2, then it may be feasible to define I_∞ and to ask whether the weak convergence result continues to hold. In subsection 3.3 we show that it does, under appropriate regularity conditions.

Finally, in subsection 3.4, we show that if $\tilde{\rho}$ is replaced by the positive definite version $\hat{\rho}$ in the definition of I_∞ then, under appropriate conditions, an identical weak convergence result holds. In this sense, $\hat{\rho}$ and $\tilde{\rho}$ are first-order equivalent.

3.2 *Integrated squared error of $\check{\rho}$ over a finite region*

We assume the following conditions on K, on the process generating the design points t_i, on h and n, and on X. Let $r \geqq 1$ be an integer, denoting the "order" of the kernel K. Of K we ask that this function from $\mathbb{R}^d$ to $\mathbb{R}$ be bounded, compactly supported and Hölder continuous, and satisfy

$$\int (u^{(i)})^j K(u)\, du = \delta_{j0} \quad \text{for } 1 \leqq i \leqq d \text{ and } 0 \leqq j \leqq r-1$$

where $\delta_{j_1 j_2}$ is the Kronecker delta.

We assume that the design points t_i may be written as $t_i = \lambda u_i$, where $\lambda > 0$ is a large constant and $u_1, u_2, \ldots$ represents a realization of a sequence $U_1, U_2, \ldots$ of independent and identically distributed random d-vectors. We suppose that the common distribution of the U_i's has a density f which is compactly supported and has r bounded derivatives. We regard h and λ as functions of n, and ask that for some $\varepsilon > 0$,

$$h + \lambda^{-1} + h^{2r}\lambda^d + (n^{1-\varepsilon} h^{d/2} \lambda^{-d})^{-1} \to 0 \tag{3.1}$$

as $n \to \infty$.

The asymptotic regime described by these conditions corresponds to neither "infill" asymptotics, where an increasing number of points is added to a single fixed region, nor to "increasing domain" asymptotics, where points separated by a fixed distance are added to successively larger regions. In the case of $d=1$ dimension, the former type of design is common in nonparametric regression (e.g. with $t_i = t_{ni} = i/n$ for $1 \leqq i \leqq n$) and the latter type is common in time series analysis (e.g. with $t_i = i$ for $1 \leqq i \leqq n$). We claim that neither of these regimes can produce consistent estimation of the covariance function in the setting of the present paper. To appreciate why, observe that while the former produces an increasingly dense set of design points it does not permit arbitrary high lags to be observed, because no design points are available beyond a certain fixed limit. Therefore, consistent estimation of the full covariance function is only possible if that quantity vanishes outside a certain radius, which condition we regard as unrealistically stringent. On the other hand, the "increasing domain" type of asymptotics for design does not allow differences of design points to be arbitrary close to one another, and so does not permit estimation without structural assumptions. In particular, nonparametric estimation of the covariance function under smoothness conditions alone is not possible. The asymptotic regime suggested in this paper is a genuine compromise between these two extreme approaches, and captures the main features of both. It is related to the continuous observation of a random field over an increasingly large region, through the assumption that the design points are increasingly dense and are scattered across an increasingly large region.

The stationary random field X is assumed to be weakly dependent, in a sense that we now make precise. Given sets $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{R}^d$, define the distance function

$$\langle \mathcal{S}_1, \mathcal{S}_2 \rangle = \inf\{|t_1 - t_2| : t_i \in \mathcal{S}_i,\ i=1, 2\},$$

let $X[\mathcal{S}_i]$ denote the σ-field generated by the random variables $\{X(t),\ t \in \mathcal{S}_i\}$, and write $Y \in X[\mathcal{S}_i]$ to mean that the random variable Y is measurable in $X[\mathcal{S}_i]$. Put

$$\chi(s) = \sup_{\substack{\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{R}^d \text{s.t.} \\ \langle \mathcal{S}_1, \mathcal{S}_2 \rangle \geqq s}} \ \sup_{\substack{Y_i \in X[\mathcal{S}_i], i=1,2, \text{s.t.} \\ E(Y_i^2) < \infty}} |\mathrm{corr}(Y_1, Y_2)| .$$

We ask that

$$\chi(s) \to 0 \quad \text{as} \quad s \to +\infty; \tag{3.2}$$

the rate of convergence in this condition is not critical to our results.

We further assume that $E(X^8) < \infty$, and

$$\sup_{|t_0|, \ldots, |t_p| \leqq c} \int \ldots \int \left| E\left[\{X(0)\, X(t_0) - \rho(t_0)\} \cdot \prod_{j=1}^{p} \{X(v_j)\, X(v_j + t_j) - \rho(t_j)\} \right] \right| d v_1 \ldots d v_p < \infty \tag{3.3}$$

for $p = 1, 3$ and each $c > 0$. We also ask that ρ have $r \geqq 1$ bounded derivatives, and satisfy $\int |\rho| < \infty$.

In the case $d = 1$, and when X is a Gaussian process, condition (3.2) is equivalent to strong mixing. See for example Ibragimov and Rozanov (1978). In the case of processes related to Gaussian fields it is often straightforward to specify circumstances under which (3.2) and (3.3) hold. For example, both conditions are satisfied if $X = g(Z)$ where Z is a Gaussian field and g is a function, satisfying $E\{g(Z)^8\}<\infty$ and $\mathrm{cov}\{Z(0), Z(t)\}=0$ for all sufficiently large $|t|$.

Let Y denote a (nonstationary) Gaussian random field with zero mean and covariance function given by

$$\mathrm{cov}\{Y(s), Y(t)\}=(\int f^2)^{-2}(\int f^4)\int E[\{X(0)\,X(s)-\rho(s)\}\cdot\{X(u)\,X(u+t)-\rho(t)\}]\,du.$$

Theorem 3.1 *Under the conditions imposed above, and for each $c>0$, the random variable*

$$\lambda^d \int_{|t|\leqq c} \{\check{\rho}(t)-\rho(t)\}^2\,dt$$

converges in distribution to

$$\int_{|t|\leqq c} Y(t)^2\,dt$$

as $n\to\infty$.

Remark 3.1 The interpretation placed on the role of the u_i's in this result may be taken to be that the stated convergence occurs for sequences u_1, u_2, ... that arise with probability one as realizations of the independent random vectors U_1, U_2, ... However, our method of proof allows a more general interpretation in terms of triangular arrays, as follows. For each $n\geqq 1$, let $U_{n1}, \ldots, U_{nn}$ be independent and identically distributed random vectors with the distribution corresponding to density f, and suppose that for each n the sequence $u_i=u_{ni}$, $1\leqq i\leqq n$, arises as a realization of $U_{n1}, \ldots, U_{nn}$. The convergence stated in Theorem 3.1 is valid for a class of triangular arrays $\{u_{ni}, 1\leqq i\leqq n<\infty\}$ which are generated with probability one from triangular arrays $\{U_{ni}, 1\leqq i\leqq n\}$ defined as above, without regard to the manner in which the rows of the array might be related.

We say that an event or a convergence property holds "with U-probability one" if it is valid for a measure-one set of $u_i=u_{ni}$'s generated in this manner.

An alternative version of Theorem 3.1, for which the support of the design is a square lattice, may be proved without difficulty. There, in all appearances of f in the definition of the limit distribution, f should be replaced by the uniform density on its support. However, we have been unable to derive, in an economical way, a general version of Theorem 3.1 which in one stroke contains both the equally-spaced and stochastic design cases. The difficulties here are related to those in the better understood, and far simpler, context of design in univariate nonparametric regression. For example, even the particularly flexible design conditions allowed by Gasser and Müller (1979, 1984) and Cheng and Lin (1981) do not apply to stochastic design as well as deterministic design in the univariate regression setting, on account of the fact that adjacent points in stochastic design can be spaced $O(n^{-1} \log n)$ rather than $O(n^{-1})$ apart. (The latter spacing is crucial to Gasser and Müller's argument.)

Remark 3.2 The condition that the design density f be compactly supported may be relaxed by incorporating a truncation argument into the proof in Sect. 4.

Remark 3.3 Condition (3.1) may be interpreted as asking that there be sufficiently many design points, or in other words that the density of design points be sufficiently high. Note that any bandwidth h which decreases sufficiently slowly to zero as $\lambda \to \infty$, satisfies $h + \lambda^{-1} + h^{2r} \lambda^d \to 0$; and then, provided only that $n \to \infty$ sufficiently rapidly, (3.1) holds.

Remark 3.4 The ratio $(\int f^2)^{-2} \int f^4$ is minimized, subject to f having a given support, when f is uniform on that support. Related work on functions of random fields in this context has been treated in the Soviet literature; see for example Ivanov and Leonenko (1989, Chap. 4) and Kovalchuk (1987). While uniform design is a desirable property from the viewpoint of optimality, it is our experience that the experimenter is almost never able to choose the type of design. Ideally, design would be on a regular lattice, as this further improves performance, but we have seen very few lattice-supported spatial designs which do not originate in an image processing context.

Remark 3.5 It is not absolutely essential to impose the assumption of r derivatives at all places in d-dimensional space. In particular, if the assumption fails at some point, but nevertheless the total contribution from the neighbourhood of that point to integrated squared bias of the estimator is of smaller order than λ^{-d}, then the validity of the theorem is not affected. In practice the only place where the assumption of r derivatives is likely to cause serious problems is at the origin. We pause briefly here to consider that situation in the special case $d = r = 2$; other cases are similar.

Suppose that all second derivatives of ρ at the origin behave like $|t|^{\alpha-2}$ as $|t| \to 0$, for some $\alpha > 1$. For example, this would be the case if $\rho(t)$ were equal to $C_1 \exp(-C_2 |t|^\alpha)$, for $1 \leqq \alpha < 2$. Then it may be shown that the squared bias of $\check{\rho}$ is bounded above by a constant multiple of $h^4 |t|^{2\alpha-4}$, and so, since $|t|^{2\alpha-4}$ is integrable over any neighbourhood of the origin, the contribution to the integral of squared bias in any such neighbourhood is equal to $O(h^4)$, as required.

The restriction that α be greater than 1 may be relaxed if we strengthen (3.1) somewhat. For example, still in the case $d = r = 2$, suppose $\rho(t)$ were equal to $C_1 \exp(-C_2 |t|^\alpha)$ for some $0 < \alpha < 1$. We claim that if we append to (3.1) the condition that $\lambda^d h^{2+2\alpha-\varepsilon} \to 0$ for some $\varepsilon > 0$, then we may choose $\beta \in (0, \alpha)$ such that the integral of squared bias is of smaller order than λ^{-d}, as required. To appreciate why, note that the squared bias of $\check{\rho}$ may be shown to be bounded above by a constant multiple of $h^{2+2\beta} |t|^{2\alpha-2\beta-2}$, for any $0<\beta<1$, in any neighbourhood of the origin. Taking $0<\beta<\alpha$ we ensure that the integral of $|t|^{2\alpha-2\beta-2}$ over any neighbourhood of the origin is finite, and that the integral of squared bias is of order $h^{2+2\beta}$. By the assumption made earlier, this is of smaller order than λ^{-d}.

These remarks are applicable without change to all the results in this paper, not just to Theorem 3.1.

3.3 *Integrated squared error of $\tilde{\rho}$*

We assume that $\tilde{\rho}=\check{\rho}w$, where the weight function w has the properties that $|w|\leqq 1$ and, for real numbers $0<q_1<q_2<\infty$,

$$w(t)=\begin{cases}1 & \text{if } |t|\leqq q_1\\ 0 & \text{if } |t|>q_2.\end{cases}$$

We do allow w to be a function of the data, although we do not permit q_1 or q_2 to depend on the data. This restriction may be removed at the expense of a longer proof of the theorem below.

We ask that $q_1, q_2 \to \infty$ in such a manner that

(3.4) $$\lambda^d \int_{|t|>q_1} \rho(t)^2 \to 0,$$

(3.5) $$\lambda^{-1} q_2 = O(n^{-\varepsilon})$$

for some $\varepsilon>0$. We further ask that for some $\alpha>d/2$ and $c>0$,

(3.6) $$\sup_{0\leqq j\leqq r} \sup_{s:|s-t|\leqq 1} |\rho_{i_1\ldots i_j}(s)| \leqq C(1+|t|)^{-\alpha},$$

all $t\in\mathbb{R}^d$, where $\rho_{i_1\ldots i_j}(t)$ denotes the r-fold derivative of ρ with respect to the components of t_1 with indices $i_1, \ldots, i_j$; and that the function

$$r_1(t, u) = \sup_{s:|s-t|\leqq 1} |r(s, u, t)|,$$

where

$$r(s, u, t) = E[\{X(0)\, X(s) - \rho(s)\}\{X(u)\, X(u+t) - \rho(t)\}],$$

be integrable. (Note that condition (3.6) implies the finiteness of the integral in (3.4).)

Define the process Y as in subsection 3.2.

Theorem 3.2 *Assume the conditions imposed in Theorem* 3.1, *and also those stated above. Then* $\int E(Y^2) < \infty$, *implying that* $\int Y^2 < \infty$ *with probability one; and*

$$\lambda^d \int (\tilde{\rho} - \rho)^2 \to \int Y^2$$

in distribution as $n \to \infty$.

Remark 3.6 As an example, note that all the conditions on ρ imposed in Theorems 3.1 and 3.2 hold if the process X is Gaussian, with $\rho(t) = 0$ for sufficiently large $|t|$ and ρ having r bounded derivatives. In particular, this assumption is sufficient for (3.2), (3.3), (3.4) and (3.6). As noted in Remark 3.3, condition (3.1) amounts to an assumption about the density of design points; and (3.5) asks only that $\tilde{\rho}$ vanish outside a sphere which grows no faster than $\lambda n^{-\varepsilon}$, for an arbitrarily small $\varepsilon > 0$.

3.4 Integrated squared error of $\hat{\rho}$

Recall from Sect. 2 that we define $\tilde{\psi}$ to be the Fourier transform of $\tilde{\rho}$, $\hat{\psi}$ to be the positive part of $\tilde{\psi}$, and $\hat{\rho}$ to be the function whose Fourier transform is $\hat{\psi}$; see (2.2). We showed in Sect. 2 that $\int (\hat{\rho} - \rho)^2 \leqq \int (\tilde{\rho} - \rho)^2$. Our aim in the present subsection is to prove that this inequality is asymptotically an equality, in the sense that

$$\left\{\int (\hat{\rho} - \rho)^2\right\} / \left\{\int (\tilde{\rho} - \rho)^2\right\} \to 1 \tag{3.7}$$

in probability. It then follows from Theorem 3.2 that $\lambda^d \int (\hat{\rho} - \rho)^2 \to \int Y^2$ in distribution, so that $\hat{\rho}$ and $\tilde{\rho}$ share the same first-order asymptotic behaviour.

Since $\tilde{\psi}$ is defined in terms of an integral of $\tilde{\rho}$, rather than an integral of the square of $\tilde{\rho}$, then we shall have to impose versions of conditions (3.4) and (3.6) which are appropriate to the L^1 rather than the L^2 metric. Therefore we ask that the numbers q_1 and q_2, introduced in subsection 3.3 to describe the weight function w, satisfy

$$\lambda^{d/2} \int_{|t| > q_1} |\rho(t)| \, dt \to 0, \tag{3.8}$$

and for some $\alpha > d$ and $C > 0$,

$$\sup_{0 \leqq j \leqq r} \; \sup_{s : |s - t| \leqq 1} |\rho_{i_1 \ldots i_j}(s)| \leqq C(1 + |t|^{-\alpha}). \tag{3.9}$$

(The latter condition implies (3.6).) Define the process Y as in subsection 3.2.

Theorem 3.3 *Assume the conditions introduced in Theorem* 3.1, *that* ψ *vanishes on a set of measure zero, and also that* (3.8) *and* (3.9) *hold. Then* (3.7) *holds, and so* $\lambda^d \int (\hat{\rho} - \rho)^2 \to \int Y^2$ *in distribution as* $n \to \infty$.

Remark 3.7 We should comment on our assumption that ψ vanishes on a set of measure zero. Our proof of Theorem 3.3 relies critically on the result that $P\{\tilde{\psi}(\theta)\leq 0\}\to 0$ for almost all θ, which holds true if and only if the set $\{\theta\colon \psi(\theta)=0\}$ is of measure zero. Should the latter condition fail then $\hat{\rho}$ can outperform $\tilde{\rho}$, in the sense of mean squared error.

4 Proofs

Proof of Theorem 3.1

Step (a). *Preliminaries.* We may assume, and do throughout our proof, that $E(X)=0$. Define

$$D_1(t)=\sum_i \sum_j K\{(t-t_{ij})/h\}\{X(t_i)\,X(t_j)-\rho(t_{ij})\},$$

$$D_2(t)=\sum_i \sum_j K\{(t-t_{ij})/h\},$$

$$D_3(t)=\sum_i \sum_j K\{(t-t_{ij})/h\}\;\rho(t_{ij}),$$

$$D_{41}(t)=\sum_i \sum_j K\{(t-t_{ij})/h\}\;X(t_i).$$

Let D_{42} have the same definition as D_{41} but with $X(t_j)$ replacing $X(t_i)$ in the double series, and put $D_4=D_{41}+D_{42}$. In this notation,

$$\breve{\rho}=(D_1+D_3+D_2\bar{X}^2-D_4\bar{X})D_2^{-1}. \tag{4.1}$$

Let $\mathscr{B}_c$ denote the d-dimensional sphere, of radius c, centred at the origin. Given a function g from $\mathbb{R}^d$ to $\mathbb{R}$, define

$$\|g\|'=\Big(\int_{\mathscr{B}_c} g^2\Big)^{1/2}.$$

In view of (4.1),

$$\big|\,\|\breve{\rho}-\rho\|'-\|D_1D_2^{-1}\|'\big|\leq\|D_3D_2^{-1}-\rho\|'+\bar{X}^2\,v(\mathscr{B}_c)^{1/2}+|\bar{X}|\;\|D_4D_2^{-1}\|',$$

where $v(\mathscr{B}_c)$ denotes the content of $\mathscr{B}_c$. Therefore, in order to prove the theorem it suffices to show that

$$\limsup_{\lambda\to\infty}\lambda^d\,E(\bar{X}^2)<\infty, \tag{4.2}$$

and as $n \to \infty$,

(4.3) $$\int_{\mathscr{B}_c} E(D_4^2) D_2^{-2} \to 0,$$

(4.4) $$\int_{\mathscr{B}_c} (D_3 D_2^{-1} - \rho)^2 = o(\lambda^{-d}),$$

(4.5) $$\lambda^d \int_{\mathscr{B}_c} (D_1 D_2^{-1})^2 \to \int_{\mathscr{B}_c} Y^2$$

in distribution.

Step (b). *Proof of* (4.2). Observe that

$$n^{-2} \sum_{i<j} \sum E[\rho\{\lambda(U_i - U_j)\}] = \tfrac{1}{2}(1-n^{-1}) \iint \rho\{\lambda(u-v)\} f(u) f(v)\, du\, dv$$
$$\sim \tfrac{1}{2} \lambda^{-d} (\textstyle\int f^2)(\int \rho) = O(\lambda^{-d}).$$

Therefore, if we prove that

(4.6) $$A \equiv n^{-2} \sum_{i<j} \sum [\rho\{\lambda(U_i - U_j)\} - E\rho\{\lambda(U_i - U_j)\}] = O(\lambda^{-d})$$

with probability one, it will follow that

$$a \equiv n^{-2} \sum_{i<j} \sum \rho\{\lambda(u_i - u_j)\} = O(\lambda^{-d}).$$

From this result, the formula $E(\bar{X}^2) = n^{-1} \rho(0) + 2a$, and the fact that $n^{-1} \lambda^d \to 0$, follows the desired result (4.2).

Result (4.6) may be proved using the martingale argument described in Step (d) below, which in fact yields $A = o(\lambda^{-3d/2})$ with probability one. The context here is analogous to that in (4.10), but with $h=1$ and $t=0$. The argument in Step (d) is rather more complex than would be required for the present, simpler case, and so it seems best to illustrate the argument there.

Step (c). *Proof of* (4.3). In Step (d) we shall derive the second of the following two results:

(4.7) $$D_2(t) = n(n-1)(\lambda^{-1} h)^d \cdot \{\textstyle\iint K(w) f(u) f(u - \lambda^{-1} t + \lambda^{-1} h w)\, du\, dw + o(\lambda^{-d/2})\}$$

(4.8) $$D_3(t) = n(n-1)(\lambda^{-1} h)^d \cdot \{\rho(t) \textstyle\iint K(w) f(u) f(u - \lambda^{-1} t + \lambda^{-1} h w)\, du\, dw + o(\lambda^{-d/2})\}$$

uniformly in $|t| \leqq c$. The first of these results may be established similarly, although more simply. In view of (4.7),

$$\int_{\mathscr{B}_c} E(D_4^2) D_2^{-2} \sim \{n^2(\lambda^{-1} h)^d \int f^2\}^{-2} \int_{\mathscr{B}_c} E(D_4^2).$$

Therefore, it suffices to prove that

$$\int_{\mathscr{B}_c} E(D_{41}^2) = o(n^4 \lambda^{-2d} h^{2d}). \tag{4.9}$$

The case of D_{42} is similar.

Observe that

$$\begin{aligned}\int_{\mathscr{B}_c} E(D_{41}^2) &= \sum_{i_1} \sum_{j_1} \sum_{i_2} \sum_{j_2} \rho(t_{i_1 i_2}) \\ &\quad \cdot \int_{|t| \leqq c} K\{(t - t_{i_1 j_1})/h\} K\{(t - t_{i_2 j_2})/h\} \, dt \\ &= h^d \sum_{i_1} \sum_{j_1} \sum_{i_2} \sum_{j_2} \rho(t_{i_1 i_2}) \\ &\quad \cdot \int_{|t_{i_1 j_1} + hw| \leqq c} K(w) K\{w + (t_{i_1 j_1} - t_{i_2 j_2}) h^{-1}\} \, dw.\end{aligned}$$

Since $t_{ij} = \lambda(u_i - u_j)$, and K is compactly supported, there exists a constant $c_1 > 0$ such that

$$\begin{aligned}\int_{\mathscr{B}_c} E(D_{41}^2) &\leqq (\sup K^2)\, h^d \sum_{i_1} \sum_{j_1} \sum_{i_2} \sum_{j_2} |\rho\{\lambda(u_{i_1} - u_{i_2})\}| \\ &\quad \cdot I(|u_{i_1} - u_{j_1}| \leqq c_1 \lambda^{-1}, |(u_{j_1} + u_{i_2} - u_{i_1}) - u_{j_2}| \leqq c_1 \lambda^{-1} h) \\ &\leqq (\sup K^2)\, h^d \{\sup_u \sum_{j_1} I(|u - u_{j_1}| \leqq c_1 \lambda^{-1})\} \\ &\quad \cdot \{\sup_u \sum_{j_2} I(|u - u_{j_2}| \leqq c_1 h \lambda^{-1})\} \sum_{i_1} \sum_{i_2} |\rho\{\lambda(u_{i_1} - u_{i_2})\}| \\ &= O\{h^d \cdot n \lambda^{-d} \cdot n (h \lambda^{-1})^d \cdot (n + n^2 \lambda^{-d})\} = O(n^4 \lambda^{-3d} h^{2d}).\end{aligned}$$

This proves (4.9).

Step (d). *Proof of* (4.4). Result (4.4) is an immediate consequence of (4.7) and (4.8). The latter two formulae may be proved similarly, and so we concentrate here on deriving (4.8).

Define

$$\begin{aligned} D_{31}(t) &= n K(t/h)\, \rho(0), \\ D_{32}(t) &= \sum_{i \neq j} \sum K[\{t - \lambda(u_i - u_j)\}/h]\, \rho\{\lambda(u_i - u_j)\}. \end{aligned} \tag{4.10}$$

Since $D_3 = D_{31} + D_{32}$, and $\sup |D_{31}| = O(n) = o(n^2 \lambda^{-3d/2} h^d)$ under the conditions of the theorem, then it suffices to prove that if each u_i is replaced by U_i,

$$(4.11)\quad E\{D_{32}(t)\} = n(n-1)(\lambda^{-1} h)^d \cdot \{\rho(t) \iint K(w) f(u) f(u - \lambda^{-1} t + \lambda^{-1} hw)\, du\, dw + o(\lambda^{-d/2})\},$$

and with U-probability one,

$$(4.12)\qquad D_{32}(t) - E\{D_{32}(t)\} = o(n^2 \lambda^{-3d/2} h^d).$$

Both (4.11) and (4.12) must hold uniformly in $|t| \leqq c$.

Result (4.11) follows from the identity

$$D_{32}(t) = n(n-1)(\lambda^{-1} h)^d \iint K(w) f(u) f(u - \lambda^{-1} t + \lambda^{-1} hw)\, \rho(t - hw)\, du\, dw,$$

after a little Taylor expansion. (Note that $h^r = o(\lambda^{-d/2})$.) It remains to verify (4.12). Define

$$a(U_i, U_j) = K[\{t - \lambda(U_i - U_j)\}/h]\, \rho\{\lambda(U_i - U_j)\},$$

$$a(U_i) = E\{a(U_i, U_j) \mid U_i\} \quad (\text{for } i<j), \qquad \alpha = E\{a(U_i)\},$$

$$b(U_i, U_j) = a(U_i, U_j) - a(U_i), \qquad b(U_i) = a(U_i) - \alpha,$$

$$Z_j = \sum_{i=1}^{j-1} b(U_i, U_j).$$

In this notation, that part of $D_{32}(t) - E\{U_{32}(t)\}$ which corresponds to summing over $i<j$ in (4.10) is given by

$$\sum_{j=2}^{n} Z_j + \sum_{i=1}^{n-1} (n-i)\, b(U_i).$$

Therefore it suffices to prove that

$$(4.13)\qquad \sum_{j=2}^{n} Z_j = o(n^2 \lambda^{-3d/2} h^d),$$

$$(4.14)\qquad \sum_{i=1}^{n-1} (n-i)\, b(U_i) = o(n^2 \lambda^{-3d/2} h^d)$$

uniformly in $|t| \leqq c$.

Next we outline the method of proof used to derive (4.13) and (4.14). Let $A = A(t)$ denote the left-side of either identity. In view of the Hölder continuity of K it suffices to prove that for any set $\mathcal{T}_n \subseteq \mathcal{B}_c$, containing at most $O(n^M)$ elements for some $M > 0$, both (4.13) and (4.14) hold uniformly in $t \in \mathscr{T}_n$. Using the Borel-Cantelli Lemma, and bounding the probability that the supremum of $|A(t)|$ over $\mathscr{T}_n$ exceeds a constant, by the sum of the probabilities that individual values of $|A(t)|$ exceed the constant, we see that it suffices to show that for each $N>0$ there exists an integer $k \geqq 1$, chosen sufficiently large, such that

$$\sup_{t \in \mathscr{B}_c} E\{A(t)^{2k}\} = O(n^2 \lambda^{-3d/2} h^d)^{2k} n^{-N}. \tag{4.15}$$

When $A(t)$ equals the left-hand side of (4.14), result (4.15) is straightforward to prove. There, by Rosenthal's inequality (Hall and Heyde 1980, p. 23), and uniformly in $t \in \mathscr{B}_c$,

$$E\{A(t)^{2k}\} = O[n^{3k} E\{b(U_1)^2\}^k] = O\{n^{3k}(\lambda^{-1} h)^{2dk}\},$$

which, in view of the conditions on n, h and λ, implies (4.15) for sufficiently large k.

Now suppose that $A(t)$ equals the left-hand side of (4.13). Since $E(Z_j | U_1, \ldots, U_{j-1}) = 0$ then the Z_j's are martingale differences. Hence, by Rosenthal's inequality for martingales, and for constants $C_j(k)$ depending only on k,

$$\begin{aligned} E\left(\sum_{j=2}^n Z_j\right)^{2k} &\leqq C_1(k)\left[E\left\{\sum_{j=2}^n E(Z_j^2 | U_1, \ldots, U_{j-1})\right\}^k + \sum_{j=2}^n E(Z_j^{2k})\right\} \\ &\leqq C_1(k)(n^{k-1}+1) \sum_{j=2}^n E(Z_j^{2k}), \end{aligned}$$

$$\begin{aligned} E(Z_j^{2k}) &= E\{E(Z_j^{2k} | U_j)\} \\ &\leqq 2^{2k-1} E\left[E\left\{\left(\sum_{i=1}^{j-1} [b(U_i, U_j) - E\{b(U_i, U_j) | U_j\}]\right)^{2k} \middle| U_j\right\}\right. \\ &\qquad \left. + [(j-1) E\{b(U_1, U_j) | U_j\}]^{2k}\right] \\ &\leqq 2^{2k-1} C_1(k)(E[(j-1) \operatorname{var}\{b(U_1, U_j) | U_j\}]^k \\ &\qquad + (j-1) E\,|b(U_1, U_j) - E\{b(U_1, U_j) | U_j\}|^{2k} \\ &\qquad + n^{2k} E[E\{b(U_1, U_j) | U_j\}]^{2k}) \\ &\leqq C_2(k)\{n^k(\lambda^{-1} h)^{dk} + n(\lambda^{-1} h)^d + n^{2k}(\lambda^{-1} h)^{2dk}\}. \end{aligned}$$

Therefore,

$$E\left(\sum_{j=2}^n Z_j\right)^{2k} \leqq C_3(k)\{n^{2k}(\lambda^{-1} h)^{dk} + n^{k+1}(\lambda^{-1} h)^d h^{-dk} + n^{3k}(\lambda^{-1} h)^{2dk}\},$$

which, in view of the conditions on n, h and λ, implies (4.15).

Step (e). *Preliminaries for proof of* (4.5). In view of (4.7),

$$\int_{\mathscr{B}_c} (D_1 D_2^{-1})^2 = \{n^2(\lambda^{-1}h)^d \textstyle\int f^2\}^{-2}\{1+o_p(1)\} \displaystyle\int_{\mathscr{B}_c} D_1^2.$$

Therefore it suffices to prove that

$$\{n^2(\lambda^{-1}h)^d \textstyle\int f^2\}^{-2}\lambda^d \displaystyle\int_{\mathscr{B}_c} D_1^2 \to \int_{\mathscr{B}_c} Y^2 \tag{4.16}$$

in distribution. The present step will show that for (4.16), it is enough to prove that the finite-dimensional distributions of $\{n^2(\lambda^{-1}h)^d \int f^2\}^{-1}\lambda^{d/2}D_1$ converge to those of Y. The latter result will be verified in Step (f).

Let $\mathscr{S}_m = \{s_1, \ldots, s_m\}$ denote a set of distinct elements of $\mathscr{B}_c$. Assume that the points s_i are chosen in such a way that as $m\to\infty$, the greatest distance between any point in $\mathscr{B}_c$ and any point in $\mathscr{S}_m$ converges to zero. Given $t\in\mathscr{B}_c$, let $s(t)$ denote that element of $\mathscr{S}_m$ which is nearest to t, with ties being broken in an arbitrary manner. Put $D_5(t)=D_1\{s(t)\}$. It may be proved, straightforwardly but with tedious algebra in the case of the second result, that

$$\sup_{t\in\mathscr{B}_c} E\{D_1(t)^2\} = O(n^4\lambda^{-3d}h^{2d}),$$

$$\lim_{m\to\infty}\limsup_{n\to\infty}(n^4\lambda^{-3d}h^{2d})^{-1}\sup_{t\in\mathscr{B}_c}E\{D_1(t)-D_5(t)\}^2=0.$$

These results, and the fact that

$$E\left|\int_{\mathscr{B}_c} D_1^2 - \int_{\mathscr{B}_c} D_5^2\right| \le \left\{\int_{\mathscr{B}_c} E(D_1-D_5)^2\right\}^{1/2}\left\{\left(\int_{\mathscr{B}_c} ED_1^2\right)^{1/2}+\left(\int_{\mathscr{B}_c} ED_5^2\right)^{1/2}\right\},$$

imply that

$$\lim_{m\to\infty}\limsup_{n\to\infty}(n^4\lambda^{-3d}h^{2d})^{-1}E\left|\int_{\mathscr{B}_c} D_1^2 - \int_{\mathscr{B}_c} D_5^2\right| = 0. \tag{4.17}$$

Let $L(W_1, W_2)$ denote a metric distance (such as distance in the Lévy metric) between the distributions of W_1 and W_2. In view of (4.17), (4.16) will follow if we prove that

$$\lim_{m\to\infty}\limsup_{n\to\infty} L\left[\{n^2(\lambda^{-1}h)^d \textstyle\int f^2\}^{-2}\lambda^d \displaystyle\int_{\mathscr{B}_c} D_5^2, \int_{\mathscr{B}_c} Y^2\right]=0.$$

A sufficient condition for the latter result is convergence of the finite-dimensional distributions of $\{n^2(\lambda^{-1}h)^d \int f^2\}^{-1}\lambda^{d/2}D_1$ to those of Y.

Step (f). *Convergence of finite-dimensional distributions of* D_1. We give the proof only in outline, beginning with two continuous approximations to D_1. Put

$$D_{11}(t)=n(n-1)\iint K[\{t-\lambda(u-v)\}/h][X(\lambda u)\,X(\lambda v)$$
$$-\rho\{\lambda(u-v)\}\,f(u)f(v)\,du\,dv$$
$$=n(n-1)(\lambda^{-1}h)^d\iint K(w)\{X(\lambda u)\,X(\lambda u-t+hw)$$
$$-\rho(t-hw)\}\,f(u)f(u-\lambda^{-1}t+\lambda^{-1}hw)\,du\,dw$$
$$D_{12}(t)=n(n-1)(\lambda^{-1}h)^d\int\{X(\lambda u)\,X(\lambda u-t)-\rho(t)\}\,f(u)^2\,du.$$

It is straightforward, although tedious, to prove that for each fixed $t\in\mathbb{R}^d$,

$$E\{D_1(t)-D_{11}(t)\}^2+E\{D_{11}(t)-D_{12}(t)\}^2=o(n^4\lambda^{-3d}h^{2d}).$$

Therefore it suffices to show that the finite-dimensional distributions of

$$(n^2\lambda^{-3d/2}h^d\int f^2)^{-1}D_{12}$$

converge to those of Y. For the sake of simplicity we confine attention to a single dimension. Derivation of convergence in distribution for a multivariate sequence is similar.

Divide $\mathbb{R}^d$ into a regular lattice of d-dimensional cubes with edge widths α, such that the faces of adjacent cubes are parallel and separated by distance β. Enumerate the cubes, with $\mathscr{C}_j$ denoting the j'th cube and $\mathscr{D}$ the union of all the strips separating the cubes. If $\mathscr{A}$ denotes either $\mathscr{C}_j$ or $\mathscr{D}$, define $\lambda^{-1}\mathscr{A}=\{\lambda^{-1}x\colon x\in\mathscr{A}\}$. Put

$$Z_j=\int_{\lambda^{-1}\mathscr{C}_j}\{X(\lambda u)\,X(\lambda u-t)-\rho(t)\}\,f(u)^2\,du,$$

$$R=\int_{\lambda^{-1}\mathscr{D}}\{X(\lambda u)\,X(\lambda u-t)-\rho(t)\}\,f(u)^2\,du.$$

In this notation,

$$\{n(n-1)(\lambda^{-1}h)^d\}^{-1}D_{12}=\sum Z_j+R. \tag{4.18}$$

The series on the right-hand side of (4.18) may be confined to the sum over $j\in\mathscr{J}$, where $\mathscr{J}$ denotes the set of indices j such that $\lambda^{-1}\mathscr{C}_j$ has nonempty intersection with the support of f. Let J be the number of elements of $\mathscr{J}$. We assume that $\beta\leqq\alpha$, in which case $J\leqq C(\lambda/\alpha)^d$, where the constant $C>0$ does not depend on α, β or λ.

Put $\beta=\lambda^{1/2}$ and choose $x=x(\lambda)$ to increase to $+\infty$ as $\lambda\to\infty$, at a rate which is sufficiently slow to ensure that $x/\lambda^{1/2}\downarrow 0$ and $x^d\chi(\beta/2)\to 0$. Define $\alpha=\lambda/x$, then

$$\begin{aligned}\alpha\lambda^{-1}+\beta\alpha^{-1}&=x^{-1}+x\lambda^{-1/2}\to 0,\\ (\lambda/\alpha)^d\,\chi(\beta/2)&=x^d\,\chi(\beta/2)\to 0.\end{aligned}\tag{4.19}$$

The first of these relations implies that the content of the intersection of $\lambda^{-1}\mathscr{D}$ with the support of f converges to zero as $\lambda\to\infty$, and so

$$\begin{aligned}E(R^2)&\leqq\int_{\lambda^{-1}\mathscr{D}}\int_{\lambda^{-1}\mathscr{D}} f(u)^2 f(v)^2\,|E[\{X(\lambda u)\,X(\lambda u-t)-\rho(t)\}\\ &\quad\cdot\{X(\lambda v)\,X(\lambda v-t)-\rho(t)\}]|\,du\,dv\\ &\leqq\lambda^{-d}(\sup f^2)\Big(\int_{\lambda^{-1}\mathscr{D}} f^2\Big)\Big[\int_{\mathbb{R}^d}|E[\{X(0)\,X(-t)-\rho(t)\}\\ &\quad\cdot\{X(w)\,X(w-t)-\rho(t)\}]|\,dw\Big]\\ &=o(\lambda^{-d}).\end{aligned}$$

That is, $\lambda^d E(R^2)\to 0$ as $\lambda\to\infty$. Hence, in view of (4.18) the desired convergence of the distribution of $D_{12}(t)$ will follow if we show that for each $s\in\mathbb{R}$,

$$E[\exp\{is\lambda^{d/2}(\textstyle\int f^2)^{-1}\sum Z_j\}]\to E[\exp\{is\,Y(t)\}],\tag{4.20}$$

where $i=\sqrt{-1}$.

Put $\xi=\lambda^{d/2}(\int f^2)^{-1}$. Let δ denote the diameter of the smallest d-dimensional sphere containing the support of f, and choose λ so large that $\beta>2(\delta+2|t|)$. Let $s\in\mathbb{R}$. If $\mathscr{K}$ is any subset of $\mathscr{J}$ and if $j_0\in\mathscr{K}$ then

$$\begin{aligned}&|E\{\exp(is\xi\sum_{j\in\mathscr{K}}Z_j)\}-E\{\exp(is\xi\sum_{j\in\mathscr{K}\setminus\{j_0\}}Z_j)\}\\ &\quad\cdot E\{\exp(is\xi Z_{j_0})\}|\leqq 2\chi(\beta-\delta-2|t|).\end{aligned}$$

Iterating over all $j_0\in\mathscr{J}$ we deduce that

$$\begin{aligned}|E\{\exp(is\xi\sum_{j\in\mathscr{J}}Z_j)\}-\prod_{j\in\mathscr{J}}E\{\exp(is\xi Z_j)\}|&\leqq 2J\chi(\beta-\delta-2|t|)\\ &\leqq 2C(\lambda/\alpha)^d\,\chi(\beta/2)\\ &\to 0,\end{aligned}$$

by (4.19). Therefore, (4.20) will follow if we prove that

$$\prod_{j\in\mathscr{J}}E\{\exp(is\xi Z_j)\}\to E[\exp\{is\,Y(t)\}],$$

or equivalently, that if Z'_j, $j \in \mathscr{J}$, are independent random variables with the respective distributions of Z_j, then

$$\xi \sum_{j \in \mathscr{J}} Z'_j \to Y(t) \tag{4.21}$$

in distribution.

To prove (4.21), observe that $E(Z'_j)=0$,

$$\begin{aligned}
\sum_{j \in \mathscr{J}} E(Z'^2_j) = \sum_{j \in \mathscr{J}} E(Z^2_j)& \\
&= \lambda^{-d} \sum_{j \in \mathscr{J}} \int_{\lambda^{-1}\mathscr{C}_j} f(u)^2 \, du \int_{\mathscr{C}_j - \lambda u} f(u+\lambda^{-1} v)^2 \\
&\quad \cdot E[\{X(0)\, X(-t) - \rho(t)\}\{X(v)\, X(v-t) - \rho(t)\}]\, dv \\
&\sim \lambda^{-d} \sum_{j \in \mathscr{J}} \Big(\int_{\lambda^{-1}\mathscr{C}_j} f^4 \Big) \int E[\{X(0)\, X(t) - \rho(t)\} \\
&\quad \cdot \{X(v)\, X(v+t) - \rho(t)\}]\, dv \\
&= \lambda^{-d} \Big(\int f^4\Big) \int E[\{X(0)\, X(t) - \rho(t)\}\{X(v)\, X(v+t) - \rho(t)\}]\, dv \\
&= \xi^{-2} E\{Y(t)^2\},
\end{aligned}$$

and

$$\begin{aligned}
\sum_{j \in \mathscr{J}} E(Z'^4_j) = \sum_{j \in \mathscr{J}} E(Z^4_j)& \\
&= \lambda^{-3d} \sum_{j \in \mathscr{J}} \int_{\lambda^{-1}\mathscr{C}_j} f(u)^2 \, du \int_{\mathscr{C}_j - \lambda u} f(u+\lambda^{-1} v_1)^2 \, dv_1 \\
&\quad \cdot \int_{\mathscr{C}_j - \lambda u} f(u+\lambda^{-1} v_2)^2 \, dv_2 \int_{\mathscr{C}_j - \lambda u} f(u+\lambda^{-1} v_3)^2 \\
&\quad \cdot E[\{X(0)\, X(-t) - \rho(t)\}\{X(v_1)\, X(v_1 - t) - \rho(t)\} \\
&\quad \cdot \{X(v_2)\, X(v_2 - t) - \rho(t)\}\{X(v_3)\, X(v_3 - t) - \rho(t)\}]\, dv_3 \\
&\leqq \lambda^{-3d} (\sup f^6) \Big(\int f^2\Big) \int | E[\{X(0)\, X(-t) - \rho(t)\} \\
&\quad \cdot \{X(v_1)\, X(v_1 - t) - \rho(t)\}\{X(v_2)\, X(v_2 - t) - \rho(t)\} \\
&\quad \cdot \{X(v_3)\, X(v_3 - t) - \rho(t)\}]|\, dv_1\, dv_2\, dv_3 \\
&= O(\lambda^{-3d}).
\end{aligned}$$

Therefore,

$$\sum_{j_1 \in \mathscr{J}} E\, |Z'_{j_1} / \{ \sum_{j_2 \in \mathscr{J}} E(Z'^2_{j_2})\}^{1/2}|^4 = O\{\lambda^{-3d} (\lambda^{-d})^{-2}\} = O(\lambda^{-d}).$$

Result (4.21) now follows via Liapunov's central limit theorem.

Proof of Theorem 3.2 We know from Theorem 3.1 and the fact that $q_1 \to \infty$, that for each fixed $c>0$ and as $n \to \infty$,

$$\lambda^d \int_{|t| \leqq c} \{\tilde{\rho}(t)-\rho(t)\}^2 \, dt = \lambda^d \int_{|t| \leqq c} \{\check{\rho}(t)-\rho(t)\}^2 \, dt \to \int_{|t| \leqq c} Y(t)^2 \, dt,$$

the convergence being in distribution. It is straightforward to prove that $\int E(Y^2) < \infty$, so that

$$\int_{|t|>c} E\{Y(t)^2\} \, dt \to 0$$

as $c \to \infty$. Therefore, the proof of Theorem 3.2 will be complete if we show that

$$\lim_{c \to \infty} \limsup_{n \to \infty} \lambda^d \int_{|t|>c} E\{\tilde{\rho}(t)-\rho(t)\}^2 \, dt = 0.$$

Given $0<a<b<\infty$, define $\mathscr{B}(a, b) = \{t \in \mathbb{R}^b : a < |t| \leqq b\}$. Then

$$\begin{aligned} \int_{|t|>c} \{\tilde{\rho}(t)-\rho(t)\}^2 \, dt &= \int_{\mathscr{B}(c,\infty)} (\tilde{\rho}-\rho)^2 \\ &\leqq 2 \int_{\mathscr{B}(c,\infty)} (\check{\rho}-\rho)^2 w^2 + 2 \int_{\mathscr{B}(c,\infty)} \rho^2 (w-1)^2 \\ &\leqq 2 \int_{\mathscr{B}(c,q_2)} (\check{\rho}-\rho)^2 + 2 \int_{\mathscr{B}(q_1,\infty)} \rho^2. \end{aligned}$$

By hypothesis, the last-written integral equals $o(\lambda^{-d})$, and so it suffices to prove that

$$\text{(4.22)} \qquad \lim_{c \to \infty} \limsup_{n \to \infty} \lambda^d \int_{\mathscr{B}(c,q_2)} E(\check{\rho}-\rho)^2 = 0.$$

Let $D_1, \ldots, D_4, D_{41}, D_{42}$ be as defined during the proof of Theorem 3.1, and recall formula (4.1) describing the decomposition of $\check{\rho}$. Arguing in the manner there we see that (4.22) will follow if we show that

$$\text{(4.23)} \qquad \lim_{n \to \infty} \lambda^{-d} \int_{\mathscr{B}(c,q_2)} 1 = 0,$$

$$\text{(4.24)} \qquad \lim_{c \to \infty} \limsup_{n \to \infty} \int_{\mathscr{B}(c,q_2)} E(D_4^2) D_2^{-2} = 0,$$

$$\text{(4.25)} \qquad \lim_{c \to \infty} \limsup_{n \to \infty} \lambda^d \int_{\mathscr{B}(c,q_2)} (D_3 D_2^{-1} - \rho)^2 D_2^{-2} = 0,$$

$$\text{(4.26)} \qquad \lim_{c \to \infty} \limsup_{n \to \infty} \lambda^d \int_{\mathscr{B}(c,q_2)} E(D_1^2) D_2^{-2} = 0.$$

Formulae (4.24)–(4.26) are of course directly analogous to (4.3)–(4.5). Result (4.23) follows from the fact that $q_2/\lambda \to 0$.

The arguments employed in Step (d) of the proof of Theorem 3.1 may be modified to show that with U-probability one,

$$D_2(t)=\{1+o(1)\}\, n^2(\lambda^{-1} h)^d \int f^2,$$
$$D_3(t)=D_2(t)\{\rho(t)+O(\lambda^{-d/2}\,|t|^{-\alpha})\} \tag{4.27}$$

uniformly in $1 \leqq |t| \leqq q_2$. Result (4.25) is an immediate consequence of these two formulae. We may also deduce from (4.27) that (4.24) and (4.26) will follow if we prove, respectively, that

$$\lim_{c\to\infty} \limsup_{n\to\infty} (n^4 \lambda^{-2d} h^{2d})^{-1} \int_{\mathscr{B}(c,q_2)} E(D_4^2)=0, \tag{4.28}$$

$$\lim_{c\to\infty} \limsup_{n\to\infty} (n^4 \lambda^{-3d} h^{2d})^{-1} \int_{\mathscr{B}(c,q_2)} E(D_1^2)=0, \tag{4.29}$$

When deriving (4.28) we may confine attention to the case where $E(D_4^2)$ on the left-hand side is replaced by $E(D_{41}^2)$. This situation may be treated by modifying the argument in Step (c) of the proof of Theorem 3.1, along the following lines. Since $t_{ij}=\lambda(u_i-u_j)$ and K is compactly supported then there exists $c_1>0$ such that, for sufficiently large n,

$$\begin{aligned}
\int_{\mathscr{B}(c,q_2)} E(D_{41}^2) &= h^d \sum_{i_1}\sum_{j_1}\sum_{i_2}\sum_{j_2} \rho(t_{i_1 i_2}) \int_{c<|t_{i_1 j_1}+hw|\leqq q_2} K(w) \\
&\quad \cdot K\{w+(t_{i_1 j_1}-t_{i_2 j_2})h^{-1}\}\, dw \\
&\leqq (\sup K^2)\, h^d \sum_{i_1}\sum_{j_1}\sum_{i_2}\sum_{j_2} |\rho\{\lambda(u_{i_1}-u_{i_2})\}| \\
&\quad \cdot I(|u_{i_1}-u_{j_1}|\leqq 2\lambda^{-1} q_2,\ |(u_{j_1}+u_{i_2}-u_{i_1})-u_{j_2}|\leqq c_1 \lambda^{-1} h) \\
&\leqq (\sup K^2)\, h^d \{\sup_u \sum_{j_1} I(|u-u_{j_1}|\leqq 2\lambda^{-1} q_2)\} \\
&\quad \cdot \{\sup_u \sum_{j_2} I(|u-u_{j_2}|\leqq c_1 \lambda^{-1} h)\} \sum_{i_1}\sum_{i_2} |\rho\{\lambda(u_{i_1}-u_{i_2})\}| \\
&= O\{h^d \cdot n(\lambda^{-1} q_2)^d \cdot n(\lambda^{-1} h)^d \cdot (n+n^2\lambda^{-d})\, n^\delta\} \\
&= O(n^{4+\delta} \lambda^{-3d} h^{2d} q_2^d)
\end{aligned}$$

for all $\delta>0$. Since $q_2/\lambda=O(n^{-\varepsilon})$ for some $\varepsilon>0$ then (4.28) is true.

To derive (4.29), observe that

$$\begin{aligned}
\int_{\mathscr{B}(c,q_2)} E(D_1^2) &= h^d \sum_{i_1}\sum_{i_2}\sum_{j_1}\sum_{j_2} r(t_{i_1 j_1}, t_{i_2 j_1}, t_{i_2 j_2}) \\
&\quad \cdot \int_{c<|t_{i_1 j_1}+hw|\leqq q_2} K(w)\, K\{w+(t_{i_1 j_1}-t_{i_2 j_2})h^{-1}\}\, dw \\
&\leqq (\sup K^2)\, h^d \sum_{i_1}\sum_{i_2}\sum_{j_1}\sum_{j_2} |r(t_{i_1 j_1}, t_{i_2 j_1}, t_{i_2 j_2})| \\
&\quad \cdot I(|t_{i_1 j_1}|>\tfrac{1}{2}c,\ |t_{i_1 j_1}-t_{i_2 j_2}|\leqq c_1 h) \\
&\leqq (\sup K^2)\, h^d \sum_{i_1}\sum_{i_2}\sum_{j_1}\sum_{j_2} r_1(t_{i_1 j_1}, t_{i_2 j_1}) \\
&\quad \cdot I(|t_{i_1 j_1}|>\tfrac{1}{2}c,\ |t_{i_1 j_1}-t_{i_2 j_2}|\leqq c_1 h).
\end{aligned} \tag{4.30}$$

If in the latter formula we were to write $t_{ij}=\lambda(u_i-u_j)$, then replace each u_i by the random variable U_i, and finally take expectations, we would obtain the quantity

$$\begin{aligned}
&\{1+o(1)\}(\sup K^2)\, n^4 h^d \int \dots \int r_1\{\lambda(u_1-u_2), \lambda(u_3-u_2)\} \\
&\quad \cdot I\{|u_1-u_2|>\tfrac{1}{2} c \lambda^{-1}, |(u_1-u_2)-(u_3-u_4)| \leqq c_1 \lambda^{-1} h\} \\
&\quad \cdot f(u_1) \dots f(u_4)\, du_1 \dots du_4 \\
&\leqq \{1+o(1)\}(\sup f)(\sup K^2)\, c_1^d n^4 \lambda^{-d} h^d \\
&\quad \cdot \int \dots \int r_1\{\lambda(u_1-u_2), \lambda(u_3-u_2)\} \\
&\quad \cdot I(|u_1-u_2|>\tfrac{1}{2} c \lambda^{-1}) f(u_1) f(u_2) f(u_3)\, du_1\, du_2\, du_3 \\
&= \{1+o(1)\}(\sup f)(\sup K^2)(\textstyle\int f^3)\, n^4 \lambda^{-3d} h^{2d} \\
&\quad \cdot \iint_{|v_1|>\frac{1}{2}c,\, v_2 \in \mathbb{R}^d} r_1(v_1, v_2)\, dv_1\, dv_2 .
\end{aligned}$$

Given any $\varepsilon>0$, we may ensure that this is less than $\varepsilon n^4 \lambda^{-3d} h^{2d}$ for all sufficiently large n, simply by choosing c large. Therefore it suffices to show that

$$\begin{aligned}
(4.31)\quad &(n^4 \lambda^{-3d} h^d)^{-1} \sum_{i_1} \sum_{i_2} \sum_{j_1} \sum_{j_2} (r_1(t_{i_1 j_1}, t_{i_2 j_2})\, I(|t_{i_1 j_1}|>\tfrac{1}{2} c, |t_{i_1 j_1}-t_{i_2 j_2}| \\
&\leqq c_1 h) - E[r_1\{\lambda(U_{i_1}-U_{i_2}), \lambda(U_{i_2}-U_{j_2})\} \\
&\quad \cdot I\{\lambda\, |U_{i_1}-U_{j_1}|>\tfrac{1}{2} c, \lambda\, |(U_{i_1}-U_{j_1})-(U_{i_2}-U_{j_2})| \leqq c_1 h\}]) \to 0
\end{aligned}$$

with U-probability one. This may be done by (a) replacing each t_{ij} by $\lambda(U_i-U_j)$, (b) computing the fourth moment of the left-hand side of (4.31), and showing that it equals $O(n^{-(1+\delta)})$ for some $\delta>0$, and (c) applying the Borel-Cantelli lemma and Markov's inequality (for a fourth moment), noting that $\sum n^{-(1+\delta)}<\infty$.

Proof of Theorem 3.3 It suffices to prove that $\|\tilde{\rho}-\hat{\rho}\|=o_p(\lambda^{-d/2})$. Observe that by Parseval's identity

$$\begin{aligned}
(2\pi)^{d/2} \|\tilde{\rho}-\hat{\rho}\| &= \|\tilde{\psi} I(\tilde{\psi} \leqq 0)\| \\
&\leqq \|(\tilde{\psi}-\psi)\, I(\tilde{\psi} \leqq 0)\| + \|\psi I(\tilde{\psi} \leqq 0)\| .
\end{aligned}$$

However, if $\tilde{\psi} \leqq 0$ then $0 \leqq \psi \leqq \psi-\tilde{\psi}$, and so $\|\psi I(\tilde{\psi} \leqq 0)\| \leqq \|(\psi-\tilde{\psi})\, I(\tilde{\psi} \leqq 0)\|$. Therefore,

$$\|\tilde{\rho}-\hat{\rho}\| \leqq 2(2\pi)^{-d/2} \|(\tilde{\psi}-\psi)\, I(\tilde{\psi} \leqq 0)\|,$$

and so it suffices to prove that $\lambda^{d/2} \|(\tilde{\psi}-\psi)\, I(\tilde{\psi} \leqq 0)\| \to 0$ in probability. For this it is adequate to prove that for each $c>0$,

$$\lambda^d \int_{|\theta| \leqq c} E[\{\tilde{\psi}(\theta)-\psi(\theta)\}^2\, I\{\tilde{\psi}(\theta) \leqq 0\}]\, d\theta \to 0,$$

and that

$$\lim_{c\to\infty} \liminf_{n\to\infty} \lambda^d \int_{|\theta|\leq c} E\{\tilde{\psi}(\theta)-\psi(\theta)\}^2 \, d\theta$$
$$\geqq \lim_{n\to\infty} \lambda^d (2\pi)^d \int E(\tilde{\rho}-\rho)^2.$$

To establish these results it suffices to prove the existence of a Gaussian random field $Z=Z(\theta)$, with zero mean, such that $\lambda^{d/2}\{\tilde{\psi}(\theta)-\psi(\theta)\}\to Z(\theta)$ in distribution, for all $\theta\in\mathbb{R}^d$; $\lambda^{d/2} E\{\tilde{\psi}(\theta)-\psi(\theta)\}^2 \to E\{Z(\theta)\}^2$ uniformly on compact sets; and $\lambda^d \int E(\tilde{\rho}-\rho)^2 \to (2\pi)^{-d} \int EZ^2$.

These results may be derived using techniques from the proofs of Theorems 3.1 and 3.2. Of course, $\int EZ^2 = (2\pi)^d \int EY^2$.

Acknowledgements. We are most grateful to two referees for their exceptionally helpful remarks.

References

Armstrong, A.G., Diamond, P.: Testing variograms for positive-definiteness. Math. Geol. **16**, 407–421 (1984)

Cheng, K.F., Lin, P.E.: Nonparametric estimation of a regression function. Z. Wahrscheinlichkeitstheor. Verw. Geb. **57**, 223–233 (1981)

Christakos, G.: On the problem of permissible covariance and variogram models. Water Resources Res. **20**, 251–265 (1984)

Cressie, N.: Statistics for Spatial Data. New York: Wiley, 1991

Devroye, L., Györfi, L.: Nonparametric Density Estimation: The L_1 View. New York: Wiley, 1985

Gasser, Th., Müller, H.G.: Kernel estimation of regression functions (Lect. Notes Math., vol. 757, pp. 23–68) Berlin Heidelberg New York: Springer, 1979

Gasser, Th., Müller, H.G.: Estimating regression functions and their derivatives by the kernel method. Scand. J. Stat. **11**, 171–185 (1984)

Hall, P., Heyde, C.C.: Martingale Limit Theory and its Application. New York: Academic Press, 1980

Hall, P., Fisher, N.I., Hoffmann, B.: On the nonparametric estimation of covariance functions. Research Report, ANU, 1992

Härdle, W.: Applied Nonparametric Regression. Cambridge: University Press, 1990

Härdle, W., Marron, J.S.: Random approximations to some measures of accuracy in nonparametric curve estimation. J. Multivariate Anal. **20**, 91–113 (1986)

Ibragimov, I.A., Rozanov, Yu.A.: Gaussian Random Processes. Berlin Heidelberg New York: Springer, 1978

Ivanov, A.V., Leonenko, N.N.: Statistical Analysis of Random Fields. Dordrecht: Kluwer, 1989

Journel, A.G., Huijbregts, Ch.J.: Mining Geostatistics. London: Academic Press, 1978

Kovalchuk, T.V.: On convergence of the distribution of functionals of the correlogram of a homogenous random field with unknown mean. Theor. Probab. Math. Stat. **35**, 45–53 (1987)

Matheron, G.: The Theory of Regionalised Variables. Les Cahiers du Centre de Morphologie Mathematique de Fontainebleau. 5 (1971)

Sampson, P.D., Guttorp, P.: Nonparametric representation of nonstationary spatial covariance structure. J. Am. Stat. Assoc. **87**, 108–119 (1992)

Schoenberg, I.J.: Metric spaces and completely monotone functions. Ann. Math. **79**, 811–841 (1938)
Shapiro, A., Botha, J.D.: Variogram fitting with a general class of conditionally nonnegative definite functions. Comput. Stat. Data Anal. **11**, 87–96 (1991)
Silverman, B.W.: Density Estimation for Statistics and Data Analysis. London: Chapman and Hall, 1986