

## Edge effects and efficient parameter estimation for stationary random fields

BY R. DAHLHAUS

*Universität Essen, Fachbereich Mathematik, D-4300 Essen 1, Federal Republic of Germany*

AND H. KÜNSCH

*Seminar für Statistik, ETH-Zentrum, CH-8092 Zurich, Switzerland*

### SUMMARY

We consider the estimation of the parameters of a stationary random field on  $d$ -dimensional lattice by minimizing the classical Whittle approximation to the Gaussian log likelihood. If the usual biased sample covariances are used, the estimate is efficient only in one dimension. To remove this edge effect, we introduce data tapers and show that the resulting modified estimate is efficient also in two and three dimensions. This avoids the use of the unbiased sample covariances which are in general not positive-definite.

*Some key words:* Edge effect; Data taper; Sample covariance; Stationary random field; Strong mixing; Whittle function.

### 1. INTRODUCTION

Edge effects are a serious problem in spatial statistics because the number of boundary points increases with the dimension (Ripley, 1984). Consider for instance a stationary random field  $X_t$ , for  $t = (t_1, \dots, t_d) \in \mathbb{Z}^d$ , observed on a rectangle  $P_N = \{1, \dots, n_1\} \times \dots \times \{1, \dots, n_d\}$  of sample size  $N = n_1 \dots n_d$ . For calculational convenience we assume that the mean of  $X_t$  is zero. If we estimate the covariances  $\gamma(k)$  by  $c_N(k) = N^{-1} \sum X_t X_{t+k}$ , where the sum is over  $t, t+k \in P_N$ , there is a bias due to the boundary. If  $n_i \rightarrow \infty, n_i/n_1 \rightarrow \alpha_i \in (0, \infty)$  ( $i = 1, \dots, d$ ), this bias is of the order  $O(N^{-1/d})$  which is for  $d \geq 2$  of the same or a higher order as the standard deviation of  $c_N$  which is usually  $O(N^{-1/2})$ . To remedy this, Guyon (1982) proposed the use of

$$c_N^*(k) = \prod_{i=1}^d (n_i - |k_i|)^{-1} \sum_{t, t+k \in P_n} X_t X_{t+k},$$

which is always unbiased.

Similar problems occur in parametric models with spectral density  $f_\theta(\lambda)$ ,  $\theta \in \Theta$ ,  $\lambda \in \Pi^d$ , where  $\Pi = (-\pi, \pi]$ . The models we have in mind include the Gaussian Markov or autonormal models (Besag, 1974), the so-called errors in variables models (Besag, 1977) and some models coming from geostatistics (Mardia & Marshall, 1984). A common approach to the estimation of  $\theta$  is to minimize Whittle's (1954) approximation to the Gaussian log likelihood

$$L_N(\theta) = \frac{1}{2} (2\pi)^{-d} \int_{\Pi} \{\log f_\theta(\lambda) + I_N^*(\lambda) f_\theta(\lambda)^{-1}\} d\lambda, \quad (1)$$

where  $I_N^*(\lambda) = (2\pi)^{-d} \sum c_N^*(k) \exp(-i\langle \lambda, k \rangle)$ . Here too, use of  $c_N(k)$  would cause a bias which is not asymptotically negligible (Guyon, 1982, § 3.3).

However,  $c_N^*$  has some other unpleasant properties. First, it is not always positive-definite so that the spectral estimates based on  $c_N^*(k)$  may be negative. Furthermore, positive-definiteness

allows one to interpret the Whittle estimator as a minimum distance estimator and this usually guarantees the existence of a  $\theta$  minimizing (1); see Woods (1976) and Künsch (1981) for the case of Markov fields. A second disadvantage of  $c_N^*$  is its large variance, particularly for larger lags. In practical situations the use of  $c_N^*(k)$  instead of  $c_N(k)$  may make a big difference, as in the last example of Mardia & Marshall (1984).

The purpose of the present paper is to show that we can avoid these disadvantages of  $c_N^*(k)$  and nevertheless get an asymptotically negligible bias by using the data tapers introduced in nonparametric time series analysis by Tukey (1967). Dahlhaus (1984) used data tapers for the parametric analysis of time series.

## 2. THE ASYMPTOTIC BIAS OF TAPERED WHITTLE ESTIMATES

Let  $w$  be a continuous increasing function with  $w(0)=0$  and  $w(1)=1$ . We define a one-dimensional taper  $h(u)$ ,  $u \in [0, 1]$ , with smoothness parameter  $\rho$ , by

$$h(u) = \begin{cases} w(2u/\rho) & (0 \leq u < \frac{1}{2}\rho), \\ 1 & (\frac{1}{2}\rho \leq u \leq \frac{1}{2}), \\ h(1-u) & (\frac{1}{2} < u \leq 1). \end{cases} \quad (2)$$

The tapered covariance estimate is then

$$c_N^h(k) = \left\{ \prod_{i=1}^d H_{2, n_i}(0) \right\}^{-1} \sum_{t, t+k \in P_N} X_t X_{t+k} \left[ \prod_{i=1}^d h\{(t_i - \frac{1}{2})/n_i\} h\{(t_i + k_i - \frac{1}{2})/n_i\} \right],$$

where

$$H_{j, n}(\alpha) = \sum_{s=1}^n h\{(s - \frac{1}{2})/n\}^j \exp(-ias).$$

A common taper in time series analysis is the Tukey-Hanning taper with  $w(u) = \frac{1}{2}\{1 - \cos(u\pi)\}$ . Replacing  $I_N^*(\lambda)$  by  $I_N^h(\lambda) = (2\pi)^{-d} \sum c_N^h(k) \exp(-i(\lambda, k))$  in (1) leads to the tapered Whittle estimate which we denote by  $\hat{\theta}_N$ . By standard arguments (Walker, 1964) one obtains consistency of  $\hat{\theta}_N$  under some regularity conditions and hence by the usual Taylor series argument

$$\begin{aligned} \hat{\theta}_N - \theta_0 &= -\Gamma(\theta_0)^{-1} \frac{\partial}{\partial \theta} L_N(\theta_0) \{1 + o_p(1)\} \\ &= -\frac{1}{2}(2\pi)^{-d} \int_{\Pi} \{I_N^h(\lambda) - f_{\theta_0}(\lambda)\} \frac{\partial}{\partial \theta} f_{\theta_0}(\lambda)^{-1} d\lambda \{1 + o_p(1)\}, \end{aligned}$$

where

$$\Gamma(\theta)_{ij} = \frac{1}{2}(2\pi)^{-d} \int_{\Pi} \frac{\partial}{\partial \theta_i} \log f_{\theta}(\lambda) \frac{\partial}{\partial \theta_j} \log f_{\theta}(\lambda) d\lambda. \quad (3)$$

We study in this section the expectation of

$$J_N(\phi) = \int_{\Pi} \{I_N^h(\lambda) - f_{\theta_0}(\lambda)\} \phi(\lambda) d\lambda,$$

where  $\phi$  is a continuous function. This gives the asymptotic bias of  $c_N^h(k)$  and typically also of  $\hat{\theta}_N$ .

For the taper we need the following assumption.

*Assumption 1.* The function  $w$  is differentiable on  $[0, 1]$  with a Lipschitz-continuous derivative.

THEOREM 1. Let  $X_t$ , for  $t \in Z^d$ , be a mean zero stationary random field with twice differentiable spectrum  $f$ . Assume the taper  $h(u)$  satisfies Assumption 1. Then

$$E\{J_N(\phi)\} = \frac{1}{2} N^{-2/d} \left( \prod_{k=1}^d \alpha_k \right)^{2/d} \left\{ \int_0^1 h'(x)^2 dx \Big/ \int_0^1 h(x)^2 dx \right\} \\ \times \left\{ \sum_{k=1}^d \alpha_k^{-2} \int_{\mathbb{I}} \phi(\beta) \frac{\partial^2}{(\partial \beta_k)^2} f(\beta) d\beta \right\} \{1 + o(1)\}$$

if  $n_i \rightarrow \infty$  and  $n_i/n_1 \rightarrow \alpha_i \in (0, \infty)$  ( $i = 1, \dots, d$ ). Here  $\rho$  may depend on  $N$  and the  $o(1)$  term is uniform in  $\rho$ , where  $\rho^{-1} = o(N^{1/2})$ .

Proof. Straightforward calculations yield

$$E(J_N(\phi)) = \int_{\mathbb{I}} \int_{\mathbb{I}} \phi(\beta) \{f(\beta + \alpha) - f(\beta)\} K_N(\alpha) d\alpha d\beta,$$

with

$$K_N(\alpha) = \prod_{k=1}^d \{2\pi H_{2,n_k}(0)\}^{-1} |H_{1,n_k}(\alpha_k)|^2.$$

Furthermore, since  $f$  is twice differentiable we have

$$f(\beta + \alpha) - f(\beta) = \sum_{k=1}^d \alpha_k \frac{\partial}{\partial \beta_k} f(\beta) + 2 \sum_{k,l=1}^d \sin(\frac{1}{2}\alpha_k) \sin(\frac{1}{2}\alpha_l) \frac{\partial^2}{\partial \beta_k \partial \beta_l} f(\beta) + R(\beta, \alpha),$$

with

$$|R(\beta, \alpha)| = o\left\{ \sum_{k=1}^d \sin^2(\frac{1}{2}\alpha_k) \right\}.$$

Since  $K_N(\alpha) = K_N(-\alpha)$ ,  $\int \alpha_k K_N(\alpha) d\alpha = 0$  and  $\int \sin(\frac{1}{2}\alpha_k) \sin(\frac{1}{2}\alpha_l) K_N(\alpha) d\alpha = 0$  for  $k \neq l$ . Summation by parts gives, with  $h(u) = 0$  for  $u \notin [0, 1]$ ,

$$H_{1,n_k}(\alpha_k) = \{\exp(i\alpha_k) - 1\}^{-1} \sum_{s=0}^{n_k} D \left\{ h\left(\frac{s - \frac{1}{2}}{n_k}\right) \right\} \exp(-i\alpha_k s),$$

where

$$D \left\{ h\left(\frac{s - \frac{1}{2}}{n_k}\right) \right\} = h\left(\frac{s + \frac{1}{2}}{n_k}\right) - h\left(\frac{s - \frac{1}{2}}{n_k}\right).$$

Hence,

$$\int \sin^2(\frac{1}{2}\alpha_k) K_N(\alpha) d\alpha = H_{2,n_k}(0)^{-1} \sum_{s=0}^{n_k} D \left\{ h\left(\frac{s - \frac{1}{2}}{n_k}\right) \right\}^2. \tag{4}$$

Since  $h(u)$  is differentiable for  $u \notin D_\rho = \{0, \frac{1}{2}\rho, 1 - \frac{1}{2}\rho, 1\}$  we obtain for  $s$  with  $[(s - \frac{1}{2})/n, (s + \frac{1}{2})/n] \cap D_\rho = \emptyset$ , by the mean value theorem,

$$D \left\{ h\left(\frac{s - \frac{1}{2}}{n}\right) \right\}^2 = n^{-1} \int h'(x)^2 dx + n^{-1} \int \{h'(\bar{x}_s)^2 - h'(x)^2\} dx,$$

where the integrals are over  $((s - \frac{1}{2})/n, (s + \frac{1}{2})/n)$ , and with  $(s - \frac{1}{2})/n \leq \bar{x}_s \leq (s + \frac{1}{2})/n$ . The result follows by a straightforward estimation of the second term, and by considering the four remaining terms in (4) separately.  $\square$

The result remains valid if we drop the assumption that the mean is known. We then subtract the tapered sample mean

$$\hat{\mu}_N = \left\{ \prod_{i=1}^d H_{1,n_i}(0) \right\}^{-1} \sum_{t \in P_n} \left[ \prod_{i=1}^d h\{(t_i - \frac{1}{2})/n_i\} \right] X_t$$

from the data.

We thus have shown that for  $d = 1, 2, 3$  the bias for the tapered estimates is of smaller order than  $N^{-1}$ . For fixed  $w$  we obtain

$$\int_0^1 \{h'(x)\}^2 dx / \int_0^1 \{h(x)\}^2 dx = \frac{4}{\rho} \int_0^1 \{w'(x)\}^2 dx / \left[ 1 + \rho \left\{ \int_0^1 \{w(x)\}^2 dx - 1 \right\} \right],$$

and the bias is therefore smallest if we choose for  $\rho$  the maximal value 1. However, asymptotically a loss of efficiency occurs if  $\rho > 0$ ; see § 3. To avoid this we let  $\rho$  depend on  $N$  with  $\rho_N \rightarrow 0$ . By Theorem 1 this can be done for  $d = 1, 2, 3$  without getting a bias of the order  $O(N^{-1})$ .

Another advantage of tapering is the reduction of the bias if some nonstationarities are present at the boundary, for instance if boundary plots in a field trial behave differently. In the literature several variants of Markov fields have been proposed which treat boundary observations differently (Ripley, 1984). This causes a severe bias for  $c_N^*(k)$ , whereas  $c_N^h(k)$  will have a small bias no matter what the covariances at the boundary look like.

### 3. EFFICIENCY OF TAPERED WHITTLE ESTIMATES

In this section we prove a central limit theorem for the estimate  $\hat{\theta}_N$  and study its efficiency. As usual we need assumptions on the parameter space, Assumption 2, e.g. identifiability conditions, assumptions on the observed random fields, Assumption 3, and assumptions on the taper, Assumption 1, which have already been stated in § 2.

*Assumption 2.* Let  $\Theta$  be a compact subset of  $\mathbb{R}^p$ , and  $X_{t,\theta}$  a family of stationary random fields with spectral densities  $f_\theta$ . Suppose  $(f_\theta)$  is bounded uniformly above and away from zero, that  $f_\theta^{-1}$  has a continuous second derivative, with respect to  $\theta$  which is continuous on  $\theta \times \pi^d$  and  $\theta_1 \neq \theta_2$  implies  $f_{\theta_1} \neq f_{\theta_2}$  on a set of positive Lebesgue measure. Suppose  $\theta_0$  is true, and  $\hat{\theta}_N$  exists in the interior of  $\Theta$ .

*Assumption 3.* We have that  $X_{t,\theta_0}$ ,  $t \in Z^d$  is a strongly mixing stationary random field with mean zero,  $E(|X_t|^{4+2\delta}) < \infty$  for some  $\delta > 0$ , and mixing coefficients  $\alpha_{k,l}(m)$  (Bolthausen, 1982) with, for all  $k \in \mathbb{N}$ ,

$$\sum_{m=1}^{\infty} m^{d-1} \alpha_{k,\infty}(m)^{\delta/(2+\delta)} < \infty, \quad \sum_{m=1}^{\infty} m^{3d} \{\alpha_{2,2}(m) + \alpha_{1,3}(m)\}^{\delta/(2+\delta)} < \infty.$$

In addition to the matrix  $\Gamma(\theta)$ , defined in (3), we need

$$B(\theta) = \frac{1}{4}(2\pi)^{-d} \int_{\mathbb{1}^{2d}} \frac{f_{4,\theta}(\alpha_1, -\alpha_1, \alpha_2)}{f_\theta(\alpha_1)f_\theta(\alpha_2)} \frac{\partial}{\partial \theta_i} \log f_\theta(\alpha_1) \frac{\partial}{\partial \theta_j} \log f_\theta(\alpha_2) d\alpha_1 d\alpha_2,$$

where  $f_{4,\theta}$  is the fourth-order cumulant spectrum of the process  $X_{t,\theta}$ .

**THEOREM 2.** *Suppose Assumptions 1, 2 and 3 hold where the smoothness parameter  $\rho$  depends on  $N$  with  $\rho_N \rightarrow \rho_0$  and  $\rho_N^{-1} = o(N^{2/d-1})$ . Then  $N^{1/2}(\hat{\theta}_N - \theta_0)$  converges weakly to*

$$\mathcal{N}(0, H\Gamma(\theta_0)^{-1} \{ \Gamma(\theta_0) + B(\theta_0) \} \Gamma(\theta_0)^{-1}),$$

where  $H = \int \{h_{\rho_0}(x)\}^4 dx / [\int \{h_{\rho_0}(x)\}^2 dx]^2$ .

*Proof.* The technique of proof which was sketched in § 2 is part of time series folklore, provided we have a central limit theorem for  $J_N(\phi)$  with  $\phi(\lambda) = \partial f_{\theta_0}(\lambda)^{-1} / \partial \theta$ . This is stated in the Appendix. □

**COROLLARY.** *Suppose in addition that  $X_{t,\theta_0}$  is Gaussian and  $\rho_N \rightarrow 0$ . Then  $\hat{\theta}_N$  is asymptotically Fisher-efficient.*

*Proof.* Since  $H = 1$  in the case  $\rho_0 = 0$ , the result follows from Theorem 4.2 and Proposition 1 of Guyon (1982), analogously to Proposition 4 of Guyon.  $\square$

Due to the requirement  $\rho_N^{-1} = o(N^{2/d-1})$  we can obtain efficient estimates in the cases  $d = 1, 2, 3$ , while in the nontapered case we only have efficiency if  $d = 1$  (Guyon, 1982, § 3.3).

The theoretical results of this paper were confirmed in a simulation study where a conditional autoregression model was fitted to the superposition of two plane waves in white noise. The use of the tapered covariances  $c_N^h$  led to a better resolution of the smaller peak than the use of the covariances  $c_N$ . The unbiased covariances  $c_N^*$  were often not positive-definite, and as a consequence the corresponding Whittle function did not always have a minimum, while for  $c_N$  and  $c_N^h$  the existence of a minimum is guaranteed automatically.

The results mean that in two and three dimensions our tapered estimate is an attractive compromise between the classical biased and the unbiased versions avoiding their bad properties and retaining many of the good ones.

APPENDIX

A central limit theorem

We state a central limit theorem for the statistic  $J_N(\phi)$  needed for the proof of Theorem 2.

**THEOREM A1.** *Suppose that  $X_t$ , for  $t \in Z^d$ , fulfils Assumption 3, Assumption 1 holds where the smoothness parameter  $\rho$  depends on  $N$  with  $\rho_N \rightarrow \rho_0$  and  $\rho_N^{-1} = o(N^{2/d-1})$ , and  $\phi_j$  ( $j = 1, \dots, k$ ) are continuous functions. Then  $\{N^{1/d} J_N(\phi_j)\}$  for  $j = 1, \dots, k$  converges weakly to a Gaussian random vector  $\{Y_j\}$  with mean zero and*

$$\begin{aligned} \text{cov}(Y_j, Y_k) &= (2\pi)^d \left( \int_0^1 \{h_{\rho_0}(x)\}^4 dx / \left[ \int_0^1 \{h_{\rho_0}(x)\}^2 dx \right]^2 \right)^d \\ &\quad \times \left[ \int_{\mathbb{R}^{2d}} \phi_i(\alpha_1) \phi_j(-\alpha_2) f_4(\alpha_1, -\alpha_1, \alpha_2) d\alpha_1 d\alpha_2 \right. \\ &\quad \left. + \int_{\mathbb{R}^d} \phi_1(\alpha) \{ \phi_2(\alpha) + \phi_2(-\alpha) \} f(\alpha)^2 d\alpha \right], \end{aligned}$$

where  $f_4$  is the fourth-order spectrum of the process.

*Proof.* The proof consists of the following two steps. By generalizing the arguments of Bolthausen (1982) we first prove a central limit theorem for the empirical covariances  $c_N^h(k)$ , and by using the Cesàro sums of the  $\phi_j$  we then conclude from this to the weak convergence of  $\{N^{1/d} J_N(\phi_j)\}$  for  $j = 1, \dots, k$ . Details may be obtained from the authors upon request.  $\square$

Alternatively, the result may also be proved by a cumulant method under the assumptions that all moments of the process exist and the cumulant spectra fulfill certain smoothness conditions, for example by using a modified version of Theorem 3.1 of Brillinger (1970).

REFERENCES

BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc. B* **36**, 192-236.  
 BESAG, J. (1977). Errors-in-variables estimation for Gaussian lattice schemes. *J. R. Statist. Soc. B* **39**, 73-8.  
 BOLTHAUSEN, E. (1982). On the central limit theorem for stationary mixing random fields. *Ann. Prob.* **10**, 1047-50.  
 BRILLINGER, D. R. (1970). The frequency analysis of relations between stationary spatial series. In *Proc. 12th Biennial Seminar on Time Series, Stochastic Processes, Convexity, and Combinatorics*, Ed. R. Pyke, pp. 39-81. Montreal: Canadian Math. Congress.  
 DAHLHAUS, R. (1984). Parameter estimation of stationary processes with spectra containing strong peaks. In *Robust and Nonlinear Time Series Analysis*, Lecture Notes in Statistics, 26, Ed. J. Franke, W. Hardle and R. D. Martin, pp. 50-68. New York: Springer.

- GUYON, X. (1982). Parameter estimation for a stationary process on an  $d$ -dimensional lattice. *Biometrika* **69**, 95-105.
- KÜNSCH, H. (1981). Thermodynamics and statistical analysis of Gaussian random fields. *Z. Wahr. verw. Geb.* **58**, 407-21.
- MARDIA, K. V. & MARSHALL, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**, 135-46.
- RIPLEY, B. D. (1984). Spatial statistics: Developments 1980-3. *Int. Statist. Rev.* **52**, 141-50.
- TUKEY, J. W. (1967). An introduction to the calculations of numerical spectrum analysis. In *Advanced Seminar on Spectral Analysis of Time Series*, Ed. B. Harris, pp. 25-46. New York; Wiley.
- WALKER, A. M. (1964). Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time-series. *J. Aust. Math. Soc.* **4**, 363-84.
- WHITTLE, P. (1954). On stationary processes in the plane. *Biometrika* **41**, 450-62.
- WOODS, J. W. (1976). Two-dimensional Markov spectral estimation. *IEEE Transf. Info. Theory* **22**, 552-9.

[Received July, 1986. Revised December 1986]