Université de Montréal

Sur la modélisation et l'estimation de la fonction de covariance d'un processus aléatoire

par

Miro Powojowski

Département de mathématiques et de statistique
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en Statistique

Décembre 1999

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée

Sur la modélisation et l'estimation de la fonction de covariance d'un processus aléatoire

présentée par

Miro Powojowski

a été évaluée par un jury composé des personnes suivantes :

Roch Roy
(président-rapporteur)

Christian Léger
(directeur de recherche)

Denis Marcotte
(co-directeur)

Marc Moore
(co-directeur)

Brenda MacGibbon
(membre du jury)

Noel Cressie
(examinateur externe)

Roch Roy
(représentant du doyen)

Thèse acceptée le :
28 juin 2000

SOMMAIRE

Cette thèse considère le problème de la modélisation et l'estimation de la fonction de covariance d'un processus aléatoire. Le travail est divisé en trois articles. Le premier article introduit le modèle additif et l'estimateur par projection. Les propriétés de l'estimateur par projection sont obtenues dans des contextes asymptotiques. Différentes généralisations de l'estimateur sont aussi considérées. Le deuxième article présente une classe importante de modèles additifs de la fonction de covariance d'un processus isotrope, la classe de modèles additifs à composantes spectrales. Les résultats théoriques du premier article permettent d'établir les propriétés de l'estimateur par projection appliqué aux modèles additifs à composantes spectrales. On démontre aussi que la classe des modèles additifs à composantes spectrales est dense dans l'ensemble des fonctions de covariance des processus isotropes ayant une densité spectrale. Le troisième article propose des méthodes de sélection optimale du modèle pour des modèles additifs à composantes spectrales. Des critères pour choisir le nombre et la forme des composantes spectrales sont obtenus et évalués par simulation. Enfin, un jeu de données réelles est analysé par les méthodes développées dans les trois articles.

REMERCIEMENTS

L'auteur remercie le directeur de recherche, Monsieur Christian Léger, ainsi que les codirecteurs, Messieurs Denis Marcotte et Marc Moore, pour leur appui et leurs conseils. L'auteur remercie également les trois Conseils (CRSNG, CRM et CRSH) pour le financement de cette recherche par leur programme conjoint de bourses.

Table des matières

Table des figures

Liste des tableaux

INTRODUCTION

Cette thèse traite de l'estimation du covariogramme d'un processus aléatoire. Pour un processus aléatoire $Y(x), x \in D$, où D est un sous-ensemble de l'espace euclidien de dimension d, le covariogramme est une fonction définie par

$$C_Y(x_1, x_2) = \text{cov}(Y(x_1), Y(x_2)). \tag{0.0.1}$$

Un problème typique en géostatistique est celui de l'estimation de la fonction C_Y basée sur une réalisation du processus Y observé sur un nombre fini d'emplacements $x_1, \ldots, x_n$ dans le domaine D. Un sommaire des méthodes traditionnelles de l'estimation du covariogramme d'un processus isotrope Y apparaît dans Cressie (1993). La procédure standard peut être décrite comme suit. Premièrement, le variogramme ou le covariogramme empirique est estimé. Pour ce faire, les paires d'observations sont divisées en classes de distance. Cette étape implique des décisions concernant le nombre et la longueur des intervalles de distance. Ensuite un modèle paramétrique est choisi parmi un nombre relativement petit de fonctions positives définies connues. Finalement, les paramètres du modèle sont estimés par une méthode d'ajustement standard qui cherche à minimiser l'écart entre le modèle et le covariogramme ou le variogramme empirique. Dans la plupart des cas, la méthode d'ajustement est une variante de la minimisation de la somme des différences au carré entre le variogramme ou le covariogramme empirique et les valeurs données par le modèle. Dans presque tous les cas, cette procédure de

minimisation implique une fonction objective non linéaire. Pour effectuer la minimisation, il est nécessaire d'appliquer des méthodes numériques qui compliquent le calcul et surtout rendent l'analyse théorique de l'estimateur difficile. Une autre approche de l'estimation du covariogramme est la méthode MINQUE (minimum norm quadratic unbiased estimator). On définit un modèle de covariance additif par

$$C_\theta = \sum_{i=1}^{q} \theta(i) C_i \tag{0.0.2}$$

où les fonctions C_i sont déterminées et connues, tandis que les paramètres $\theta(i)$ sont à estimer. Le modèle C_θ dépend des paramètres $\theta(i)$ de façon linéaire. La méthode MINQUE s'applique seulement aux modèles additifs. Cependant, elle offre certains avantages par rapport aux méthodes traditionnelles. Le besoin de choisir le nombre et la longueur des intervalles de distance est éliminé. Les méthodes d'algèbre matricielle sont suffisantes pour calculer l'estimateur. Des expressions relativement simples existent pour la moyenne et la variance de l'estimateur. En plus, certains résultats asymptotiques ont été établis pour l'estimateur MINQUE par Stein (1989). D'autre part, l'estimateur MINQUE présente des désavantages considérables. Le calcul de l'estimateur, bien qu'utilisant seulement l'agèbre matricielle, exige le calcul des inverses et des décompositions spectrales de grandes matrices ($n \times n$, où n est le nombre d'observations). Les jeux de données contenant des centaines ou des milliers d'observations sont commun en géostatistique. Le fardeau computationel est considérable et les difficultés liées à l'instabilité numérique peuvent se faire sentir avec des matrices de cette taille. La méthode MINQUE comporte un problème plus fondamental. Elle exige un point de départ (un estimé préalable) C_0 de la fonction C_Y à estimer. Les résultats théoriques

obtenus par Stein (1989), ainsi que les résultats généraux concernant la méthode MINQUE (dont un sommaire se trouve dans Rao et Kleffe, 1988) exigent des hypothèses voulant que C_0 et C_Y soient "proches" dans un certain sens. Ces hypothèses ne sont pas vérifiables en pratique. Par contre, une violation relativement peu sévère de ces hypothèses peut produire une instabilité importante de l'estimateur (Powojowski, non publié).

Les difficultés de mise en oeuvre et de l'analyse théorique de l'estimateur MINQUE ont mené au développement de l'estimateur par projection que j'introduis dans cette thèse. L'estimateur par projection emprunte tant aux approches traditionnelles qu'à la méthode MINQUE. Tout comme dans l'approche traditionnelle, l'estimateur par projection résulte de la minimisation d'une certaine distance entre certains moments expérimentaux et ceux du modèle. Tout comme l'estimateur MINQUE, l'estimateur par projection s'applique aux modèles additifs. Bien que la dérivation de l'estimateur par projection, plutôt naturelle, n'exige pas une familiarité avec la méthode MINQUE, on peut démontrer que l'estimateur par projection peut être considéré comme un cas particulier de l'estimateur MINQUE, avec un choix particulier de C_0. Ce choix de C_0 n'est généralement pas "proche" de C_Y et donc les résultats asymptotiques de Stein (1989) et Rao et Kleffe (1988) ne s'appliquent pas.

La thèse est divisée en trois articles. Le premier article traite des résultats généraux pertinents à la méthode d'estimation par projection qui est présentée pour la première fois dans cette thèse. Les méthodes standards de projections orthogonales dans un espace vectoriel muni d'un produit scalaire sont utilisées pour obtenir l'estimateur par projection $\hat{\theta}_n$ (où n est le nombre d'observations). Par définition, l'estimateur est sans biais pour le paramètre θ si le vrai modèle

C_Y est de la forme

$$C_Y = \sum_{i=1}^{q} \theta(i) C_i. \tag{0.0.3}$$

Différentes généralisations de l'estimateur sont considérées. Il est ensuite démontré que l'estimateur par projection possède certaines propriétés utiles. Considérons un processus Y qui aurait la forme

$$Y = X\beta + \eta \tag{0.0.4}$$

où X est une matrice connue, β est un vecteur inconnu et η est un processus dont la moyenne est zéro. L'estimateur demeure sans biais même si le paramètre β de la moyenne du processus Y n'est pas connu et doit être estimé. Ceci est un avantage par rapport aux méthodes traditionnelles, où la nécessité d'estimer la moyenne du processus introduit un biais dans l'estimation du covariogramme empirique (Cressie, 1993). Le calcul du variogramme ou covariogramme empirique n'est pas nécessaire. En plus, le calcul de l'estimateur implique seulement des sommes et produits, ainsi que le calcul des inverses de petites matrices ($q \times q$, où q est le nombre de composantes dans le modèle). Une expression simple existe pour la moyenne de l'estimateur. Si le processus est gaussien, une expression simple existe également pour la variance. Pour des processuus non gaussiens, une hypothèse supplémentaire produit une borne supérieure pour la variance de l'estimateur. De plus, l'estimateur par projection n'exige pas d'hypothèse de stationnarité de Y ni celle de $Y - \mathbf{E}[Y]$ et donc peut être défini et calculé pour un processus non stationnaire.

Les propriétés asymptotiques de l'estimateur par projection sont ensuite explorées dans le contexte asymptotique où les observations sont placées dans un domaine compact avec une densite croissante (in-fill asymptotics). Dans ce cas, le processus Y est observé à une séquence infinie de points $\{x_j\}_{j=1}^{\infty}$ comprise dans

un domaine compact D. On démontre qu'avec des hypothèses de régularité plutôt faibles, lorsque le nombre d'observations n augmente, la moyenne de l'estimateur par projection converge vers une expression simple impliquant la vraie fonction de covariance C_Y. De plus, on obtient une borne supérieure $E(D)$ pour la variance de l'estimateur. Ces résultats ont des conséquences pratiques car ils peuvent guider le choix efficace du nombre d'observations dans un domaine compact donné. Il est démontré que la moyenne du covariogramme estimé par l'estimateur par projection est toujours le plus "proche" modèle dans sa classe en fonction de la distance utilisée. On démontre aussi qu'il est généralement impossible d'estimer le covariogramme de façon convergente à partir des observations comprises dans un domaine compact.

Dans la dernière partie du premier article, l'attention est concentrée sur un processus de la forme (0.0.4) où η est un processus isotrope (et donc stationnaire). On démontre que la taille du domaine D influence la borne $E(D)$ de la variance de l'estimateur par projection et que des hypothèses de régularité relativement faibles sont suffisantes pour obtenir une convergence de $E(D)$ vers zéro lorsque le domaine D grandit indéfiniment. Cela ouvre la possibilité d'une estimation convergente du covariogramme à partir d'une séquence infinie d'observations qui couvrent un domaine croissant avec une densité qui augmente. Le cas particulier de l'échantillonnage uniforme d'un domaine croissant indéfiniment est considéré en détail. On démontre que si le nombre d'observations augmente avec la taille du domaine à un certain taux minimal, l'estimateur par projection converge en probabilité vers la vraie valeur du paramètre θ.

Les résultats du premier article s'appliquent aux modèles additifs seulement. Cependant, la plupart des modèles utilisés en géostatistique ne dépendent pas de leurs paramètres de façon linéaire. L'utilité de l'estimateur par projection va donc

dépendre de la disponibilité de modèles additifs adéquats. Le deuxième article fournit une vaste classe de modèles additifs pour un processus de la forme (0.0.4) où η est isotrope. Pour différencier entre les fonctions de covariance $C_Y(x_1, x_2)$, $C_\theta(x_1, x_2)$, $C_i(x_1, x_2)$ et leur versions isotropes, ces dernières seront dénotées par $\mathbf{C}_Y(\rho)$, $\mathbf{C}_\theta(\rho)$ et $\mathbf{C}_i(\rho)$ où $\mathbf{C}_Y(\rho) = C_Y(x_1, x_2)$, $\mathbf{C}_Y(\theta) = C_\theta(x_1, x_2)$ et $\mathbf{C}_i(\rho) = C_i(x_1, x_2)$ et où $\rho = \| x_1 - x_2 \|$. La classe proposée est basée sur la représentation spectrale de la fonction de covariance d'un processus isotrope. Le modèle s'écrit comme suit :

$$\mathbf{C}_\theta(\rho) = \sum_{i=1}^{q} \theta(i)\mathbf{C}_i(\rho) \tag{0.0.5}$$

où

$$\mathbf{C}_i(\rho) = \rho^{(2-d)/2} \int_{a_i}^{b_i} \lambda J_{(d-2)/2}(\lambda\rho)\Psi(\lambda)d\lambda, \tag{0.0.6}$$

et Ψ est une fonction non négative intégrable quelconque, les intervalles $[a_i, b_i]$ sont finis et disjoints et J_ν est la fonction de Bessel d'ordre ν de la première espèce. Les modèles de ce type seront nommés modèles additifs à composantes spectrales. Ces modèles sont inspirés de ceux introduits par Shapiro et Botha (1991). On démontre que les modèles proposés satisfont aux hypothèses de régularité nécessaires pour obtenir les résultats asymptotiques du premier article, contrairement à ceux de Shapiro et Botha (1991). Ensuite on démontre que la classe des modèles additifs à composantes spectrales est dense dans l'ensemble des fonctions de covariance isotropes ayant une densité spectrale. Il s'ensuit que pour n'importe quel modèle $\mathbf{C}_Y$ ayant une densité spectrale, il existe un modèle additif à composantes spectrales qui produit un estimé $\mathbf{C}_{\hat{\theta}}$ (où $\hat{\theta}$ est trouvé par l'estimateur par projection) dont le biais est arbitrairement petit. Cela se compare favorablement avec l'approche paramétrique, où un mauvais choix de modèle

peut introduire un biais important. On démontre que dans le cas de modèles additifs à composantes spectrales, les expressions pour la moyenne et la variance de l'estimateur par projection prennent des formes très particulières, révélant ainsi une connection avec la théorie spectrale. On démontre que l'estimateur par projection peut être utilisé pour estimer la densité spectrale et les expressions pour sa variance fournissent des expressions pour la variance de $\hat{\Psi}_Y = \Psi_{\hat{\theta}}$. Il semble que ce soit la seule méthode qui permette d'estimer la densité spectrale à partir de données qui ne proviennent pas d'une grille régulière. On démontre ensuite que les corrélations asymptotiques des composantes $\hat{\theta}_n(i)$ de l'estimateur égalent zéro si le processus est gaussien. Une autre connection utile avec la théorie spectrale est une expression équivalente pour la norme dans l'espace des fonctions de covariance en terme des densités spectrales comme ceci :

$$\int_0^\infty (\mathbf{C}_Y(\rho) - \mathbf{C}_\theta(\rho))^2 \rho^{d-1} d\rho = \int_0^\infty (\Psi_Y(\lambda) - \Psi_\theta(\lambda))^2 \lambda d\lambda \qquad (0.0.7)$$

Cette relation sera essentielle au traitement des questions liées à la sélection optimale du modèle. Le restant du deuxième article a pour but d'illustrer la forme des composantes spectrales et le biais très faible qui résulte de l'ajustement d'un modèle additif à composantes spectrales dans les cas où le vrai modèle est un modèle paramétrique tel que gaussien, exponentiel, sphérique ou "hole-effect". Le même modèle additif est ajusté à des vrais modèles très différents. Dans tous les cas, le biais introduit par l'application du modèle additif plutôt que du vrai modèle est négligeable.

Afin de construire un modèle additif à composantes spectrales, il faut choisir l'ordre q du modèle, ainsi que les extrémités des intervalles de support $[a_i, b_i], i = 1, \ldots, q$. Le troisième article traite des méthodes de sélection du modèle. Une méthode qui permet l'estimation approximative du support spectral est développée. Elle est elle-même basée sur un estimateur par projection. Le support de

la densité spectrale Ψ_Y peut être considéré comme l'ensemble maximal qui doit être couvert par les intervalles $[a_i, b_i], i = 1, \ldots, q$. La méthode est testée par simulation dans deux études où elle s'avère adéquate pour l'estimation du support spectral de Ψ_Y. Des critères pour choisir les intervalles $[a_i, b_i]$ lorsque l'ordre q et le support spectral ont déjà été choisis sont considérés dans une section subséquente. Finalement, un critère est développé pour permettre la comparaison de deux modèles. Le critère est conçu pour minimiser l'erreur quadratique moyenne (EQM) de l'estimateur par projection et il vise un équilibre entre des modèles avec un nombre élevé de paramètres, un biais faible et une variance plus importante et des modèles avec un faible nombre de paramètres, un biais plus important mais une plus faible variance. Le critère est étudié par simulation et il s'avère adéquat. Les méthodes de sélection du modèle sont ensuite appliquées afin de construire un modèle pour le jeu de données de Davis (1973). L'estimateur par projection est employé pour estimer les paramètres du modèle. L'estimé $\mathbf{C}_{\hat{\theta}}$ obtenu est comparé avec les résultats obtenus par d'autres chercheurs.

L'ordre des articles reflète leur emplacement dans le développement de la méthodologie présentée. Le premier article présente des résultats généraux. Le deuxième propose une classe de modèles flexible. Lorsque les résultats du premier article sont appliqués à cette classe de modèles, les représentations spectrales sont obtenues. Elles sont essentielles pour le développement des critères de sélection du modèle, présentés dans le troisième article.

Chapitre 1

ADDITIVE COVARIOGRAM MODELS AND ESTIMATION THROUGH PROJECTIONS

1.1. Abstract

The paper considers the problem of estimating the covariogram of a stationary process. The main ideas explored are additive covariance models and their estimation in terms of projections in the inner product space of sufficiently regular functions. Asymptotic properties of the resulting estimators are worked out, without explicit assumptions about the functional form of model components or that of the true covariogram. Expressions for bias of the estimator in misspecified models, expressions for the estimator's variance in the normal case and bounds for variance of the estimator under relaxed assumptions are derived. It is demonstrated through asymptotic analysis that the inclusion of drift and nugget effect does not significantly affect the estimator's performance. Both in-fill asymptotics and expanding-domain asymptotics are considered. The approach can also be applied to the estimation of non-stationary covariance structures and the in-fill asymptotic results hold. The techniques are applied to a data set of Davis (1973).

1.2. INTRODUCTION

For a random process $Y(x), x \in D$, where D is a subset of a d-dimensional Euclidean space, the covariogram is defined as $C(x_1, x_2) = \mathrm{cov}(Y(x_1), Y(x_2))$, the semivariogram is defined as $\gamma(x_1, x_2) = (1/2)\,\mathrm{var}(Y(x_1) - Y(x_2))$, and the variogram is defined as 2γ. These definitions do not require the process to be stationary. For a second-order stationary process, the two are related through $\gamma(x_1, x_2) = C(0, 0) - C(x_1, x_2)$ (Cressie, 1993). A common problem in geostatistics is one of estimating the functions C and γ based on one realisation of the process Y observed at a finite number of locations $x_1, x_2, \ldots, x_n$ in D. It is important to note that the knowledge of function values $C(x_1, x_2)$ for arbitrary $(x_1, x_2) \in D^2$ is required, and not simply the covariances of Y at lags observed in the sample. The fact of observing only one realisation forces one to make certain assumptions about the process Y, which translate into restrictions on the form of C and γ. There also exist theoretical reasons for restricting the function families considered. The covariogram has to be a positive definite function, whereas the variogram has to be conditionally negative definite. Further restrictions may be desirable. The process Y may be assumed second-order stationary, or even isotropic, requiring $C(x_1, x_2)$ and $\gamma(x_1, x_2)$ to depend only on $x_1 - x_2$ or its length, respectively. In a typical covariogram estimation problem it is supposed that the observed process Y follows the model

$$Y = X\beta + \eta.$$

The known regressor X usually contains terms corresponding to the mean of the process and any trend that is modelled, while the parameter β is unknown and the random term η is assumed to have zero mean and a covariogram C_Y. The

covariogram C_Y is modelled by a covariance function C_θ, known up to the value of a finite-dimensional vector θ, to be estimated.

The paper is organised as follows. Firstly the notation is introduced and some notions of projections in inner product spaces are summarised. The projection estimator and its extensions are then defined. Subsequently, the asymptotic in-fill setting is defined, where observations of the process Y are collected on a finite domain. In this setting, the projection estimator's asymptotic properties are investigated. It is shown that in general it is impossible to estimate the covariogram consistently based on observations from a finite domain, but an upper bound for the asymptotic variance of the projector estimator is obtained. Subsequently, it will be shown that as the size of the domain increases indefinitely, the upper bound for the variance of the estimator derived earlier converges to zero. Finally, an application of the projection estimator is illustrated with a data set of Davis (1973).

It should be noted that Sections 1.4 - 1.5.2 do not assume that η is a second-order stationary process. Thus the estimator can be computed for more general processes and its in-fill asymptotics are unaffected by non-stationarity.

Most technical proofs are given in Appendix A.

1.3. Traditional approaches

A rather exhaustive discussion of the traditional methods of covariogram and variogram estimation is contained in Cressie(1993). Two broad classes of methods can be distinguished: methods requiring parametric distributional assumptions concerning the underlying process, such as ML or REML methods, and methods which avoid making such precise parametric hypotheses. Among the methods of the latter category reviewed by Cressie, all but the MINQUE method involve the

computation of the so-called empirical variogram or covariogram. The empirical covariogram can be meaningfully computed only for second-order stationary processes, while the empirical variogram can be meaningfully computed only for intrinsically stationary processes. For a second-order stationary process, the empirical semivariogram is usually defined as the quantity $\hat{\gamma}(h)$

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{N(h)} (Y(x_i) - Y(x_j))^2, \tag{1.3.1}$$

where

$$N(h) = |\{(x_i, x_j) : x_i - x_j = h; i, j = 1, \ldots, n\}|,$$

while the empirical covariogram is defined as

$$\hat{C}(h) = \frac{1}{N(h)} \sum_{N(h)} (Y(x_i) - \hat{\mu})(Y(x_j) - \hat{\mu}), \tag{1.3.2}$$

where $\hat{\mu}$ is some estimator of the mean of the process. In practice, the observations are usually not regularly spaced and the $n(n-1)/2$ lags h are binned in order to obtain a sufficient number of observed lags falling into each bin. Since the ultimate goal of the (co)variogram estimation usually involves modelling variances at unobserved lags, a parametric curve from some valid family of (co)variogram functions is then fitted to the empirical (co)variogram. This last step may be performed in a number of ways. For example, ordinary or weighted least squares fitting may be applied. Usually this step involves optimisation of non-linear and non-quadratic functions. There appears to be a preference among practitioners for estimating the variogram rather than the covariogram. This is usually justified by the fact that the empirical variogram can be meaningfully defined for a broader class of stochastic processes (intrinsically stationary stochastic processes) than the empirical covariogram (which requires the process to be second-order

stationary) and that the empirical variogram is unaffected by a constant non-zero mean in Y, while the empirical covariogram is a biased estimate in this case. It is also known that the empirical variogram is less sensitive to a linear trend contamination (Cressie, 1993). It will be seen that the covariogram estimation through projections eliminates the problems associated with the empirical covariogram. The estimate makes sense even for non-stationary processes and for a stationary process with any polynomial trend (of known order) the covariogram can be estimated without bias.

Covariance function estimation based on empirical (co)variogram estimation suffers from a number of drawbacks. Binning the observations introduces an element of arbitrariness and is sometimes difficult to perform if the number of observations is low or the process is not isotropic. The empirical covariogram is meaningless if the observed process is non-stationary and the empirical variogram is sensitive to departures from intrinsic stationarity (Cressie, 1993). Finally, the fitting procedure is usually difficult to assess from a statistical point of view. Most known theoretical results (whose comprehensive summary may be found in Cressie, 1993), are concerned only with the properties of the empirical (co)variogram and not those of the fitted (co)variogram function. It appears that the problem of obtaining the properties of the fitted (co)variogram function from the empirical (co)variogram has not been extensively studied. In contrast, the projection-based estimation yields the mean and variance expressions for the parameters of the estimated covariance curve.

1.4. Covariogram estimation through projections

This section describes the notation, reviews standard notions of inner product spaces and introduces the model and the estimator considered in the remainder of the paper, as well as some extensions of the estimator.

1.4.1. Notation

To avoid confusion which might arise due to the frequent occurrence of multiple subscripts, the following notation will be used throughout the paper: if A is a matrix, its entries will be denoted by $A(i,j)$, while $A_{i,j}$ may denote a matrix from some (doubly) indexed set of matrices. Similarly, if θ is a vector, its components will be denoted by $\theta(i)$, while θ_i may denote a vector from some indexed set of vectors. Given a set of scalars $A(i,j), 1 \leq i \leq n, 1 \leq j \leq m$ the notation $[A(i,j)]$ will denote the $n \times m$ matrix whose (i,j)-th entry is $A(i,j)$. This notation will be used only in situations where the scalars $A(i,j)$ and the ranges for i and j are clearly defined. Similarly, given a set of scalars $B(i), 1 \leq i \leq n$, $[B(i)]$ will denote a (column) vector whose i-th entry is B(i).

In the most general setting, one considers a random process Y on the domain D, a subset of a d-dimensional Euclidean space. The process Y is observed at n locations $\{x_i\}_{i=1}^n$, $x_i \in D$. Let $Y_n = (Y(x_1), \ldots, Y(x_n))'$ and $Y_n(i) = Y(x_i), 1 \leq i \leq n$. It will be further assumed that

$$Y_n = X_n\beta + \eta_n \tag{1.4.1}$$

with $\mathbf{E}[\eta_n] = 0$. It will be assumed that X_n has p columns corresponding to different regression terms. Thus $X_n(l,k) = r_k(x_l)$, $1 \leq k \leq p$, $1 \leq l \leq n$, where x_l is the l-th location in the sample and r_k is a continuous function defined on D and it is the k-th regression term in the mean of Y. If present, the term

$r_1 \equiv 1$ corresponds to the (non-zero) constant term in the mean of Y. The term $r_k(x_l) = x_l(1)$, where $x_l(1)$ is the first component of the d-dimensional vector $x_l \in D$, would correspond to a linear trend in the mean of $Y(x)$ in the direction of the first component of x. The matrix X_n will always be known, while the $p \times 1$ vector β may have to be estimated. The function $C_Y(x_1, x_2) = \text{cov}(Y(x_1), Y(x_2)) = \text{cov}(\eta(x_1), \eta(x_2))$ is called the covariance function of the process Y (and of the zero-mean process η). Let $K_{Y,n} = \text{var}(Y_n)$. Thus $K_{Y,n}$ is a symmetric matrix whose entries are $K_{Y,n}(i,j) = C_Y(x_i, x_j)$. If C_θ is a given covariance function model, one defines the symmetric matrix $K_{\theta,n}$ in a similar way, by putting $K_{\theta,n}(i,j) = C_\theta(x_i, x_j)$. Thus $K_{\theta,n}$ is a fixed matrix depending only on the model C_θ and on the set of locations $\{x_i\}_{i=1}^n$, $x_i \in D$.

The model C_θ will always be assumed to be additive, that is of the form

$$C_\theta = \sum_{i=1}^{q} \theta(i) C_i, \tag{1.4.2}$$

where the components C_i are fully specified valid covariance functions and the only parameters to be estimated are the $\theta(i)$. Throughout the paper the components C_i as well as C_Y will be assumed continuous. In Section 1.5.2.3 a discontinuous component W will be introduced, which will result in (possibly discontinuous) models of the form $C_{(\gamma,\theta)} = \gamma W + \sum_{i=1}^{q} \theta(i) C_i$. The difference in notation is meant to emphasise the different nature of the functions involved. In all sections preceding 1.5.3 no stationarity or isotropy assumptions are made about the processes Y or η. In Section 1.5.3 and the remainder of the paper the process η will be assumed isotropic (hence in particular second-order stationary). Thus $\mathbf{E}[\eta(x)] = 0$ for all $x \in D$ and the covariance function C_Y of η (and Y) depends only on $\rho = \| x_1 - x_2 \|$. It will then be convenient to introduce explicitly isotropic

versions of C_Y, C_i and C_θ defined by $\mathbf{C}_Y(\rho) = C_Y(x_1, x_2)$, $\mathbf{C}_i(\rho) = C_i(x_1, x_2)$ and $\mathbf{C}_\theta(\rho) = C_\theta(x_1, x_2)$.

Given a model C_θ of the form (1.4.2), it will be said that the true covariance function C_Y is in the span of C_θ (or, equivalently, in the span of the components $C_i, 1 \leq i \leq q$) if and only if there exists a vector θ_Y such that

$$C_Y = \sum_{i=1}^{q} \theta_Y(i) C_i. \tag{1.4.3}$$

If A is a symmetric matrix, the shorthands $A > 0$ and $A \geq 0$ will mean that A is positive definite and non-negative definite, respectively. Similarly if B is another symmetric matrix of the same size as A, $A > B$ and $A \geq B$ will mean $A - B > 0$ and $A - B \geq 0$, respectively.

1.4.2. Orthogonal projections and estimation with additive models

The goal of this section is to summarise the relationship between orthogonal projections in inner product spaces and linear estimation. The discussion will be rather general and the results presented are well known. However, further sections will often use these results and it may be useful to establish the notation and terminology before proceeding. Issues specific to the estimation of covariogram models will be discussed later. Let $\mathbf{V}_n$ be the linear space of symmetric matrices of size $n \times n$ over the field of real numbers. (One could make this discussion entirely general by considering any vector space.) Let $K_1, \ldots, K_q$ be fixed, linearly independent elements of $\mathbf{V}_n$. In particular, it follows that $q \leq \dim(\mathbf{V}_n) = n(n+1)/2$. Furthermore, one considers the vector subspace $\text{span}(K_1, \ldots, K_q)$ of $\mathbf{V}_n$ ($\text{span}(K_1, \ldots, K_q)$ denotes the space of linear combinations of the elements $K_1, \ldots, K_q$). Let

$$\langle K, J \rangle_V, \qquad K \in \mathbf{V}_n, J \in \mathbf{V}_n \tag{1.4.4}$$

be any inner product defined on the vector space $\mathbf{V}_n$, thus making it into an inner product space. The inner product (1.4.4) gives rise to a norm on $\mathbf{V}_n$ defined by

$$\| K - J \|_V = \langle K - J, K - J \rangle_V^{1/2}. \tag{1.4.5}$$

Let $P(J)$ denote the orthogonal projection of J onto the subspace $\text{span}(K_1, \ldots, K_q)$. Thus P is a linear transformation satisfying

$$\langle K_i, J - P(J) \rangle_V = 0, \qquad i = 1, \ldots, q. \tag{1.4.6}$$

Since $P(J) \in \text{span}(K_1, \ldots, K_q)$, one can write $P(J) = \sum_{i=1}^{q} \theta(i) K_i$. Together with (1.4.6) one obtains

$$\sum_{j=1}^{q} \langle K_i, K_j \rangle_V \, \theta(j) = \langle K_i, J \rangle_V, \qquad i = 1, \ldots, q \tag{1.4.7}$$

or, in matrix form

$$[\langle K_i, K_j \rangle_V] \, \theta = [\langle K_i, J \rangle_V] \tag{1.4.8}$$

where $[\langle K_i, K_j \rangle_V]$ denotes a $q \times q$ matrix whose (i,j)-th element is $\langle K_i, K_j \rangle_V$, $[\langle K_i, J \rangle_V]$ denotes a vector of length q whose i-th element is $\langle K_i, J \rangle_V$ and θ is a vector of length q whose i-th element is $\theta(i)$. It is easy to see that the matrix $[\langle K_i, K_j \rangle_V]$ is invertible under the assumption of linear independence of $K_1, \ldots, K_q$. This implies that the equation (1.4.8) has exactly one solution, given by

$$\theta = [\langle K_i, K_j \rangle_V]^{-1} [\langle K_i, J \rangle_V]. \tag{1.4.9}$$

If Y is an n-dimensional random variable with $\mathbf{E}[Y] = 0$ and $\text{var}(Y) = \mathbf{E}[YY'] = K_Y$, let the subset S_Y of $\mathbf{V}_n$ be defined by $S_Y = \{YY', Y \in \mathbf{R}^n\}$. The random process Y gives rise to a probability measure on S_Y. Therefore,

$$\hat{\theta} = [\langle K_i, K_j \rangle_V]^{-1} [\langle K_i, YY' \rangle_V] \tag{1.4.10}$$

is a random variable. One has

$$\mathbf{E}[\hat{\theta}] = [\langle K_i, K_j \rangle_V]^{-1}[\langle K_i, \mathbf{E}[YY'] \rangle_V] = [\langle K_i, K_j \rangle_V]^{-1}[\langle K_i, K_Y \rangle_V]. \tag{1.4.11}$$

Therefore by the uniqueness of the solution of (1.4.8), one concludes that

$$P(K_Y) = \sum_{i=1}^{q} \mathbf{E}[\hat{\theta}(i)] K_i. \tag{1.4.12}$$

On the other hand, if $K_Y \in \text{span}(K_1, \ldots, K_q)$, then for some vector θ_Y one has

$$K_Y = \sum_{i=1}^{q} \theta_Y(i) K_i. \tag{1.4.13}$$

By elementary properties of projections, one immediately obtains $P(K_Y) = K_Y$ and, again by the uniqueness of the solution of (1.4.8) it follows that $\mathbf{E}[\hat{\theta}(i)] = \theta_Y(i)$.

Summing up, it follows that for any choice of inner product $\langle ., . \rangle_V$, the random variable $\hat{\theta}$ of (1.4.10) is the vector minimising $\| YY' - \sum_{i=1}^{q} \alpha(i) K_i \|_V$. The mean vector $\mathbf{E}[\hat{\theta}]$ is the vector minimising $\| K_Y - \sum_{i=1}^{q} \alpha(i) K_i \|_V$. Furthermore, if K_Y is of the form (1.4.13), then $\mathbf{E}[\hat{\theta}] = \theta_Y$. If on the other hand K_Y is not of the form (1.4.13), the vector $\theta = \mathbf{E}[\hat{\theta}]$, given by (1.4.11) is still a meaningful parameter, since it defines the orthogonal projection $\sum_{i=1}^{q} \theta(i) K_i$ of K_Y onto $\text{span}(K_1, \ldots, K_q)$.

1.4.3. The estimator

The goal is to estimate the unknown covariance function of the process Y from the observations Y_n. If β in the equation (1.4.1) is unknown, it may also be necessary to estimate it, otherwise one can work directly with η_n. In this sense, knowing β is equivalent to putting $X = 0$. To motivate the discussion, it

is initially assumed that $X = 0$ and hence $Y_n = \eta_n$. One observes that

$$\mathbf{E}[Y_n Y_n'] = K_{Y,n}. \tag{1.4.14}$$

Furthermore, valid covariance functions $C_i, 1 \leq i \leq q$ are assumed to be fully specified. The functions C_i give rise to the symmetric matrices $K_{i,n}$. One considers the class of covariance function models in (1.4.2), which results in the covariance matrix models

$$K_{\theta,n} = \sum_{i=1}^{q} \theta_n(i) K_{i,n}, \tag{1.4.15}$$

where it will be assumed that the $\theta(i)$ are such that C_θ is a valid covariance function. A member of the class (1.4.2) is sought which will be in some way closest to the unknown true covariance function C_Y. The approach proposed here is based on the observation that the symmetric $n \times n$ matrices form an $(n(n+1)/2$ - dimensional) inner product space with the inner product

$$\langle A, B \rangle = \mathrm{tr}(AB). \tag{1.4.16}$$

The resulting norm $\| A - B \| = \langle A - B, A - B \rangle^{1/2}$ is the square root of the sum of squares of elements of $A - B$. Following the general approach outlined in Section 1.4.2, in order to estimate the $\theta(i)$, the matrix $Y_n Y_n'$ is projected onto the linear space spanned by the matrices $K_{i,n}$. Equivalently, the $\hat{\theta}_n(i)$ are selected so as to minimise $\| Y_n Y_n' - \sum_{i=1}^{q} \hat{\theta}_n(i) K_{i,n} \|$. The resulting estimator is

$$\hat{\theta}_n = [\mathrm{tr}(K_{i,n} K_{j,n})]^{-1} [\mathrm{tr}(K_{i,n} Y_n Y_n')] = [\mathrm{tr}(K_{i,n} K_{j,n})]^{-1} [Y_n' K_{i,n} Y_n] \tag{1.4.17}$$

where $\hat{\theta}_n = (\hat{\theta}_n(1), \ldots, \hat{\theta}_n(q))'$. (The notation $[\mathrm{tr}(K_{i,n} K_{j,n})]$ denotes a $q \times q$ matrix whose (i,j) - th entry is $\mathrm{tr}(K_{i,n} K_{j,n})$. Similarly, $[\mathrm{tr}(K_{i,n} Y_n Y_n')]$ denotes a $q \times 1$ vector.) The expression (1.4.17) should be compared to the general form (1.4.10). In particular, it follows that if the true covariogram C_Y is of the form (1.4.3) for

some θ_Y, then the estimator is unbiased for θ_Y. Otherwise $\hat{\theta}_n$ is still a meaningful parameter in the sense that $\sum_{i=1}^{q} \hat{\theta}_n(i)K_i$ is the closest (in the sense of the norm defined by the inner product (1.4.16)) matrix of the form $\sum_{i=1}^{q} \alpha_i K_i$ to the matrix YY'. Similarly, $\theta_n = \mathbf{E}[\hat{\theta}_n]$ defines $\sum_{i=1}^{q} \theta_n(i)K_i$, which is the closest matrix of the form $\sum_{i=1}^{q} \alpha_i K_i$ to the matrix K_Y.

In the more general case of unknown β, one considers the residuals $e_n = Y_n - X_n\hat{\beta}$ of the regression model (1.4.1), where $\hat{\beta}$ is the least-squares estimator of β. Denoting the orthogonal projection $I_n - X_n(X_n'X_n)^{-1}X_n'$ by P_n, one obtains the following expression as a generalisation of (1.4.14):

$$\mathbf{E}[e_n e_n'] = P_n K_{Y,n} P_n = U_{Y,n} \tag{1.4.18}$$

where

$$e_n = (I_n - X_n(X_n'X_n)^{-1}X_n')Y_n = P_n Y_n.$$

Hence to estimate the $\theta(i)$, the matrix $e_n e_n'$ is projected onto the linear space spanned by the matrices $U_{i,n} = P_n K_{i,n} P_n$. Equivalently, the $\hat{\theta}_{i,n}$ are selected so as to minimise $\| e_n e_n' - \sum_{i=1}^{q} \hat{\theta}_{i,n} U_{i,n} \|$. The resulting estimator is

$$\hat{\theta}_n = [\mathrm{tr}(U_{i,n}U_{j,n})]^{-1}[\mathrm{tr}(U_{i,n}e_n e_n')] = [\mathrm{tr}(U_{i,n}U_{j,n})]^{-1}[e_n' U_{i,n} e_n]. \tag{1.4.19}$$

Again, if the true covariogram C_Y is of the form (1.4.3) for some θ_Y, then the estimator (1.4.19) is unbiased for θ_Y, by an argument similar to that of Section 1.4.2. The resulting estimate of the covariance function C_Y based on the observations Y_n is $\hat{C}_\theta = \sum_{i=1}^{q} \hat{\theta}_n(i)C_i$.

Brown (1978), Rao (1971), Rao and Kleffe (1988) and Verdooren (1988) considered the MINQUE (minimum norm quadratic unbiased estimator) estimator in a somewhat more restrictive context than that defined by the combination of (1.4.1) and (1.4.3).

The MINQUE methods were designed for the purpose of estimation of variance components and have been studied quite extensively in the context of analysis of variance . Their applications to the problem of estimating the covariogram of a random process have been investigated from a practical point of view by Kitanidis (1985), among others, and from a theoretical point of view by Stein (1987, 1989). A MINQUE estimator of the covariance matrix K_Y of the process Y requires a tuning parameter K_0, which can be thought of as an initial guess of the matrix K_Y to be estimated. The estimator then has the form

$$\hat{\theta}_{MINQUE} = [tr(R_n' K_{0,n}^{-1} K_{i,n} K_{0,n}^{-1} R_n K_{j,n})]^{-1} [Y_n' R_n' K_{0,n}^{-1} K_{i,n} K_{0,n}^{-1} R_n Y_n]$$

where

$$R_n = I_n - X_n (X_n' K_{0,n}^{-1} X_n)^{-1} X_n' K_{0,n}^{-1}.$$

It can be shown that the estimator (1.4.19) is a special case of MINQUE, where $K_0 = I$, the identity matrix. Existing theoretical results establishing properties of MINQUE estimators impose certain assumptions on the relationship between the true covariance matrix K_Y and the initial guess K_0 (in some sense, the two have to be close - for details the reader is referred to Rao and Kleffe, 1988 and Stein, 1989). In the practice of geostatistics, these hypotheses are difficult to establish. Indeed, in the particular case of $K_0 = I$ it is not even clear that the hypotheses can be satisfied for any random process Y. For such a choice of K_0 most available theoretical results for MINQUE do not apply. Kitanidis (1985) considers the MINQUE estimator with $K_0 = I$ in a simulation study comparing it to the MINQUE estimator with $K_0 = \sum_{i=1}^{q} \alpha(i) K_i$ for some fixed values $\alpha(i)$ where the fitted model and the true model were of the form $\sum_{i=1}^{q} \theta(i) K_i$. Not surprisingly, the latter model performed better under those circumstances. However, this estimator can become quite unstable if the true model is not of

the form $\sum_{i=1}^{q} \theta(i)K_i$ (Powojowski, unpublished) and establishing its theoretical properties in such a case is quite difficult. On the other hand, the choice $K_0 = I$ leads to an estimator which among models of the form $K_\theta = \sum_{i=1}^{q} \theta(i)K_i$ gives the one closest (in the sense of the sum of squares of elements of $K_Y - K_\theta$) to the true model. In addition, the computation of MINQUE with K_0 other than I involves inverting large matrices ($n \times n$, where n is the number of observations), while MINQUE with $K_0 = I$, or the projection estimator, requires inverting only small matrices ($q \times q$, where q is the number of components in the model).

As has been pointed out by many authors (for a review, see, for example Rao and Kleffe, 1988), the problem at hand imposes certain constraints on the values of $\theta(i)$ if the resulting estimate is to be a valid covariance function, namely $\sum \theta(i)K_i > 0$ is required. In many situations even more severe constraints may be necessary. It may be required that $\theta(i) > 0$ for all i, and in the covariogram estimation it is necessary that $\sum \theta(i)C_i(x_j, x_k)$ be a positive definite function. When these constraints are imposed, the optimisation may have to be carried out in a convex cone and not the entire vector space. Possible ways of addressing these difficulties include truncated estimators or quadratic optimization with linear constraints. In general, such methods tend to introduce a bias, but they often reduce the estimator's MSE. These concerns are not relevant to the asymptotic results derived in the remainder of this paper and will be ignored in the analysis. However, to apply the estimator (1.4.19) in practice for a finite sample one will have to address these issues.

1.4.4. A class of estimators

In the previous section it was seen that in the $n(n+1)/2$-dimensional vector space of symmetric matrices of size n, the expression $\mathrm{tr}(AB)$ amounts to an

inner product and the element $\sum_{i=1}^{q} \hat{\theta}(i) U_{i,n}$ is the orthogonal projection of the element ee' onto the vector space spanned by the matrices $U_{1,n}, U_{2,n}, \ldots, U_{q,n}$, where the inner product is defined as $\langle A, B \rangle = \mathrm{tr}(AB)$. This may be taken as the definition of $\hat{\theta}_n$ and it suggests a possibility of extending the estimation method to other inner products. In particular, the class of inner products considered can be expressed as

$$\langle A, B \rangle_V = \sum_{k=1}^{n} \sum_{l=1}^{n} A(k,l) B(k,l) V(k,l) \tag{1.4.20}$$

where V is an $n \times n$ symmetric matrix with positive entries. If $V(k,l) \equiv 1$ the new inner product reduces to the old one. Throughout this paper $\langle A, B \rangle$ will denote $\mathrm{tr}(AB)$ unless it is explicitly redefined. Moreover, one easily verifies that

$$\langle A, B \rangle_V = \langle A * V, B \rangle = \langle A, B * V \rangle$$

where $(A * B)(k,l) = A(k,l)B(k,l)$ is the Hadamard matrix product. The resulting estimator may now be expressed as

$$\begin{aligned} \hat{\theta}_{V,n} &= [\mathrm{tr}((U_{i,n} * V_n) U_{j,n})]^{-1} [\mathrm{tr}((U_{i,n} * V_n) e_n e_n')] \\ &= [\mathrm{tr}((U_{i,n} * V_n) U_{j,n})]^{-1} [e_n' (U_{i,n} * V_n) e_n]. \end{aligned} \tag{1.4.21}$$

The element $\sum_{i=1}^{q} \hat{\theta}_{V,n}(i) U_i * V_n$ is the orthogonal projection of ee' onto the space spanned by the elements $U_{1,n} * V_n, U_{2,n} * V_n, \ldots, U_{q,n} * V_n$, where the inner product is $\langle ., . \rangle$ and this fact defines $\hat{\theta}_{V,n}$. Equivalently, $\hat{\theta}_{V,n}$ may be defined by the fact that $\sum_{i=1}^{q} \hat{\theta}_{V,n}(i) U_i$ is the orthogonal projection of ee' onto the space spanned by $U_{1,n}, U_{2,n}, \ldots, U_{q,n}$, this time with the inner product $\langle ., . \rangle_V$.

There may be very good reasons for considering such a modified inner product. For example, in geostatistics the covariogram estimation is only an intermediate step in some spatial prediction procedure such as kriging. In such cases

it is often more important to estimate the covariogram more accurately at short distances, while inaccuracies at greater distances may not be so important. Thus the matrix V may be defined by $V(k,l) = v(\| x_k - x_l \|)$ where v is some non-increasing, positive function of distance.

Another reason for considering a modified estimator will become more apparent later. Briefly, it will be seen that the asymptotic convergence properties of the estimator may be improved by a careful choice of V.

The discussion in Section 1.4.2 shows that in the case where

$$\mathbf{E}[ee'] = U_\theta = \sum_{i=1}^{q} \theta(i) U_i, \tag{1.4.22}$$

that is C_Y is of the form (1.4.3), the estimator $\hat{\theta}_V$ is unbiased for θ_Y regardless of the choice of V.

1.4.5. The moments of $\hat{\theta}_{V,n}$

It is easy to see that the mean and variance of $\hat{\theta}_{V,n}$ are given by

$$\mathbf{E}[\hat{\theta}_{V,n}] = [\mathrm{tr}((U_{i,n} * V_n)U_{j,n})]^{-1}[\mathrm{tr}((U_{i,n} * V_n)U_{Y,n})] \tag{1.4.23}$$

and

$$\mathrm{var}(\hat{\theta}_{V,n}) = [\mathrm{tr}((U_{i,n} * V_n)U_{j,n})]^{-1} \,\mathrm{var}([e_n'(U_{i,n} * V_n)e_n])[\mathrm{tr}((U_{i,n} * V_n)U_{j,n})]^{-1}. \tag{1.4.24}$$

Later it will often be useful to make the assumption

$$\sup_{n, A_n \neq 0} \frac{\mathrm{var}[Y_n' A_n Y_n]}{\mathrm{tr}(A_n K_{Y,n} A_n K_{Y,n})} = c < \infty \tag{1.4.25}$$

where the matrices A_n are symmetric. To ensure that the denominator does not vanish, it will be assumed that $K_{Y,n}$ is nonsingular for all n. In particular, the

condition in (1.4.25) yields a bound for $\text{var}(Y_n' A_n Y_n)$:

$$(1.4.25) \Rightarrow \text{var}(Y_n' A_n Y_n) \leq c \, \text{tr}(A_n K_{Y,n} A_n K_{Y,n}).$$

If Y is a Gaussian process, Y_n is multinormal and $c = 2$, in which case (1.4.25) holds by the relation

$$\text{var}[Y_n' A_n Y_n] = 2 \, \text{tr}(A_n K_{Y,n} A_n K_{Y,n})$$

1.5. Asymptotic results

This section contains the main results of the paper. Firstly, the in-fill asymptotic setting considered throughout the paper is defined. Assuming in-fill sampling on a finite domain, one obtains an expression for the asymptotic mean and a bound for the asymptotic variance of the projection estimator. Subsequently, the effect of varying the size of the domain on which the in-fill sequence is defined on the obtained asymptotic variance bound is investigated. Finally, an example is constructed in which observations are placed with an increasing density on a growing domain in such a way as to produce an estimator which converges in probability to the true covariance function. Various extensions are considered, such as the presence of regression terms in the mean of the process (unknown β in (1.4.1)), or the presence of the so-called nugget effect.

1.5.1. Asymptotic settings

Various asymptotic settings are possible in geostatistics. In all cases it will be assumed that (1.4.1) holds. The number of observations n will be allowed to tend to infinity. However, additional considerations arise in defining a setting for an asymptotic theory. These have to do with the relative locations of the observations x_i in the domain of the process, and the size and shape of the

domain itself. Many authors (e.g., Cressie, 1993) have distinguished between two basic asymptotic settings: the so-called in-fill asymptotics, in which the observations are placed within a compact domain D, and the expanding-domain asymptotics, where the observations are spread over an increasing family of domains $\{D_i\}_{i=1,\ldots}, D_i \subseteq D_{i+1}$. Clearly, even this does not fully define the problem. In both settings the observations may be placed on some regular grid or in any geometric arrangement whatsoever. Each such arrangement generally gives rise to a different model as in (1.4.1), even though the underlying process Y is the same.

Various precise definitions of asymptotic settings have been used by many authors. Infill configurations have been considered, among others, by Stein (1987, 1989), Stein and Handcock (1989), and Lahiri (1996). Expanding-domain schemes in which the minimal distance between observations remains bounded from below by a positive value have been considered by Cressie and Grondona (1992) and Cressie (1993) and others. Finally, schemes combining both in-fill and expanding-domain properties in a sampling configuration have been considered by Hall and Patil (1994) and Lahiri et al. (1999).

Precise meaning will now be given to the notion of in-fill asymptotics used in subsequent discussion. The in-fill asymptotic context differs from that considered by other authors in that it does not require the observations to be equally spaced (as opposed to, for example, Stein, 1987) but nevertheless specifies a precise limiting coverage (as opposed to Stein, 1989, where the limiting coverage is not explicitly characterised). It will later be found that some naturally occurring sampling schemes can be captured by this definition.

Let $\{x_i\}_{i=1}^{\infty} \in D$ denote a sequence of sampling locations within the domain. The vector $Y_n = (Y(x_1), \ldots, Y(x_n))'$ will be referred to as the sample of size

n, while $\{x_1, \ldots, x_n\}$ will be called the sampling configuration of size n. The following technical definition will facilitate further discussion.

Definition 1.5.1. *For a given positive integer k, a collection $\mathbf{A}_k = \{A_1^k, \ldots, A_k^k\}$ of measurable, disjoint, connected subsets of D with equal Lebesgue measure $\mu(A_i^k) = \mu(D)/k$ will be called a regular partition of D. A sequence of regular partitions $\{\mathbf{A}_k\}_{k=1}^{\infty}$ satisfying*

$$\lim_{k\to\infty} \max\{\text{diam}(A_i^k),\ 1 \leq i \leq k\} = 0$$

will be referred to as an in-fill partition sequence.

Now the in-fill asymptotic setting will be introduced.

Definition 1.5.2. *Let D be a compact domain in $\mathbf{R}^d$. Let $\{x_j\}_{j=1}^{\infty}$ be a sequence of points in D. For any subset A of $\mathbf{R}^d$, let $\nu_n(A) = |\{l : 1 \leq l \leq n \text{ and } x_l \in A\}|$, that is, let ν_n be a counting measure with atoms at the points $\{x_l\}_{l=1}^{n}$. The sequence $\{x_j\}_{j=1}^{\infty}$ will be called an in-fill sampling sequence if there exists a continuous function f on D such that for any in-fill partition sequence $\{\mathbf{A}_k\}_{k=1}^{\infty}$ the following condition holds:*

for any positive integer k and any $\epsilon > 0$ there exists N such that for all $1 \leq i \leq k$

$$n > N \Rightarrow \left| \frac{\nu_n(A_i^k)}{n} - \int_{A_i^k} f(u)du \right| < \epsilon. \tag{1.5.1}$$

The function f will be called the sampling intensity function. The triple $(D, \{x_j\}_{j=1}^{\infty}, f)$ will be called an in-fill sampling domain.

Next, in-fill sampling and expanding-domains will be combined. The asymptotic setup here differs slightly from that considered by Hall and Patil (1994) in that the sampling locations are not required to be a realisation of repeated

sampling from a sufficiently regular random variable on D. However, at least in some cases it can be shown that with probability one the sampling configuration obtained by their construction will be a special case of the scheme considered here. The following setup also differs from that of Lahiri et al. (1999) in that the sampling locations are not placed on any regular grid.

One considers a sequence of in-fill sampling domains. Let $D \subseteq \mathbf{R}^d$ be a compact domain on which an in-fill sampling sequence with sampling intensity f is defined. To simplify notation it will be assumed that $\mu(D) = 1$, where μ is the Lebesgue measure. Let $\{r_m\}_{m=1}^{\infty}$ be an increasing and unbounded sequence of real numbers with $r_1 = 1$. Let $T_m(x) = r_m x$, $x \in \mathbf{R}^d$ be the dilation operator. Furthermore, let

$$D_m = T_m(D) \tag{1.5.2}$$

$$\{x_{m,j}\}_{j=1}^{\infty} = \{T_m(x_j)\}_{j=1}^{\infty} \tag{1.5.3}$$

$$f_m(r_m x) = r_m^{-d} f(x), \qquad x \in D \tag{1.5.4}$$

It is easily seen that for a fixed $m = 1, 2, \ldots$, the collection of sets $\mathbf{A}_{m,k} = \{A_{m,1}^k, \ldots, A_{m,k}^k\} = \{T_m(A_1^k), \ldots, T_m(A_k^k)\}$ is a regular partition of D_m and $\{\mathbf{A}_{m,k}\}_{k=1}^{\infty}$ is an in-fill partition sequence. Similarly, for a fixed m the sequence $\{x_{m,j}\}_{j=1}^{\infty}$ is an in-fill sampling sequence on D_m with the sampling intensity function f_m. The following definition will be useful in discussing situations where the sampled domain is allowed to expand.

Definition 1.5.3. *The sequence* $\{(D_m, \{x_{m,j}\}_{j=1}^{\infty}, f_m)\}_{m=1}^{\infty}$, *given by (1.5.2), (1.5.3) and (1.5.4), will be called a sequence of expanding in-fill domains.*

The following example provides an easy way of constructing an in-fill sampling sequence on a given compact, connected domain D. It is a special case of the sampling scheme introduced by Hall and Patil (1994).

Theorem 1.5.1. *If the sequence* $\{x_j\}_{j=1}^{\infty}$ *is a realisation of the sequence of independent random variables* $\{X_j\}_{j=1}^{\infty}$ *having the uniform density on* D*, then with probability one the sequence* $\{x_j\}_{j=1}^{\infty}$ *is an in-fill sampling sequence with* $f \equiv \mu(D)^{-1}$.

The proof of the result follows from a d-dimensional version of the Glivenko-Cantelli theorem (Parthasaraty, 1967).

1.5.2. In-Fill asymptotics

In this section the asymptotic properties of the estimators (1.4.17), (1.4.19) and (1.4.21) are established in the in-fill asymptotic setting, in the sense described in the previous section. The following lemmas will be useful.

Lemma 1.5.1. *Let* $\{x_j\}_{j=1}^{\infty}$ *be an in-fill sampling sequence on* D*, with the intensity sampling function* f*, and* ϕ_1 *and* ϕ_2 *be continuous functions on* $D \times D$*. Then the following convergence is uniform on* $D \times D$*: for any* $(x, y) \in D \times D$

$$\lim_{n\to\infty} \sum_{l=1}^{n} \frac{1}{n} \phi_1(x, x_l)\phi_2(y, x_l) = \int_D \phi_1(x, \xi)\phi_2(y, \xi) f(\xi) d\xi. \tag{1.5.5}$$

Lemma 1.5.2. *Let* $\{x_j\}_{j=1}^{\infty}$ *be an in-fill sampling sequence on* D*, with the intensity sampling function* f*, and* ϕ_1 *and* ϕ_2 *be continuous functions on* $D \times D$*. Then*

$$\lim_{n\to\infty} \sum_{k=1}^{n} \sum_{l=1}^{n} \frac{1}{n^2} \phi_1(x_k, x_l)\phi_2(x_k, x_l) = \int_D \int_D \phi_1(\xi, \eta)\phi_2(\xi, \eta) f(\xi) f(\eta) d\xi d\eta. \tag{1.5.6}$$

Lemma 1.5.3. *Let $\{x_j\}_{j=1}^{\infty}$ be an in-fill sampling sequence on D, with the intensity sampling function f, and $\{\phi_{1,n}\}$ and $\{\phi_{2,n}\}$ be sequences of continuous functions converging uniformly on $D \times D$ to the continuous functions ϕ_1 and ϕ_2, respectively. Then*

$$\lim_{n\to\infty} \sum_{k=1}^{n}\sum_{l=1}^{n} \frac{1}{n^2}\phi_{1,n}(x_k, x_l)\phi_{2,n}(x_k, x_l) = \int_D \int_D \phi_1(\xi,\eta)\phi_2(\xi,\eta)f(\xi)f(\eta)d\xi d\eta. \tag{1.5.7}$$

Lemma 1.5.4. *Let $\{x_j\}_{j=1}^{\infty}$ be an in-fill sampling sequence on D, with the intensity sampling function f, and ϕ_1, ϕ_2, ϕ_3 and ϕ_4 be continuous functions on $D \times D$. Then*

$$\begin{aligned}\lim_{n\to\infty} \sum_{k_1=1}^{n}\sum_{k_2=1}^{n}\sum_{k_3=1}^{n}\sum_{k_4=1}^{n} \frac{1}{n^4}\phi_1(x_{k_1}, x_{k_2})\phi_2(x_{k_2}, x_{k_3})\phi_3(x_{k_3}, x_{k_4})\phi_4(x_{k_4}, x_{k_1}) \\ = \int_D \int_D h_1(\xi,\eta)h_2(\xi,\eta)f(\xi)f(\eta)d\xi d\eta\end{aligned} \tag{1.5.8}$$

where

$$h_1(\xi,\eta) = \int_D \phi_1(\xi,\lambda)\phi_2(\lambda,\eta)f(\lambda)d\lambda$$

$$h_2(\xi,\eta) = \int_D \phi_3(\xi,\lambda)\phi_4(\lambda,\eta)f(\lambda)d\lambda.$$

1.5.2.1. *Stationary process with known mean (case $X = 0$)*

This section explores the properties of the projection estimator as an increasing number of observations from an in-fill sampling sequence on a finite domain become available. It will be initially assumed that $X = 0$ in (1.4.1) and the standard inner product will be used ($V(i,j) \equiv 1$) to obtain the estimator (1.4.17).

Let

$$A_n = (1/n^2)[\mathrm{tr}(K_{i,n}K_{j,n})], \tag{1.5.9}$$

$$M_n = (1/n^2)[\mathrm{tr}(K_{i,n}K_{Y,n})] \tag{1.5.10}$$

and

$$B_n = (1/n^4)[\mathrm{tr}(K_{i,n}K_{Y,n}K_{j,n}K_{Y,n})]. \tag{1.5.11}$$

Furthermore, let

$$E_n = A_n^{-1}\,\mathrm{diag}(B_n)A_n^{-1}, \tag{1.5.12}$$

where $\mathrm{diag}(B_n)$ is a matrix whose diagonal elements are the same as those of B_n, while the off-diagonal elements are zero. Now (1.4.23) may be expressed as

$$\mathbf{E}[\hat{\theta}_n] = A_n^{-1}M_n, \tag{1.5.13}$$

while (1.4.24) may be expressed as

$$\begin{aligned}\mathrm{var}(\hat{\theta}_n) &= A_n^{-1}(1/n^4)\,\mathrm{var}([Y_n'(K_{i,n})Y_n])A_n^{-1} \\ &= A_n^{-1}(1/n^4)[\mathrm{cov}(Y_n'K_{i,n}Y_n, Y_n'K_{j,n}Y_n)]A_n^{-1}. \end{aligned}\tag{1.5.14}$$

The following result follows:

Lemma 1.5.5. *If (1.4.25) holds, then*

$$\mathrm{var}(\hat{\theta}_n(i)) \le qcE_n(i,i), \qquad i = 1, \ldots, q. \tag{1.5.15}$$

where c is given by (1.4.25).

This result provides an upper bound for the variance of the components of $\hat{\theta}_n$. Of course, in the case of a Gaussian process Y, one has a more precise relationship, namely

$$\mathrm{var}(\hat{\theta}_n) = 2A_n^{-1}B_nA_n^{-1}. \tag{1.5.16}$$

If the sequence of models in (1.4.1) originates from an in-fill sampling sequence and the model component functions C_i, as well as the true covariance function C_Y are all continuous, the following limits result from the lemmas of the previous section:

$$\begin{aligned} A(i,j) = \lim_{n\to\infty} \frac{1}{n^2}\,\mathrm{tr}(K_{i,n}K_{j,n}) &= \lim_{n\to\infty} \sum_{k=1}^{n}\sum_{l=1}^{n} \frac{1}{n^2} C_i(x_k,x_l)C_j(x_l,x_k) \\ &= \int_D\int_D C_i(\xi_1,\xi_2)C_j(\xi_2,\xi_1)f(\xi_1)f(\xi_2)d\xi_1 d\xi_2, \end{aligned} \tag{1.5.17}$$

$$\begin{aligned} M(i) = \lim_{n\to\infty} \frac{1}{n^2}\,\mathrm{tr}(K_{i,n}K_{Y,n}) &= \lim_{n\to\infty} \sum_{k=1}^{n}\sum_{l=1}^{n} \frac{1}{n^2} C_i(x_k,x_l)C_Y(x_l,x_k) \\ &= \int_D\int_D C_i(\xi_1,\xi_2)C_Y(\xi_2,\xi_1)f(\xi_1)f(\xi_2)d\xi_1 d\xi_2 \end{aligned} \tag{1.5.18}$$

and similarly

$$\begin{aligned} B(i,j) &= \lim_{n\to\infty} \frac{1}{n^4}\,\mathrm{tr}(K_{i,n}K_{Y,n}K_{j,n}K_{Y,n}) \\ &= \lim_{n\to\infty} \sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{m=1}^{n}\sum_{r=1}^{n} \frac{1}{n^4} C_i(x_k,x_l)C_Y(x_l,x_m)C_j(x_m,x_r)C_Y(x_r,x_k) \\ &= \int_D\int_D h_i(\xi,\eta)h_j(\eta,\xi)f(\xi)f(\eta)d\xi d\eta \end{aligned} \tag{1.5.19}$$

where

$$h_i(\xi,\eta) = \int_D C_i(\xi,\lambda)C_Y(\lambda,\eta)f(\lambda)d\lambda.$$

The following result is an easy consequence.

Theorem 1.5.2. *Under the model (1.4.1), with known β, or $X = 0$, for observations coming from an in-fill sampling sequence, if the matrix A of (1.5.17) is invertible, the projection estimator defined by (1.4.17) has the limiting mean of*

$$\lim_{n\to\infty} \mathbf{E}[\hat{\theta}_n] = A^{-1}M = \theta. \tag{1.5.20}$$

In addition, the following limit exists:

$$\lim_{n\to\infty} E_n = A^{-1}\,\mathrm{diag}(B)A^{-1} = E \tag{1.5.21}$$

where the matrices B and M are given by (1.5.19) and (1.5.18). If, moreover, (1.4.25) holds, the limiting variance of the estimator satisfies

$$\lim_{n\to\infty} \mathrm{var}(\hat{\theta}_n(i)) \le qc\ E(i,i). \tag{1.5.22}$$

If C_Y is of the form (1.4.3), its parameter vector θ_Y is given by (1.5.20) (in that case $\mathbf{E}[\hat{\theta}_n] = \theta_Y$ for all n). The first part of this results follows from

$$\mathbf{E}[\hat{\theta}_n] = [\mathrm{tr}(K_{i,n}K_{j,n})]^{-1}\mathbf{E}[Y_n'K_{i,n}Y_n] = [\mathrm{tr}(K_{i,n}K_{j,n})]^{-1}[\mathrm{tr}(K_{i,n}K_{Y,n})]$$

combined with earlier definitions, while assertion (1.5.22) is a consequence of Lemma 1.5.5.

Thus it is seen that as the number of observations increases, the variance of the components of $\hat{\theta}_n$ remains bounded by the diagonal entries of the matrix E. It can also be easily seen that at least in the Gaussian case the variance of $\hat{\theta}_n$ cannot be made arbitrarily small by increasing the number of observations n. The matrix B will generally not be zero. To see this, it is sufficient to consider a

case where the covariance functions C_Y and C_i are non-negative. The following corollary follows from (1.5.16) and (1.5.21):

Corollary 1.5.1. *Under the model (1.4.1), where the observations come from an in-fill sampling sequence, the process Y is Gaussian and the matrix B in (1.5.19) is not zero, the projection estimator of θ (defined by (1.4.17)) is inconsistent (has a positive limiting variance).*

To see this, one recalls that in the Gaussian case $\lim_{n\to\infty} \mathrm{var}(\hat{\theta}_n) = 2A^{-1}BA^{-1}$. Since A is invertible, $A^{-1}BA^{-1}$ has the same rank as B and it is non-negative definite, since it is a variance matrix. If it were possible to estimate $\hat{\theta}_n$ consistently, this would imply $B = 0$. Corollary 1.5.1 is not particularly surprising and it should be compared with the results of Matheron (1965), who shows the impossibility of consistent estimation of the empirical variogram based on complete information about the process over a finite domain.

By construction, the estimator (1.4.17) is unbiased for θ_Y if C_Y is of the form (1.4.3). In this case, the resulting $C_{\hat{\theta}_n}(x_1, x_2)$ is an unbiased estimator of $C_Y(x_1, x_2)$ for all n. If, on the other hand, C_Y is not of the form (1.4.3), then the parameter θ can still be defined as the limit of $\theta_n = \mathbf{E}[\hat{\theta}_n]$. From earlier discussion and from the discussion in Section 1.4.2 it follows that

$$\lim_{n\to\infty} \mathbf{E}[C_{\hat{\theta}_n}(x, y)] = \sum_{i=1}^{q} (\lim_{n\to\infty} \mathbf{E}[\hat{\theta}_n(i)]) C_i(x, y) = \sum_{i=1}^{q} \theta(i) C_i(x, y) \qquad (1.5.23)$$

viewed as a function of (x, y) is the orthogonal projection of the function $C_Y(x, y)$ onto the space spanned by the functions $C_i(x, y)$, where the inner product between two functions ϕ_1, ϕ_2 on D^2 is defined by

$$\langle \phi_1, \phi_2 \rangle = \int_D \int_D \phi_1(\xi, \eta) \phi_2(\eta, \xi) f(\xi) f(\eta) d\xi d\eta. \qquad (1.5.24)$$

This is somewhat comforting, since it means that if the functions C_i are selected so that the space they span is sufficiently rich to contain elements close to C_Y, the obtained estimator's bias should be small.

The results of this section hold with slight modifications for the more general estimator $\hat{\theta}_{V,n}$ if the matrix V_n is of the form

$$V_n(k,l) = v(x_k, x_l) \tag{1.5.25}$$

where v is a continuous positive function on D^2. It suffices to define the matrices A, M and B as

$$\begin{aligned} A(i,j) &= \lim_{n\to\infty} \frac{1}{n^2} \operatorname{tr}((K_{i,n} * V_n)K_{j,n}) \\ &= \int_D \int_D C_i(\xi_1,\xi_2)C_j(\xi_2,\xi_1)v(\xi_1,\xi_2)f(\xi_1)f(\xi_2)d\xi_1 d\xi_2, \end{aligned} \tag{1.5.26}$$

$$\begin{aligned} M(i) &= \lim_{n\to\infty} \frac{1}{n^2} \operatorname{tr}((K_{i,n} * V_n)K_{Y,n}) \\ &= \int_D \int_D C_i(\xi_1,\xi_2)C_Y(\xi_2,\xi_1)v(\xi_1,\xi_2)f(\xi_1)f(\xi_2)d\xi_1 d\xi_2 \end{aligned} \tag{1.5.27}$$

and similarly

$$\begin{aligned} B(i,j) &= \lim_{n\to\infty} \frac{1}{n^4} \operatorname{tr}((K_{i,n} * V_n)K_{Y,n}(K_{j,n} * V_n)K_{Y,n}) \\ &= \int_D \int_D h_i(\xi,\eta)h_j(\eta,\xi)f(\xi)f(\eta)d\xi d\eta \end{aligned} \tag{1.5.28}$$

where

$$h_i(\xi,\eta) = \int_D C_i(\xi,\lambda)C_Y(\lambda,\eta)v(\xi,\lambda)f(\lambda)d\lambda.$$

Theorem 1.5.2 holds and its proof is virtually unchanged.

The remarks following Theorem 1.5.2 also hold, but the inner product in (1.5.24) is replaced by

$$\langle \phi_1, \phi_2 \rangle_V = \int_D \int_D \phi_1(\xi,\eta)\phi_2(\eta,\xi)v(\xi,\eta)f(\xi)f(\eta)d\xi d\eta. \tag{1.5.29}$$

1.5.2.2. *Process with unknown mean or a trend ($X \neq 0$)*

This section extends the results of the previous section to the case of a process with unknown mean. The presence of the unknown parameter β results in the estimator of the form (1.4.19). Let X_n be as described in Section 1.4.1. Let $R_n = (1/n)X_n'X_n$ and $S_n = n(X_n'X_n)^{-1}$. It follows from Lemma 1.5.1 that

$$\begin{aligned}\lim_{n\to\infty} R_n(k_1,k_2) &= \lim_{n\to\infty} \frac{1}{n}\sum_{l=1}^{n} r_{k_1}(x_l)r_{k_2}(x_l) \\ &= \int_D r_{k_1}(\xi)r_{k_2}(\xi)f(\xi)d\xi = R(k_1,k_2) \end{aligned} \tag{1.5.30}$$

for $1 \leq k_1, k_2 \leq p$. If the functions r_k are such that the matrix R is invertible, then

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty} n(X_n'X_n)^{-1} = R^{-1} = S.$$

If x_l is a row vector of dimension d, representing a point in the domain D (and l-th row in the matrix X_n), let Q denote the function $Q(x_{l_1}, x_{l_2}) = x_{l_1} S x_{l_2}'$, continuous on D^2. In order to generalise the results of the previous section to the case of a process with unknown mean, auxiliary functions will be defined as follows:

$$\begin{aligned}\phi_i(x_{l_1},x_{l_2}) &= C_i(x_{l_1},x_{l_2}) \\ &\quad - \int_D Q(x_{l_1},\xi)C_i(\xi,x_{l_2})f(\xi)d\xi - \int_D Q(x_{l_2},\xi)C_i(x_{l_1},\xi)f(\xi)d\xi \\ &\quad + \int_D \int_D Q(x_{l_1},\xi)Q(x_{l_2},\eta)C_i(\xi,\eta)f(\xi)f(\eta)d\xi d\eta. \end{aligned} \tag{1.5.31}$$

One defines ϕ_Y by replacing C_i by C_Y in the equation above. These definitions are rather technical in nature. Their precise role can be seen in the proofs of the results of this section, but intuitively they arise from considering $\mathrm{cov}(Y(x_1) - \hat{Y}(x_1), Y(x_2) - \hat{Y}(x_2))$ instead of $\mathrm{cov}(Y(x_1), Y(x_2))$, where $\hat{Y} = X\hat{\beta}$. The terms involving integrals can be associated with the covariances of the predictors $\hat{Y}$ with Y and with themselves. The following matrices will play a role similar to that of the matrices A, M and B defined in (1.5.17), (1.5.18) and (1.5.19):

$$A_1(i,j) = \int_D \int_D \phi_i(\xi_1, \xi_2)\phi_j(\xi_2, \xi_1) f(\xi_1) f(\xi_2) d\xi_1 d\xi_2, \tag{1.5.32}$$

$$M_1(i) = \int_D \int_D \phi_i(\xi_1, \xi_2)\phi_Y(\xi_2, \xi_1) f(\xi_1) f(\xi_2) d\xi_1 d\xi_2 \tag{1.5.33}$$

and

$$B_1(i,j) = \int_D \int_D h_{1,i}(\xi, \eta) h_{1,j}(\eta, \xi) f(\xi) f(\eta) d\xi d\eta \tag{1.5.34}$$

with

$$h_{1,i}(\xi, \eta) = \int_D \phi_i(\xi, \lambda)\phi_Y(\lambda, \eta) f(\lambda) d\lambda. \tag{1.5.35}$$

The following result generalises Theorem 1.5.2 to the case of unknown β.

Theorem 1.5.3. *Under the model (1.4.1), with X containing continuous regressor functions and such that the matrix R of (1.5.30) is invertible, if the observations come from an in-fill sampling sequence with sampling density f and the matrix A_1 of (1.5.32) is invertible, the projection estimator defined by (1.4.19) has the limiting mean of*

$$\lim_{n \to \infty} \mathbf{E}[\hat{\theta}_n] = A_1^{-1} M_1 = \theta. \tag{1.5.36}$$

In addition, the following limit exists:

$$\lim_{n\to\infty} E_{1,n} = A_1^{-1}\,\mathrm{diag}(B_1)A_1^{-1} = E_1 \tag{1.5.37}$$

where M_1 and B_1 are given by (1.5.33) and (1.5.34). If, moreover, (1.4.25) holds, the limiting variance of the estimator satisfies

$$\lim_{n\to\infty} \mathrm{var}(\hat{\theta}_n(i)) \leq qc\; E_1(i,i). \tag{1.5.38}$$

Thus in the case of unknown β, the situation is similar to the previously considered case of zero mean. If (1.4.25) holds, as n increases, the limit variance of the estimator of $\hat{\theta}(i)$ is bounded from above by the diagonal elements of the matrix

$$qc\; E_1 = qcA_1^{-1}\;\mathrm{diag}(B_1)\;A_1^{-1} \tag{1.5.39}$$

which may be compared with Lemma 1.5.5. An argument similar to that in Corollary 1.5.1 shows that consistent estimation of θ is in general impossible based on observations drawn from a finite domain.

The discussion of Section 1.4.2 shows again that the estimator is unbiased if C_Y is of the form (1.4.3). If the term $X\beta$ is absent from (1.4.1) the functions defined in (1.5.31) reduce to $C_i(x_{l_1}, x_{l_2})$ and the matrices A_1, M_1 and B_1 reduce to A, M and B, respectively.

As in the previous section, nothing changes substantially if $\hat{\theta}_V$ is considered instead of $\hat{\theta}$. It is only necessary to redefine

$$A_1(i,j) = \int_D \int_D \phi_i(\xi_1,\xi_2)\phi_j(\xi_2,\xi_1)v(\xi_1,\xi_2)f(\xi_1)f(\xi_2)d\xi_1 d\xi_2, \tag{1.5.40}$$

$$M_1(i) = \int_D \int_D \phi_i(\xi_1,\xi_2)\phi_Y(\xi_2,\xi_1)v(\xi_1,\xi_2)f(\xi_1)f(\xi_2)d\xi_1 d\xi_2 \tag{1.5.41}$$

and

$$h_{1,i}(\xi,\eta) = \int_D \phi_i(\xi,\lambda)\phi_Y(\lambda,\eta)v(\xi,\lambda)f(\lambda)d\lambda. \tag{1.5.42}$$

1.5.2.3. *The nugget effect*

In the practice of geostatistics it is common to consider processes of the form

$$Y_\epsilon(x) = Y(x) + \epsilon(x) \tag{1.5.43}$$

where $Y(x)$ is as in previous sections and where $\epsilon(x)$ is a zero-mean random variable with a finite variance γ and where for $x_1 \neq x_2$ the random variables $\epsilon(x_1)$ and $\epsilon(x_2)$ are uncorrelated. The processes $Y(x)$ and $\epsilon(x)$ are assumed uncorrelated as well. The variance of the term $\epsilon(x)$ is traditionally called the nugget effect in geostatistics. This section reviews the effect of the presence of a nugget effect in the model in the setting of the previous section. Firstly, one observes that if (1.5.43) holds, but the model (1.4.2) is not modified to reflect this, there is no effect on the asymptotic mean and variance of $\hat{\theta}_n$. To see this, one examines the formulae (1.5.17), (1.5.18) and (1.5.19). Clearly, (1.5.17) is unaffected, while (1.5.18) yields

$$\begin{aligned} M(i) &= \lim_{n\to\infty} \frac{1}{n^2} \operatorname{tr}(K_{i,n}(K_{Y,n} + \gamma I_n)) \\ &= \lim_{n\to\infty} \frac{1}{n^2} \left(\sum_{k=1}^{n} \sum_{l=1}^{n} (C_i(x_k,x_l)C_Y(x_l,x_k) + \gamma C_i(x_k,x_l)\delta(x_l - x_k)) \right) \end{aligned}$$

where $\delta(x_l - x_k)$ is a discontinuous function equal to one if $x_l = x_k$ and zero otherwise. Hence

$$M(i) = \lim_{n\to\infty} \frac{1}{n^2}\left(\sum_{k=1}^{n}\sum_{l=1}^{n} C_i(x_k, x_l)C_Y(x_l, x_k) + \gamma \sum_{k=1}^{n} C_i(x_k, x_k)\right)$$
$$= \lim_{n\to\infty} \frac{1}{n^2}\left(\sum_{k=1}^{n}\sum_{l=1}^{n} C_i(x_k, x_l)C_Y(x_l, x_k)\right)$$

and it is seen that the limit is unchanged. A similar argument shows that (1.5.19) is also unaffected.

However, the estimation of the importance of the nugget effect is of interest in its own right. To carry out this estimation, a discontinuous covariance component is added to the model (1.4.2). This discontinuous component will be called the nugget effect covariance component and it will be denoted by W in order to differentiate it from the continuous components C_i. The component W is defined as

$$W(\xi, \eta) = \begin{cases} 1 & \text{if } \xi = \eta \\ 0 & \text{otherwise.} \end{cases} \tag{1.5.44}$$

It will be assumed that the model (1.5.43) holds and hence the true covariance function of Y_ϵ is

$$C_{Y,\epsilon} = C_Y + \gamma W \tag{1.5.45}$$

where W is discontinuous as defined above, while C_Y, the covariance function of the process Y, is a continuous function as in the previous sections. The model to be fitted will be of the form

$$C_{\gamma,\theta} = \gamma W + \sum_{i=1}^{q} \theta(i)C_i = \gamma W + C_\theta. \tag{1.5.46}$$

The projection estimator is easily obtained:

$$\begin{bmatrix} \hat{\gamma}_n \\ \hat{\theta}_n \end{bmatrix} = \begin{bmatrix} n-p & [\mathrm{tr}(U_{i,n})]' \\ [\mathrm{tr}(U_{i,n})] & [\mathrm{tr}(U_{i,n}U_{j,n})] \end{bmatrix}^{-1} \begin{bmatrix} e_n'e_n \\ [e_n'U_{i,n}e_n] \end{bmatrix} \tag{1.5.47}$$

where p is the number of columns in the regression matrix X, $[\mathrm{tr}(U_{i,n})]$ is a $q \times 1$ matrix and $[\mathrm{tr}(U_{i,n}U_{j,n})]$ is a $q \times q$ matrix and where the matrices $U_{i,n}$ are defined as in (1.4.19). To extend the results of the previous section to the case with the nugget effect, the following definitions will be useful:

$$A_{ne} = \begin{bmatrix} 1 & a' \\ 0 & A_1 \end{bmatrix} \tag{1.5.48}$$

where A_1 is a $q \times q$ matrix with entries given by (1.5.32) for $1 \leq i \leq q$, $1 \leq j \leq q$ and the vector a, of size q has the entries:

$$a(i) = \int_D \phi_i(\xi, \xi) f(\xi) d\xi \tag{1.5.49}$$

with ϕ_i given by (1.5.31), while

$$M_{ne} = \begin{bmatrix} m_0 \\ M_1 \end{bmatrix} \tag{1.5.50}$$

where M_1 is a $q \times 1$ matrix with entries given by (1.5.33) with $1 \leq q$, and

$$m_0 = \int_D C_{Y,\epsilon}(\xi, \xi) f(\xi) d\xi.$$

The matrix B_{ne} is defined as

$$B_{ne} = \begin{bmatrix} b_0 & b' \\ b & B_1 \end{bmatrix} \tag{1.5.51}$$

where the $q \times q$ matrix B_1 has entries given by (1.5.34). The $q \times 1$ matrix b has entries defined by

$$b(i) = \int_D \int_D h_{1,i}(\xi, \eta)\phi_Y(\eta, \xi)f(\xi)f(\eta)d\xi d\eta$$

with $h_{1,i}$ as in (1.5.35), and

$$b_0 = \int_D \int_D \phi_Y(\xi, \eta)\phi_Y(\eta, \xi)f(\xi)f(\eta)d\xi d\eta.$$

Finally, the matrix $E_{ne,n}$ will be defined as

$$E_{ne} = A_{ne}^{-1}\,\mathrm{diag}(B_{ne})(A'_{ne})^{-1}. \tag{1.5.52}$$

The following result holds:

Theorem 1.5.4. *Under the model (1.5.43) with Y as in (1.4.1), where X contains continuous regressor functions and such that the matrix R of (1.5.30) is invertible, if the observations come from an in-fill sampling sequence with sampling density f and the matrix A_{ne} of (1.5.48) is invertible, the projection estimator defined by (1.5.47) has the limiting mean of*

$$\lim_{n\to\infty} \mathbf{E}\begin{bmatrix}\hat{\gamma}_n \\ \hat{\theta}_n\end{bmatrix} = A_{ne}^{-1}M_{ne} = \begin{bmatrix}m_0 - a'\theta \\ \theta\end{bmatrix} \tag{1.5.53}$$

where θ is given by (1.5.36). In addition, the following limit exists:

$$\lim_{n\to\infty} E_{ne,n} = A_{ne}^{-1}\,\mathrm{diag}(B_{ne})(A'_{ne})^{-1} = E_{ne} \tag{1.5.54}$$

where M_{ne} and B_{ne} are given by (1.5.50) and (1.5.51). If, moreover, (1.4.25) holds, the limiting variance of the estimator satisfies

$$\lim_{n\to\infty} \mathrm{var}(\hat{\gamma}_n) \leq qc\; E_{ne}(1,1) \tag{1.5.55}$$

$$\lim_{n\to\infty} \mathrm{var}(\hat{\theta}_n(i)) \leq qc\; E_{ne}(i+1, i+1), \qquad i = 1, \ldots, q.$$

As in the case without the nugget effect, the limiting variance of the estimator is finite and non-zero in general. For any finite n, the estimator is unbiased if the model (1.5.46) is correct. It is easy to see that if the process is stationary, $m_0 = \text{var}(Y_\epsilon(x))$.

It is a simple matter to generalise the results of this section to the estimator $\hat{\theta}_V$. In order to do so, one defines

$$A_{ne} = \begin{bmatrix} a_0 & a' \\ 0 & A_1 \end{bmatrix} \tag{1.5.56}$$

where A_1 is a $q \times q$ matrix with entries given by (1.5.40) with $1 \leq i \leq q$, $1 \leq j \leq q$ and where

$$a(i) = \int_D \phi_i(\xi, \xi) v(\xi, \xi) f(\xi) d\xi \tag{1.5.57}$$

and

$$a_0 = \int_D v(\xi, \xi) f(\xi) d\xi \tag{1.5.58}$$

while

$$M_{ne} = \begin{bmatrix} m_0 \\ M_1 \end{bmatrix} \tag{1.5.59}$$

where M_1 is a $q \times 1$ matrix with entries given by (1.5.41), and

$$m_0 = \int_D C_Y(\xi, \xi) v(\xi, \xi) f(\xi) d\xi.$$

The matrix B_{ne} is defined as

$$B_{ne} = \begin{bmatrix} b_0 & b' \\ b & B_1 \end{bmatrix} \tag{1.5.60}$$

where the $q \times q$ matrix B_1 has entries

$$B_1(i,j) = \int_D \int_D h_{1,i}(\xi,\eta) h_{1,j}(\eta,\xi) f(\xi) f(\eta) d\xi d\eta$$

with $h_{1,i}(\xi,\eta)$ given by (1.5.42), while the $q \times 1$ matrix b has entries defined by

$$b(i) = \int_D \int_D h_{1,i}(\xi,\eta) \phi_Y(\eta,\xi) v(\eta,\eta) f(\xi) f(\eta) d\xi d\eta$$

with $h_{1,i}$ is as above, and

$$b_0 = \int_D \int_D \phi_Y(\xi,\eta) \phi_Y(\eta,\xi) v(\xi,\xi) v(\eta,\eta) f(\xi) f(\eta) d\xi d\eta.$$

With these changes Theorem 1.5.4 holds and its proof is essentially unchanged. In most applications $v(\xi,\xi)$ will be a constant and in that case there is no loss of generality in assuming $v(\xi,\xi) = 1$, which simplifies the expressions for a_0, m_0, b_0 and b.

1.5.2.4. *The convergence rate for uniform random sampling sequence*

If the true covariance function C_Y is of the form (1.4.3), the mean of the estimator $\hat{\theta}_n$ in (1.4.19) is θ_Y and does not depend on the sampling sequence. Otherwise it depends on the matrices A_n and M_n of (1.5.9) and (1.5.10). The elements of the matrices A_n and M_n can be viewed as approximations of the elements of the matrices A and M defined by the integrals in (1.5.17) and (1.5.18). In fact, for a fixed in-fill sampling sequence $\{x_j\}$, the elements of M_n and A_n are Riemann sums for the integrals in A and M. The rate of convergence will in general depend on the in-fill sampling sequence. In the case of the sampling sequence described in Theorem 1.5.1, one can view the elements of A_n and B_n as a kind of quasi-Monte Carlo estimates - random variables, and it is possible to establish the convergence rate for this approximation. Unlike in the previous

sections, in this section the in-fill sampling sequence will be considered as a sequence of random variables taking values in D. These ideas are summarised in the following lemma.

Lemma 1.5.6. *Let $\{X_j\}$ be a sequence of independent random variables having the uniform density on D. Let ϕ_1 and ϕ_2 be continuous functions and let $\phi_i(x,y) = \phi_i(y,x)$, for $(x,y) \in D^2, i = 1,2$. Let*

$$Z_n = \sum_{k=1}^{n}\sum_{l=1}^{n} \frac{1}{n(n-1)}\phi_1(X_k, X_l)\phi_2(X_l, X_k)$$

Then

$$\lim_{n\to\infty} \mathbf{E}[Z_n] = \int_D \int_D \phi_1(\xi,\eta)\phi_2(\eta,\xi)f(\xi)f(\eta)d\xi d\eta = \kappa(D).$$

Moreover,

$$n^{1/2}\left(\frac{Z_n - \kappa(D)}{2\varsigma_1^{1/2}}\right) \to_d \mathbf{N}(0,1) \tag{1.5.61}$$

where

$$\varsigma_1 = \int_D \left(\int_D \phi_1(\xi,\eta)\phi_2(\eta,\xi)f(\xi)f(\eta)d\xi\right)^2 f(\eta)d\eta - (\kappa(D))^2.$$

To emphasise that the entries of A_n and B_n are now random variables depending on the realisation of the process $X = (X_1, X_2, \dots)$, they will be denoted by $A_n(X)(i,j), 1 \le i \le q, 1 \le j \le q$ and $M_n(X)(i), 1 \le i \le q$. The following result follows:

Theorem 1.5.5. *Let $\{X_j\}$ be a sequence of random variables as in Lemma 1.5.6. Then*

$$n^{1/2}(A_n(X)^{-1}M_n(X) - A^{-1}M)$$

converges in distribution to a q-dimensional multinormal random variable with zero mean.

1.5.3. Asymptotics on expanding domains

In the previous sections it was shown that under some rather general conditions, if the observations come from an in-fill sampling sequence on a compact domain, the variances of the components $\hat{\theta}(i)$ of the estimator (1.4.19) are bounded by a multiple of the diagonal elements of the matrix E (where the precise definition of E may be given by (1.5.21), (1.5.37) or (1.5.54) depending on context). It was also seen that those bounds are generally strictly positive and that even in the simple case of a Gaussian process, $\text{var}(\hat{\theta}_n(i))$ remains bounded away from zero.

In this section attention will be focused on isotropic processes and the effect of the size of the finite domain on the limiting variance will be considered. The limit matrix E will be found to depend on the size of the domain D. The main result derived in the subsequent sections states that under fairly mild regularity conditions the entries of the matrix E converge to zero as the sampled domain D is allowed to grow indefinitely. It will therefore follow that $\text{var}(\hat{\theta}_n(i))$ can be made arbitrarily small by sampling a sufficiently large domain at a sufficient number of locations.

It will be assumed that a sequence $\{(D_m, \{x_{m,j}\}_{j=1}^{\infty}, f_m)\}_{m=1}^{\infty}$ of expanding in-fill domains is given, as in Definition 1.5.3. The subscript m runs over domains in the expanding sequence $\{D_m\}$. This will be the meaning of all subscripts m in this section and it is different from the notation used in the previous sections considering a fixed compact domain, where the subscript n ran over observations in a sampling sequence.

The results obtained so far do not make the assumption of stationarity of η in (1.4.1). Throughout the remaining sections it will be assumed that the random process η, as well as the component covariance models C_i, are isotropic

and therefore (second order) stationary. In the case of an isotropic process the covariance functions $C_i(x_1, x_2)$, $C_\theta(x_1, x_2)$ and $C_Y(x_1, x_2)$ depend only on $\rho = \| x_1 - x_2 \|$ and the following notation will be used along with the current notation: $\mathbf{C}_i(\rho) = C_i(x_1, x_2)$, $\mathbf{C}_\theta(\rho) = C_\theta(x_1, x_2)$ and $\mathbf{C}_Y(\rho) = C_Y(x_1, x_2)$. Thus the model (1.4.1) will remain unchanged, but (1.4.2) will now have an equivalent isotropic version

$$\mathbf{C}_\theta = \sum_{i=1}^{q} \theta(i)\mathbf{C}_i. \tag{1.5.62}$$

1.5.3.1. *Isotropic fields*

In the case of the isotropic field, the equations (1.5.17) and (1.5.18) take a simplified form. The double integrals of (1.5.17) and (1.5.18) can be replaced by single integrals if the general covariance forms C_Y, C_θ and C_i are replaced by their isotropic counterparts $\mathbf{C}_Y$, $\mathbf{C}_\theta$ and $\mathbf{C}_i$. These new expressions will be essential in subsequent considerations deriving the limit of the matrix E as the size of the domain D grows indefinitely. This section establishes some fairly technical details necessary to obtain the new expressions.

One considers the measure

$$\mathbf{F}_2(B) = \int_B f(\xi) f(\eta) d\xi d\eta \tag{1.5.63}$$

where B is any Lebesgue-measurable subset of D^2. The following defines a measure on $[0, \mathrm{diam}(D)]$:

$$\mathbf{G}(A) = \mathbf{F}_2(\{(\xi, \eta) : \| \xi - \eta \| \in A\}) \tag{1.5.64}$$

where A is any Lebesgue-measurable subset of $[0, \mathrm{diam}(D)]$. (The fact that the distance function is continuous guarantees that the set $\{(\xi, \eta) : \| \xi - \eta \| \in A\}$ is measurable.) It is easily seen that if $\phi_i(\xi, \eta) = \Phi_i(\rho), i =$

1, 2 are measurable and isotropic, then

$$\int_D \int_D \phi_1(\xi,\eta)\phi_2(\eta,\xi)d\mathbf{F}_2(\xi,\eta) = \int_0^{\mathrm{diam}(D)} \Phi_1(\rho)\Phi_2(\rho)d\mathbf{G}(\rho). \tag{1.5.65}$$

The measure $\mathbf{G}$ will generally depend on the size and shape of D, as well as on the nature of f, but the following lemma holds in general.

Lemma 1.5.7. *If $\mathbf{F}_2$ is absolutely continuous (with respect to the Lebesgue measure) then so is $\mathbf{G}$.*

The non-negative function $G(\rho)$, $\rho > 0$ will be defined by

$$G(\rho) = \mathbf{F}_2(\{(\xi,\eta) :\| \xi - \eta \| \leq \rho\}) = \mathbf{G}([0,\rho]). \tag{1.5.66}$$

Examples of the function G for some regular domains for $d = 2$ can be found in Bartlett (1964) or Diggle (1983).

Next one considers a sequence of expanding in-fill domains $\{\{D_m\}, \{x_{m,j}\}_{j=1}^{\infty}, f_m\}_{m=1}^{\infty}$ as in Definition 1.5.3, with constant sampling density functions f_m, which gives rise to the sequences of measures $\{\mathbf{F}_{2,m}\}_{m=1}^{\infty}$ and functions $\{G_m\}_{m=1}^{\infty}$, where $\mathbf{F}_{2,m}$ is defined by (1.5.63) with f replaced by f_m, and G_m is given by (1.5.66) with $\mathbf{F}_2 = \mathbf{F}_{2,m}$. The sequence $\{G_m\}_{m=1}^{\infty}$ will play an important role in the next lemma. The following definition will also be needed. Let

$$B_\rho(D) = \{x \in D :\| x - y \| < \rho \Rightarrow y \in D\} \tag{1.5.67}$$

that is, $B_\rho(D)$ is the set of all points in D with the property that the interior of the ball of radius ρ around the point is contained in D. Loosely speaking, the following lemma decomposes the function G_m into two components: one which depends only on the dimensionality d of the embedding space, and another, which

depends on the geometry of the domain and which may be associated with the edge effect.

Lemma 1.5.8. *Let $D \subset \mathbf{R}^d$ be a compact domain and $\{x_j\}_{j=1}^{\infty}$ be an in-fill sampling sequence on D, with constant sampling intensity function $f \equiv \mu(D)^{-1}$. Let*

$$\{(D_m, \{x_{m,j}\}_{j=1}^{\infty}, f_m)\}_{m=1}^{\infty}$$

be a sequence of expanding in-fill domains. Let D satisfy the following condition:

$$\sup_{\rho>0} \frac{\mu(D \setminus B_\rho(D))}{\rho\nu(\partial D)} = c < \infty \tag{1.5.68}$$

where μ is the Lebesgue measure on $\mathbf{R}^d$, c is a constant, ∂D is the boundary of D and ν is the Lebesgue measure in $\mathbf{R}^{d-1}$. Then there exists a positive constant α, determined entirely by D, such that

$$G_m(\rho) = \frac{d\pi^{d/2}}{\Gamma(d/2+1)} r_m^{-d}\rho^d - R_m(\rho) \tag{1.5.69}$$

where

$$0 \le R_m(\rho) \le \alpha r_m^{-d-1}\rho^{d+1}.$$

The last lemma gives bounds for the function G. In addition to these bounds two other properties are easy consequences of the definition of G and will be of interest. Firstly, G is non-decreasing. Secondly, if $\mathbf{F}_2$ is absolutely continuous, G is continuous. As a result, the function G may be used to define a Stieltjes integral. The following lemma will be useful

Lemma 1.5.9. *Let ϕ be a measurable real function such that*

$$\lim_{r\to\infty} \frac{1}{r} \int_0^r |\phi(\rho)|\rho^d d\rho = 0. \tag{1.5.70}$$

Then

$$\lim_{r_m\to\infty} r_m^d \int_0^{\mathrm{diam}(D_m)} \phi(\rho) dG_m(\rho) = \alpha_G \int_0^{\infty} \phi(\rho)\rho^{d-1} d\rho \tag{1.5.71}$$

where $\alpha_G = (d\pi^{d/2})/\Gamma(d/2+1)$.

1.5.3.2. *Regularity conditions*

This section summarises most important assumptions which will be made about the observed process Y, its covariogram, the model, its components and the sequence of expanding in-fill domains which will be considered. These restrictions are mostly technical and are invoked to guarantee the existence of various limits appearing in the subsequent results. The subsequent sections will make references to the assumptions enumerated here. For an isotropic function ϕ defined on $\mathbf{R}^d \times \mathbf{R}^d$ and function Φ defined on $\mathbf{R}_+$, such that $\Phi(\| x_1 - x_2 \|) = \phi(x_1, x_2)$, the following two conditions are equivalent

$$\int_{\mathbf{R}^d} |\phi(\xi, 0)|^p d\xi < \infty \iff \int_0^\infty |\Phi(\rho)|^p \rho^{d-1} d\rho < \infty. \tag{1.5.72}$$

It will be assumed that the function $v(x_1, x_2)$ of (1.5.25) is bounded by one and isotropic. The function $\nu(\rho)$ will be defined by

$$\nu(\| x_1 - x_2 \|) = v(x_1, x_2) \leq 1. \tag{1.5.73}$$

In all of the results in the remainder of this paper, the following will be assumed.

Condition 1.5.1. *The sequence of expanding in-fill domains* $\{D_m, \{x_{m,j}\}_{j=1}^\infty, f_m\}$ *is such that* $f_m \equiv \mu(D_m)^{-1}$ *and the conclusion of Lemma (1.5.8) holds.*

Condition 1.5.2. *The covariance function of the process* Y, $\mathbf{C}_Y$ *satisfies (1.5.72) with* $p = 1$.

Condition 1.5.3. *For all covariance component functions* $\mathbf{C}_i$ *in the model, the functions* $\mathbf{C}_i(\rho)\nu(\rho)^{1/2}$ *satisfy (1.5.72) with* $p = 2$.

The last two conditions impose some restrictions on the rate of decay of the true covariance function and the covariance component functions in the model. These restrictions are not very severe. In particular, they are satisfied if the true model decays at the rate of $\rho^{-(1+\epsilon)}$, where $\epsilon > 0$, $\nu \equiv 1$, and the covariance component functions decay at the rate of $\rho^{-(1/2+\epsilon)}$. These rates are much slower than those of most traditional covariogram models considered in practice. The following conditions will be assumed in some, but not all, results of the following sections.

Condition 1.5.4. *For all covariance component functions* $\mathbf{C}_i$ *in the model, the functions* $\mathbf{C}_i(\rho)$ *satisfy (1.5.72) with* $p = 2$.

Clearly, the last condition implies Condition 1.5.3 with any choice of function $\nu(\rho)$.

Condition 1.5.5. *The functions* $\mathbf{C}_i(\rho)\nu(\rho)$ *satisfy (1.5.72) with* $p = 1$.

Clearly, the last condition implies Condition 1.5.3. If the components $\mathbf{C}_i(\rho)$ decay sufficiently fast, this last condition holds for any choice of function $\nu(\rho)$.

Condition 1.5.6. *The function* ν *of (1.5.73) satisfies*

$$\nu(\rho) \leq k_\nu \rho^{-t}$$

for some $k_\nu > 0$ *and some* $t \geq 0$.

The last condition is not very restrictive, indeed setting $v \equiv 1$ corresponds to the case $k_\nu = 1$, $t = 0$. A similar condition with respect to the functions $\mathbf{C}_i$ will also be considered.

Condition 1.5.7. *For some* $\alpha > 0$ *and* $z > d/2$

$$|\mathbf{C}_i(\rho)| \leq \alpha \rho^{-z}, \qquad i = 1, \ldots, q$$

1.5.3.3. *Process with known mean*

This section considers the behaviour of the bound (1.5.22) for the variance of the estimator (1.4.17) as the size of the domain increases. For an isotropic process observed on the domain D_m, the expression (1.5.26) becomes

$$A_m(i,j) = \int_0^{\text{diam}(D_m)} \mathbf{C}_i(\rho)\mathbf{C}_j(\rho)\nu(\rho)dG_m(\rho) \tag{1.5.74}$$

while (1.5.27) becomes

$$M_m(i) = \int_0^{\text{diam}(D_m)} \mathbf{C}_i(\rho)\mathbf{C}_Y(\rho)\nu(\rho)dG_m(\rho). \tag{1.5.75}$$

The limit vector θ in (1.5.20) generalises to

$$\theta_m = A_m^{-1}M_m. \tag{1.5.76}$$

Definition (1.5.28) becomes

$$B_m(i,j) = \int_{D_m}\int_{D_m}\left(\int_{D_m} C_i(\xi,\lambda)C_Y(\lambda,\eta)v(\xi,\lambda)f_m(\lambda)d\lambda\right) \left(\int_{D_m} C_j(\xi,\lambda)C_Y(\lambda,\eta)v(\xi,\lambda)f_m(\lambda)d\lambda\right)f_m(\xi)f_m(\eta)d\xi d\eta. \tag{1.5.77}$$

The matrix E defined by (1.5.21) for the domain D_m becomes

$$E_m = A_m^{-1}\,\text{diag}(B_m)A_m^{-1}. \tag{1.5.78}$$

The following result describes the behaviour of the matrices A_m, M_m, B_m and E_m as the domain size increases in the case where the mean of the process Y is known (in which case it may be assumed to be zero).

Theorem 1.5.6. *For an isotropic process with known β, with a sequence of expanding in-fill domains satisfying Condition 1.5.1, with the true covariance satisfying Condition 1.5.2 and with the model components satisfying Condition 1.5.3,*

the following limits exist

$$\lim_{r_m \to \infty} r_m^d A_m = A, \tag{1.5.79}$$

$$\lim_{r_m \to \infty} r_m^d M_m = M \tag{1.5.80}$$

and

$$\lim_{r_m \to \infty} r_m^{3d} B_m = B. \tag{1.5.81}$$

Moreover, if the matrix A is invertible then the following limits exist:

$$\lim_{r_m \to \infty} \theta_m = \lim_{r_m \to \infty} A_m^{-1} M_m = A^{-1} M = \theta \tag{1.5.82}$$

and

$$\lim_{r_m \to \infty} r_m^d E_m = A^{-1} B A^{-1} = E. \tag{1.5.83}$$

The entries of the matrices A, M and B of Theorem 1.5.6 are given by

$$A(i,j) = \lim_{r_m \to \infty} r_m^d A_m(i,j) = \alpha_G \int_0^\infty \mathbf{C}_i(\rho)\mathbf{C}_j(\rho)\nu(\rho)\rho^{d-1}d\rho$$

$$M(i) = \lim_{r_m \to \infty} r_m^d M_m(i,j) = \alpha_G \int_0^\infty \mathbf{C}_i(\rho)\mathbf{C}_Y(\rho)\nu(\rho)\rho^{d-1}d\rho$$

and

$$B(i,j) = \lim_{r_m \to \infty} r_m^{3d} B_m(i,j) = \alpha_G \int_0^\infty H_i(\rho)H_j(\rho)\rho^{d-1}d\rho$$

where

$$H_i(\| \xi - \eta \|) = h_i(\xi,\eta) = \int_{\mathbf{R}^d} C_i(\xi,\lambda)C_Y(\lambda,\eta)v(\xi,\lambda)d\lambda$$
$$= \int_{\mathbf{R}^d} \psi_i(\xi - \eta - \lambda)\psi_Y(\lambda)d\lambda = \zeta_i(\xi - \eta)$$

where ζ_i is the convolution of the functions ψ_i and ψ_Y defined by

$$\psi_i(\xi) = C_i(\xi, 0)v(\xi, 0)$$

and

$$\psi_Y(\xi) = C_Y(\xi, 0).$$

The convergence of $r_m^d E_m$ leads to the following result:

Corollary 1.5.2. *The bound (1.5.22) for the variance of the estimator (1.4.17) on domain D_m converges to zero at the rate r_m^{-d}, i.e. inversely proportional to the linear size of the domain raised to the power of d, the dimension of the embedding Euclidean space.*

The following corollary is a simple consequence of Theorems 1.5.2 and 1.5.6.

Corollary 1.5.3. *Under the assumptions of Theorem 1.5.6, for a process satisfying (1.4.25), for any $\epsilon > 0$ there exist integers m_ϵ and n_ϵ, such that if the process Y is observed at the locations $x_{m_\epsilon,1}, x_{m_\epsilon,2}, \ldots, x_{m_\epsilon,n_\epsilon}$, the projection estimator (1.4.17) computed from this sampling configuration, $\hat{\theta}_{m_\epsilon,n_\epsilon}$ satisfies*

$$\text{var}(\hat{\theta}_{m_\epsilon,n_\epsilon}(i)) \leq \epsilon, \qquad i = 1, \ldots, q. \tag{1.5.84}$$

This result states that even though in-fill sampling on a compact domain leads to an estimator (1.4.17) whose limiting variance is positive, this limiting variance vanishes as the size of the domain tends to infinity. Therefore, sampling a sufficiently large domain sufficiently densely produces an estimator with arbitrarily small variance. The value of m_ϵ may be obtained from Theorem 1.5.6 and the bound in Lemma 1.5.5. It suffices to take the smallest m_ϵ such that $qcr_{m_\epsilon}^{-d} E_{i,i} < \epsilon, i = 1, \ldots, q$. The choice of n_ϵ depends on the rate of convergence in (1.5.21) in Theorem 1.5.2 applied to the domain D_{m_ϵ}, which ultimately depends on the geometry of the in-fill sampling sequence. The case of the in-fill

sampling described in Theorem 1.5.1 (uniform random sampling) is considered in detail in Section 1.5.4. In that case it is shown how to obtain the number of observations $n(m)$ required on the domain D_m in order to guarantee the convergence in probability of $\hat{\theta}_{m,n}$.

1.5.3.4. *Process with unknown mean*

Theorem 1.5.6 can be extended to a process where the regression term $X\beta$ is present. However, certain restrictions will be imposed to guarantee convergence. Firstly, X will be assumed to contain homogeneous regression functions. Thus if the k-th column of X contains elements $r_k(x_l), x_l \in D$, it will be required that $r_k(rx) = r^\gamma r_k(x)$ for some constant γ and for all $x \in \mathbf{R}^d$ and for all $r > 0$. In particular, any monomial involving any number of components of $x \in \mathbf{R}^d$ is a homogeneous function. This means that any polynomial trend can be modelled. Secondly, the rate of convergence of the bound (1.5.21) for the estimator's variance may depend on the rate at which $\mathbf{C}_i(\rho)$ decreases with increasing ρ. If this rate is too slow, the convergence at the rate r_m^{-d} may still be attained for a modified estimator $\hat{\theta}_V$, for a suitable function ν defined by (1.5.73). In the generalisation of Theorem 1.5.6 to the case of Y with unknown mean, the matrices A_m, M_m, B_m and E_m will be replaced by their generalised counterparts $A_{1,m}$, $M_{1,m}$, $B_{1,m}$ and $E_{1,m}$. The entries of the matrix $A_{1,m}$ are given by

$$A_{1,m}(i,j) = \int_{D_m}\int_{D_m} \phi_{i,m}(\xi_1,\xi_2)\phi_{j,m}(\xi_2,\xi_1)v(\xi_1,\xi_2)f_m(\xi_1)f_m(\xi_2)d\xi_1 d\xi_2 \quad (1.5.85)$$

where

$$\phi_{i,m}(x_{l_1}, x_{l_2}) = C_i(x_{l_1}, x_{l_2}) - \int_{D_m} Q(x_{l_1}, \xi) C_i(\xi, x_{l_2}) f_m(\xi) d\xi - \int_{D_m} Q(x_{l_2}, \xi) C_i(x_{l_1}, \xi) f_m(\xi) d\xi + \int_{D_m} \int_{D_m} Q(x_{l_1}, \xi) Q(x_{l_2}, \eta) C_i(\xi, \eta) f_m(\xi) f_m(\eta) d\xi d\eta \quad (1.5.86)$$

and where one defines $\phi_{Y,m}$ by replacing C_i by C_Y in the equation above. The entries of the matrix $M_{1,m}$ are given by

$$M_{1,m}(i) = \int_{D_m} \int_{D_m} \phi_{i,m}(\xi_1, \xi_2) \phi_{Y,m}(\xi_2, \xi_1) v(\xi_1, \xi_2) f_m(\xi_1) f_m(\xi_2) d\xi_1 d\xi_2, \quad (1.5.87)$$

those of $B_{1,m}$ are given by

$$B_{1,m}(i,j) = \int_{D_m} \int_{D_m} h_{1,i,m}(\xi, \eta) h_{1,j,m}(\eta, \xi) f_m(\xi) f_m(\eta) d\xi d\eta \quad (1.5.88)$$

with

$$h_{1,i,m}(\xi, \eta) = \int_{D_m} \phi_{i,m}(\xi, \lambda) \phi_{Y,m}(\lambda, \eta) v(\xi, \lambda) f(\lambda) d\lambda, \quad (1.5.89)$$

and finally, the matrix $E_{1,m}$ is given by

$$E_{1,m} = A_{1,m}^{-1} \operatorname{diag}(B_{1,m}) A_{1,m}^{-1}. \quad (1.5.90)$$

The following result generalises Theorem 1.5.6 to the case of a process with unknown mean.

Theorem 1.5.7. *For an isotropic process with X composed of homogeneous regressors, with a sequence of expanding in-fill domains satisfying Condition 1.5.1, with the true covariance satisfying Condition 1.5.2, the model components satisfying Condition 1.5.3, and with the matrices A, M, B and E defined in Theorem*

1.5.6 the following limits hold:

$$\lim_{r_m \to \infty} r_m^d A_{1,m} = A \tag{1.5.91}$$

$$\lim_{r_m \to \infty} r_m^d M_{1,m} = M. \tag{1.5.92}$$

If one of the following conditions holds:

i) Condition 1.5.5 holds,

ii) both Conditions 1.5.6 and 1.5.7 hold, with $z + t \geq d$

then

$$\lim_{r_m \to \infty} r_m^{3d} B_{1,m} = B \tag{1.5.93}$$

and

$$\lim_{r_m \to \infty} r_m^d E_{1,m} = A^{-1} B A^{-1} = E. \tag{1.5.94}$$

If neither i) nor ii) holds, but both Conditions 1.5.6 and 1.5.7 hold, then

$$\lim_{r_m \to \infty} r_m^{2d+z+t} B_{1,m} = 0 \tag{1.5.95}$$

and

$$\lim_{r_m \to \infty} r_m^{z+t} E_{1,m} = 0. \tag{1.5.96}$$

If (1.5.94) holds, the bound (1.5.38) for the variance of the estimator (1.4.21) on domain D_m converges to zero at the rate r_m^{-d}, whereas if (1.5.96) holds, the rate is r_m^{-z-t}, which is slower since $z + t < d$ in this case. Thus Corollary 1.5.3 generalises to the case of a process with unknown mean in an obvious way.

1.5.3.5. *The nugget effect*

The last result can be generalised to include the case where the nugget effect is included in the model. Thus it will be assumed that the observed process Y_ϵ is of the form (1.5.43) and the fitted model is of the form (1.5.45). To simplify the notation it will be assumed that $v(\xi, \xi) = 1$. One defines

$$A_{ne,m} = \begin{bmatrix} 1 & a'_m \\ 0 & A_{1,m} \end{bmatrix} \tag{1.5.97}$$

where $A_{1,m}$ is a $q \times q$ matrix given by (1.5.85), and the entries of the (q-dimensional) vector a_m are

$$a_m(i) = \int_{D_m} \phi_{i,m}(\xi, \xi) f_m(\xi) d\xi \tag{1.5.98}$$

with $\phi_{i,m}$ defined as in (1.5.86), while

$$M_{ne,m} = \begin{bmatrix} m_{0,m} \\ M_{1,m} \end{bmatrix} \tag{1.5.99}$$

with the entries of the $q \times 1$ matrix $M_{1,m}$ given by (1.5.87) and

$$m_{0,m} = \int_{D_m} C_{Y,\epsilon}(\xi, \xi) f_m(\xi) d\xi = \gamma + \mathbf{C}_Y(0)$$

if $f_m = 1/\mu(D_m)$ for all m. The matrix $B_{ne,m}$ is defined as

$$B_{ne,m} = \begin{bmatrix} b_{0,m} & b'_m \\ b_m & B_{1,m} \end{bmatrix} \tag{1.5.100}$$

where the $q \times q$ matrix $B_{1,m}$ is given by (1.5.88), while

$$b_m(i) = \int_{D_m} \int_{D_m} h_{1,i,m}(\xi, \eta) \phi_{Y,m}(\eta, \xi) f_m(\xi) f_m(\eta) d\xi d\eta \tag{1.5.101}$$

and

$$b_{0,m} = \int_{D_m} \int_{D_m} \phi_{Y,m}(\xi,\eta)\phi_{Y,m}(\eta,\xi) f_m(\xi) f_m(\eta) d\xi d\eta. \tag{1.5.102}$$

The matrix $E_{ne,m}$ is given by

$$E_{ne,m} = A_{ne,m}^{-1} \operatorname{diag}(B_{ne,m})(A'_{ne,m})^{-1}. \tag{1.5.103}$$

Let a be a vector of length q with entries given by $a(i) = C_i(0)$. The following result generalises Theorem 1.5.7 to the case of a process with nugget effect.

Theorem 1.5.8. *For a process Y_ϵ defined as in (1.5.43), where Y an isotropic process with X composed of homogeneous regressors, with a sequence of expanding in-fill domains satisfying Condition 1.5.1, with the true covariance satisfying Condition 1.5.2 and with the model components satisfying Condition 1.5.3, the following limits hold, with the matrices A, M, B and E defined in Theorem 1.5.6*

$$\lim_{r_m \to \infty} \begin{bmatrix} 1 & 0 \\ 0 & r_m^d I_q \end{bmatrix} A_{ne,m} = \begin{bmatrix} 1 & a' \\ 0 & A \end{bmatrix} \tag{1.5.104}$$

$$\lim_{r_m \to \infty} \begin{bmatrix} 1 & 0 \\ 0 & r_m^d I_q \end{bmatrix} M_{ne,m} = \begin{bmatrix} \gamma + \mathbf{C}_Y(0) \\ M \end{bmatrix}. \tag{1.5.105}$$

If one of the following conditions holds:

i) Condition 1.5.5 holds,

ii) both Conditions 1.5.6 and 1.5.7 hold, with $z + t \geq d$

then

$$\lim_{r_m \to \infty} r_m^d \begin{bmatrix} 1 & 0 \\ 0 & r_m^d I_q \end{bmatrix} B_{ne,m} \begin{bmatrix} 1 & 0 \\ 0 & r_m^d I_q \end{bmatrix} = \begin{bmatrix} b_0 & b' \\ b & B \end{bmatrix} \tag{1.5.106}$$

where b_0 and the elements of the vector b are finite, and

$$\lim_{r_m \to \infty} r_m^d E_{ne,m} = \begin{bmatrix} 1 & a' \\ 0 & A \end{bmatrix}^{-1} \begin{bmatrix} b_0 & b' \\ b & B \end{bmatrix} \begin{bmatrix} 1 & 0 \\ a & A \end{bmatrix}^{-1} = E_{ne}. \tag{1.5.107}$$

If neither i) nor ii) holds, but both Conditions 1.5.6 and 1.5.7 hold, then

$$\lim_{r_m \to \infty} r_m^{z+t} \begin{bmatrix} 1 & 0 \\ 0 & r_m^d I_q \end{bmatrix} B_{ne,m} \begin{bmatrix} 1 & 0 \\ 0 & r_m^d I_q \end{bmatrix} = 0 \tag{1.5.108}$$

and

$$\lim_{r_m \to \infty} r_m^{z+t} E_{ne,m} = 0. \tag{1.5.109}$$

The implications for the limiting variance of the estimator $\hat{\theta}_V$ are the same as in the case without the nugget effect.

1.5.3.6. *Improving the convergence properties of the estimator*

This section discusses issues arising in situations where the covariance component functions $\mathbf{C}_i$ decay insufficiently fast for Conditions 1.5.3 and 1.5.5 to hold for all choices of function $\nu(\rho)$. In those cases, certain choices of function $\nu(\rho)$ will produce estimators for which Theorems 1.5.6, 1.5.7 and 1.5.8 apply, while others will not. In particular, if Condition 1.5.4 is not satisfied, then Condition 1.5.3 is not satisfied when $\nu(\rho) \equiv 1$ and Theorems 1.5.6, 1.5.7 and 1.5.8 do not apply. Even if Condition 1.5.3 is satisfied, Condition 1.5.5 may not be satisfied, in which case Theorem 1.5.6 applies, but Theorem 1.5.7 does not. Thus consistent estimation of the covariogram would be possible if the mean of the process were known, but perhaps impossible otherwise. It is obvious that for any covariance components $\mathbf{C}_i$, one can always select a function $\nu(\rho)$ such that both Conditions 1.5.3 and 1.5.5 are satisfied. It is not entirely clear how a particular choice of the function $\nu(\rho)$ affects the properties of the estimator.

The remainder of this section considers a particular case where Condition 1.5.4 is satisfied, but Condition 1.5.5 with $\nu(\rho) \equiv 1$ is not. In this case, if one were to attempt to estimate the covariogram function while at the same time estimating the unknown mean of the process, one might consider selecting a function $\nu(\rho)$ decaying sufficiently fast to guarantee that Condition 1.5.5 is satisfied.

The limiting behaviour of the resulting estimator $\hat{\theta}_V$ is determined by the matrices A, M and B given in Theorem 1.5.6. The choice of the function $\nu(\rho)$ affects the matrices A, M and B. It will be shown that it is possible to select $\nu(\rho)$ in such a way as to obtain an estimator whose limiting behaviour is arbitrarily close, in a certain sense, to the behaviour of the estimator obtained with $\nu(\rho) \equiv 1$ applied to a process with known mean. With this choice of $\nu(\rho)$, Theorems 1.5.7 and 1.5.8 will in fact apply while the matrices A, M and B can be made arbitrarily close to those obtained with $\nu(\rho) \equiv 1$.

In particular, the following family of functions $v(x_1, x_2)$ (defining $\nu(\rho)$ according to (1.5.73)) will be considered

$$v_R(x_1, x_2) = \nu_R(\| x_1 - x_2 \|) = \begin{cases} 1 & \text{if } \| x_1 - x_2 \| \leq R \\ g(\| x_1 - x_2 \|) & \text{otherwise} \end{cases} \tag{1.5.110}$$

where $R \geq 0$ and $0 < g(\rho) \leq 1$ is a non-increasing continuous function with $g(R) = 1$. Given any set of covariance components $\mathbf{C}_i$ it is possible to choose $g(\rho)$ such that the functions $\mathbf{C}_i(\rho)\nu_R(\rho)$ satisfy (1.5.72) with $p = 1$ for any positive R. One such choice would be $g(\rho) = \exp(R - \rho)$. This will guarantee that Theorem 1.5.7 applies and (1.5.93) holds. The limiting matrices A, M and B of Theorem 1.5.7 will now depend on R

$$A_R(i, j) = \alpha_G \int_0^\infty \mathbf{C}_i(\rho)\mathbf{C}_j(\rho)\nu_R(\rho)\rho^{d-1}d\rho \tag{1.5.111}$$

$$M_R(i) = \alpha_G \int_0^\infty \mathbf{C}_i(\rho)\mathbf{C}_Y(\rho)\nu_R(\rho)\rho^{d-1}d\rho \tag{1.5.112}$$

and

$$B_R(i,j) = \alpha_G \int_0^\infty H_{R,i}(\rho)H_{R,j}(\rho)\rho^{d-1}d\rho \tag{1.5.113}$$

where

$$\begin{aligned} H_{R,i}(\| \xi - \eta \|) = h_{R,i}(\xi,\eta) &= \int_{\mathbf{R}^d} C_i(\xi,\lambda)C_Y(\lambda,\eta)v_R(\xi,\lambda)d\lambda \\ &= \int_{\mathbf{R}^d} \psi_{R,i}(\xi-\eta-\lambda)\psi_Y(\lambda)d\lambda = \zeta_{R,i}(\xi-\eta) \end{aligned} \tag{1.5.114}$$

where $\zeta_{R,i}$ is the convolution of the functions $\psi_{R,i}$ and ψ_Y defined by

$$\psi_{R,i}(\xi) = C_i(\xi,0)v_R(\xi,0)$$

and

$$\psi_Y(\xi) = C_Y(\xi,0).$$

Finally, let the matrix E_R be given by

$$E_R = A_R^{-1}\,\mathrm{diag}(B_R)A_R^{-1} \tag{1.5.115}$$

while the vector θ_R is given by

$$\theta_R = A_R^{-1}M_R. \tag{1.5.116}$$

The following result holds

Theorem 1.5.9. *If Conditions 1.5.2 and 1.5.4 hold and for all R the functions $\mathbf{C}_i(\rho)\nu_R(\rho)$ satisfy Condition 1.5.5, then the following limits hold*

$$\begin{aligned}
\lim_{R\to\infty} A_R &= A \\
\lim_{R\to\infty} M_R &= M \\
\lim_{R\to\infty} B_R &= B \\
\lim_{R\to\infty} \theta_R &= \theta \\
\lim_{R\to\infty} E_R &= E
\end{aligned} \tag{1.5.117}$$

where A, B, M, θ and E are defined in (1.5.79), (1.5.80), (1.5.81), (1.5.82) and (1.5.83), respectively.

This result means that even if the components $\mathbf{C}_i$ do not decay sufficiently fast to guarantee the same behaviour of the estimator $\hat{\theta}$ for the case of known β as for the case of unknown β to be estimated, a suitable choice of inner product yields an estimator $\hat{\theta}_V$ with limiting properties under unknown β arbitrarily close to those in the case of known β.

1.5.4. Mixed in-fill and expanding-domain asymptotics with the uniform random sampling sequence

The asymptotic context considered in this section may be viewed as a special case of the sampling configuration of Hall and Patil (1994). Throughout this section it will be assumed that the process Y is Gaussian with zero mean. To consider a situation where both the number of observations and the domain size vary, one considers the sequence $\{(D_m, \{x_{m,n}\}_{n=1}^{\infty}, f_m)\}_{m=1}^{\infty}$ of expanding domains, in-fill sampling sequences and their sampling densities. Given the number of

observations n on the domain D_m, one obtains the estimator $\hat{\theta}_{m,n}$. If the true covariance function is in the set of valid covariance functions of the form (1.4.3), the estimator $\hat{\theta}_{m,n}$ is unbiased for θ_Y, otherwise Theorem 1.5.2 gives, for any fixed m, the expression for $\theta_m = \lim_{n\to\infty} \theta_{m,n}(X)$ where $\theta_{m,n}(X) = \mathbf{E}[\hat{\theta}_{m,n}|X]$ and where X denotes the (possibly random) in-fill sampling sequence on D_m. Theorem 1.5.5 shows that the stochastic convergence rate for this limit in the case of the sampling sequence described in Theorem 1.5.1 is $\mathrm{O_p}(n^{-1/2})$. Theorem 1.5.2 also shows that for a fixed m, $\mathrm{var}(\hat{\theta}_{m,n}(i)|X) \leq c_m(i), 1 \leq i \leq q$, that is, the variance is bounded from above by positive constants. Theorem 1.5.6 shows that $c_m = \mathrm{O}(r_m^{-d})$ for some constant c, and that the limit $\theta = \lim_{m\to\infty} \theta_m$ exists. The goal of this section is to construct a subsequence $n(m)$ such that in some sense $\hat{\theta}_{m,n(m)} \to \theta$ as $m \to \infty$. The mode of convergence considered will be convergence in probability. Naturally, the dependence of $n(m)$ on m will be affected by the rate of growth of r_m and the in-fill sampling sequences $\{x_{m,n}\}_{n=1}^{\infty}$. The case considered here will be that of the sampling sequence described in Theorem 1.5.1. Let $\{(D_m, \{x_{m,n}\}_{n=1}^{\infty}, f_m)\}_{m=1}^{\infty}$ be a sequence of expanding domains with the in-fill sampling sequences $\{x_{m,n}\}_{n=1}^{\infty}$ as in Theorem 1.5.1, let $\mathbf{X}_m = \{X_{m,n}\}_{n=1}^{\infty}$ be the random processes that generate them and let $\mathbf{X} = \{\mathbf{X}_m\}_{m=1}^{\infty}$. Let $\hat{\theta}_{m,n(m)}(\omega, \mathbf{X})$ be a random variable defined on the product space $\Omega_Y \times \Omega_{\mathbf{X}}$ of the probability space Ω_Y of $Y(\omega)(x), x \in \mathbf{R}^d$ and the probability space Ω_X of $\mathbf{X}$, given by

$$\begin{aligned}&\hat{\theta}_{m,n(m)}(\omega, \mathbf{X}) \\ &= [\mathrm{tr}(K_{i,m,n(m)}(\mathbf{X})K_{j,m,n(m)}(\mathbf{X}))]^{-1}[Y_{m,n(m)}(\omega, \mathbf{X})'K_{i,m,n(m)}(\mathbf{X})Y_{m,n(m)}(\omega, \mathbf{X})]\end{aligned} \tag{1.5.118}$$

where $K_{i,m,n(m)}(\mathbf{X})$ is an $n(m) \times n(m)$ matrix whose (l_1, l_2) element is the random variable $C_i(X_{m,l_1}, X_{m,l_2}), 1 \leq l_1, l_2 \leq n(m)$ and $Y_{m,n(m)}(\omega, \mathbf{X})$ is an $n(m) \times 1$

vector whose l-th element is the random variable $Y(\omega)(X_{m,l}), 1 \leq l \leq n(m)$. The random processes $\mathbf{X}$ and $Y(\omega)(x), x \in \mathbf{R}^d$ are assumed independent.

Theorem 1.5.10. *If* $\lim_{m\to\infty} r_m^{3d} n(m)^{-1} = 0$, *then*

$$\hat{\theta}_{m,n(m)}(\omega, \mathbf{X}) - \theta \to_{p(\omega,\mathbf{X})} \mathbf{0}$$

as $m \to \infty$.

The required rate of increase in $n(m)$ as r_m increases seems rather high. The results presented here are meant to illustrate the principle rather than to provide guidance in optimising the sampling configuration. In practice, the sampling configuration will most often not be random. In those cases it should be possible to come up with configurations requiring fewer sampling points.

1.6. Surface elevation data

In this section a class of models suggested by Shapiro and Botha (1991) is used to model the data of Davis (1973). The data consists of $n = 52$ measurements of surface elevation and it was collected on a square. Thus the domain is two-dimensional. The class of models suggested in Shapiro and Botha (1991) are of the form

$$\mathbf{C}_\theta(\rho) = \sum_{i=1}^{q} \theta(i) J_0(\lambda_i \rho) \tag{1.6.1}$$

where J_0 is the Bessel function of order zero of the first kind and the λ_i are fixed positive numbers, while the $\theta(i)$ are positive numbers to be estimated. Thus the model is additive, hence there are no difficulties in applying the estimator (1.4.19) to this model for a finite sample. However, the covariance component functions in the model decay very slowly (on the order of $\rho^{-1/2}$). This is insufficient to satisfy Condition 1.5.3 with $\nu(\rho) \equiv 1$, which was seen to complicate asymptotic

results (Theorems 1.5.6, 1.5.7 and 1.5.8 do not apply with $\nu(\rho) \equiv 1$, where ν is defined in (1.5.73)). In order to illustrate how a different choice of inner product may be used to obtain an estimator to which the asymptotic results of Theorems 1.5.6, 1.5.7 and 1.5.8 apply, the estimator (1.4.21) for non-constant V will also be computed. The matrix V_n of (1.4.21) will be computed using function $\nu(\rho)$ defined by (1.5.110) where the function g is given by $g(\rho) = \exp(R-\rho)$. The value of $R = 4$ will be used, which roughly equals one half of the maximal distance found in the dataset. The particular model (1.6.1) considered is defined by $q = 4$, with $\lambda_1 = 1$, $\lambda_2 = 2$, $\lambda_3 = 3$ and $\lambda_4 = 4$. Criteria for choosing the model parameters will not be discussed here. They are considered in greater detail by Powojowski (1999). Since the mean of the sampled process Y is not known, it needs to be modelled. Two separate models are used: one in which Y has a constant unknown mean, leading to a regressor matrix X_A, and another allowing for an arbitrary linear trend over the sampled square, yielding the regressor matrix X_B. The resulting estimates are given in Table 1.1. Figures 1.1 - 1.4 show the estimated covariograms. In all figures, the products of residuals are marked as individual points with coordinates $(\| x_k - x_l \|, e(k)e(l))$, where $1 \leq k \leq 52$, $1 \leq l \leq 52$ and $e(k) = Y(k) - \hat{Y}(k)$ is the residual computed as in (1.4.18). The estimated models themselves are plotted in Figures 1.1 and 1.2 along with the products of residuals. They are the curves $(\rho, \mathbf{C}_{\hat{\theta}}(\rho))$ and $(\rho, \mathbf{C}_{\hat{\theta}_V}(\rho))$. While the plots seem to indicate the estimated model follows the data fairly well, Figures 1.1 and 1.2 may be misleading. Let P denote the projection in (1.4.18). The mean of $e(k)e(l)$ is $(PK_YP)(k,l)$, rather than $K_Y(k,l)$, which may be estimated by $(PK_{\hat{\theta}}P)(k,l)$ rather than by $K_{\hat{\theta}}(k,l)$. It may therefore be informative to plot the points $(\| x_k - x_l \|, (PK_{\hat{\theta}}P)(k,l))$ along with the products of residuals. The resulting plots are shown in Figures 1.3 and 1.4. This is done only for the

model	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
constant mean	3908.54	638.644	219.84	0.6702
constant mean (non-constant V)	3892.6	383.423	73.1235	6.98203
linear trend	1123.54	359.73	106.796	49.7274
linear trend (non-constant V)	1173.14	388.69	96.827	63.0404

Table 1.1: *Estimated coefficients for the two mean models.*

estimator (1.4.19), where the matrix V is constant. One observes that in this case the message from the plots of fitted covariograms (Figures 1.1 and 1.2) and the plots of estimated covariances of residuals (Figures 1.3 and 1.4) are similar. One also observes that given a model for the mean, the different choices of matrix V lead to estimated covariograms which are very similar.

Comparing the results of this section with models previously fitted to the same data (e.g. Ripley, 1988, Wackernagel, 1995), one observes two main differences. Firstly, the models for the mean of the process used by those authors are linear trends or quadratic surfaces. It seems that the constant mean model was not attempted. No reasons are given for this omission, but it appears that the products of the residuals from the constant mean model are not very well estimated by a covariance model which is positive everywhere, such as the exponential or the Gaussian model used by the authors. This leads to the other major difference, namely the presence of negative covariances in the estimated model. Visual examination of Figures 1.1 - 1.4 suggests that negative covariances are plausible. The model given in (1.6.1) is capable of capturing negative covariances, while the exponential or the Gaussian model is not.

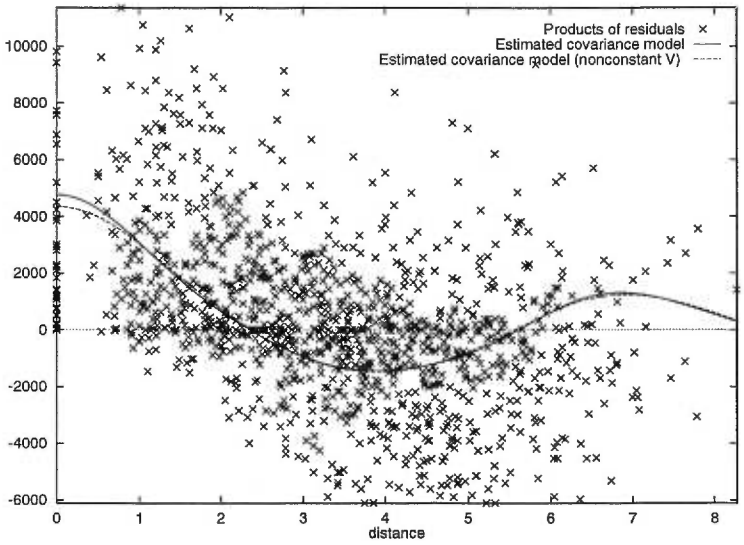

Figure 1.1: *Products of residuals and estimated covariance function, constant mean model.*

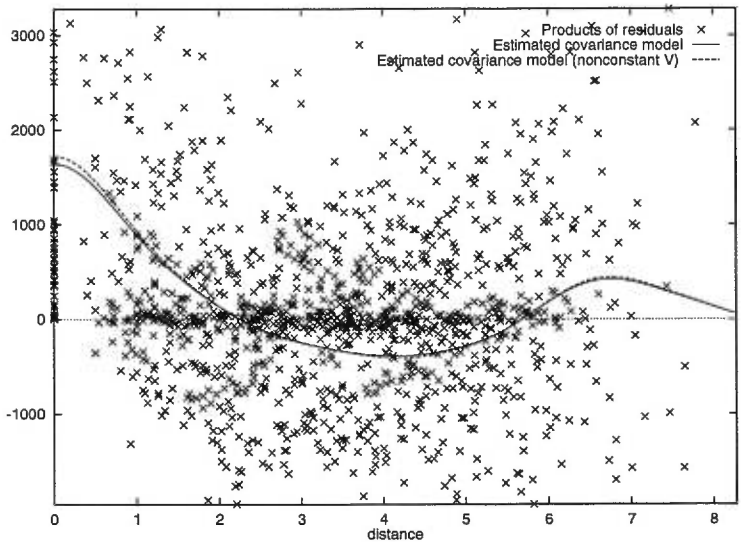

Figure 1.2: *Products of residuals and estimated covariance function, linear trend mean model.*

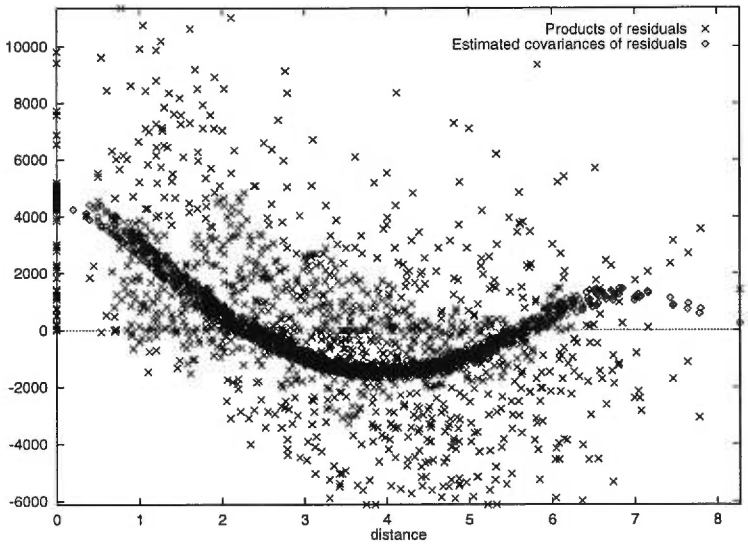

Figure 1.3: *Products of residuals and estimated covariances of residuals, constant mean model.*

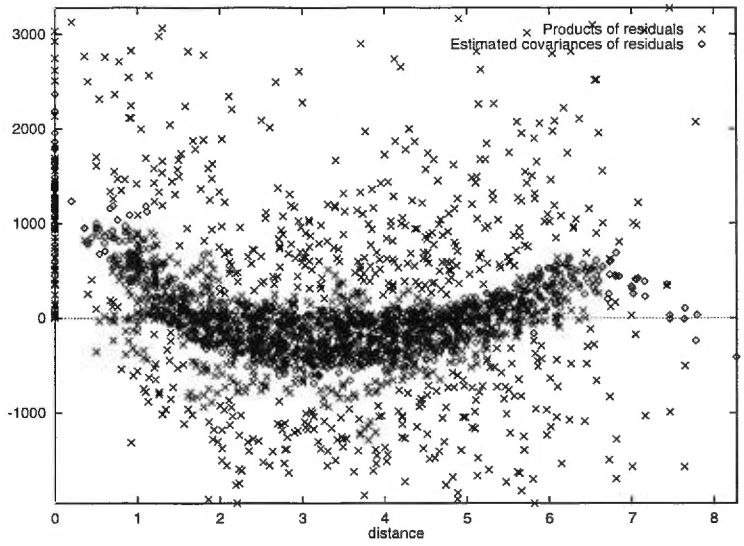

Figure 1.4: *Products of residuals and estimated covariances of residuals, linear trend mean model.*

1.7. Conclusion

The paper proposes a new approach to the problem of estimating the covariance function of a stochastic process. The approach combines two main ideas: using additive models and estimation based on orthogonal projections. The approach may be viewed as a particular case of MINQUE estimation. The case is presented that this approach offers many distinct advantages over traditional approaches involving the empirical (co)variogram, as well as over general MINQUE estimation.

In comparison with traditional methods the need to estimate the empirical (co)variogram is eliminated and so is the arbitrariness of the bin selection. Since the empirical (co)variograms are meaningless for non-stationary processes and hard to compute for non-isotropic processes, the approach presented is more generally applicable than the traditional procedures. The estimator by projection is unbiased if the mean of the process needs to be estimated from the data (assuming that the model for the mean is correct and the true covariance function is in the class of models considered). The mean and trends in the mean can be estimated simultaneously without much complication and usually without compromising the estimator's properties. The estimation procedure involves only linear algebra, and thus all problems, both theoretical and practical, associated with non-linear, non-quadratic optimisation are avoided. The properties of the estimator can be more easily understood than is the case in the traditional approach.

In contrast with the general MINQUE estimation, (which is also unbiased and requires only linear algebra to compute) the stability and asymptotic properties of the projection estimator are not dependent on the relationship between the initial guess K_0 required by MINQUE and the true covariance matrix K_Y. The

linear algebra computations required for the projection estimator involve inverses of much smaller matrices than those required for any other MINQUE estimator.

It is shown that in the in-fill asymptotic context (sampling of finite domain) it is generally impossible to estimate the covariogram consistently. An upper bound may, however, be obtained for the asymptotic variance of the estimator in this context. Furthermore, it is shown that for isotropic processes this upper bound can be made arbitrarily small if the sampled domain is sufficiently large. Some additional results are given for the case where the process is observed at points resulting from uniform random sampling of the domain.

The estimator is illustrated using a data set of Davis (1973), where the model used is that proposed by Shapiro and Botha (1991) (who use a different estimation procedure).

In order to apply the projection estimator in practice, an adequate class of additive models is required. One such class is the class of Shapiro and Botha (1991), as seen in Section 1.6. Other flexible classes of models will be explored in a separate study.

REFERENCES

BARTLETT, M.S. (1964). Spectral analysis of two-dimensional point processes. *Biometrika* **44** 299-311.

BROWN, K.G. (1978). Estimation of variance components using residuals. *Journal of the American Statistical Association* **73** 141 - 146.

CRESSIE, N.A.C. (1993). *Statistics for Spatial Data.* Revised edition. John Wiley & Sons.

CRESSIE, N.A.C. and GRONDONA, M. O. (1992). A comparison of variogram estimation with covariogram estimation. In *The Art of Statistical Science*, edited by K.V. Mardia, Wiley, Chichester, 191-208.

DAVIS, J.C. (1973). *Statistics and Data Analysis in Geology.* John Wiley & Sons.

DIGGLE, P.J. (1983). *Statistical Analysis of Spatial Point Patterns.* Academic Press.

HALL, P. and PATIL, P. (1994). Properties of nonparametric estimators of autocovariance for stationary random field. *Probability Theory and Related Fields* **99** 399-424.

KITANIDIS, P. K. (1985). Minimum variance unbiased quadratic estimation of covariances of regionalized variables. *Journal of the International Association for Mathematical Geology* **17** 195-208.

LAHIRI, S. N. (1996). On inconsistency of estimators under infill asymptotics for spatial data. *Sankhya* Ser. A, **58** 403-417.

LAHIRI, S. N., KAISER, M. S. , CRESSIE, N. and HSU, N. (1999). Prediction of spatial cumulative distribution functions using subsampling. *Journal of the American Statistical Association* **94** 86-97.

MATHERON, G. (1965). *Les variables régionalisées et leur estimation.* Masson et Cie.

PARTHASARATY, K. (1967). *Probability Measures on Metric Spaces.* Academic Press.

POWOJOWSKI, M. (1999). Model Selection in Covariogram Estimation. In preparation.

RAO, C.R. (1971). Estimation of variance and covariance components - MINQUE theory. *Journal of Multivariate Analysis* **1** 257 - 275.

RAO, C.R. and KLEFFE, J. (1988). *Estimation of Variance Components and Applications.* North-Holland.

RIPLEY, B.D. (1988). *Statistical Inference for Spatial Processes.* Cambridge University Press.

SERFLING, R. J (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons.

SHAPIRO, A. and BOTHA, J.D. (1991). Variogram fitting with a general class of conditionally nonnegative definite functions. *Computational Statistics and Data Analysis* **11** 87-96.

STEIN, M. L. (1987). Minimum norm quadratic estimation of spatial variograms. *Journal of the American Statistical Association* **82** 765-772.

STEIN, M. L. (1989). Asymptotic distribution of minimum norm quadratic estimators of the covariance function of a Gaussian random field. *The Annals of Statistics* **17** 980-1000.

STEIN, M. L. and HANDCOCK, M. S. (1989). Some asymptotic properties of kriging when the covariance function is misspecified. *Journal of the International Association for Mathematical Geology* **21** 839-861.

VERDOOREN, L.R. (1988). Least squares estimators and non-negative estimators of variance components. *Communications in Statistics A - Theory Methods* **17** 1287 - 1309.

WACKERNAGEL, H. (1995). *Multivariate geostatistics.* Springer-Verlag.

Chapitre 2

ESTIMATION OF THE COVARIOGRAM OF AN ISOTROPIC PROCESS THROUGH SPECTRAL COMPONENT ADDITIVE MODELS

2.1. Abstract

The main idea explored is that of spectral component additive models for the covariance function of an isotropic random process. A class of such models is proposed and it is shown to satisfy the conditions required by the projection estimation methods of Powojowski (1999a). The application of projection estimation methods is discussed and it is shown that when applied to spectral component additive models, the moments of the estimator have convenient expressions in terms of the spectral density of the covariance function of the process. It is also shown that the class of spectral component additive models is dense in the set of valid covariance functions and hence approximately unbiased spectral component additive model exists for any covariance function. Theoretical results are supported with simulation studies showing the approximate lack of bias of the proposed method. In addition, it is seen how an estimate of the spectral density of the covariance function may be obtained.

2.2. INTRODUCTION

For a random process $Y(x), x \in D$, where D is a subset of a d-dimensional Euclidean space, the covariogram is defined as $C_Y(x_1, x_2) = \text{cov}(Y(x_1), Y(x_2))$. A common problem in geostatistics is one of estimating the function C_Y based on one realisation of the process Y observed at a finite number of locations $x_1, x_2, \ldots, x_n$ in D. It is important to note that the knowledge of function values $C(x_1, x_2)$ for arbitrary $(x_1, x_2) \in D^2$ is required, and not simply the covariances of Y at lags observed in the sample. The fact of observing only one realisation forces one to make certain assumptions about the process Y, which translate into restrictions on the form of C_Y. There also exist theoretical reasons for restricting the function families considered. The covariogram has to be a positive definite function. Further restrictions may be desirable. In a typical covariogram estimation problem it is supposed that the observed process Y follows the model

$$Y = X\beta + \eta.$$

The known regressor X contains terms corresponding to the mean of the process and any trend that is allowed for, while the parameter β is unknown and the random term η is assumed to have zero mean and an unknown covariogram C_Y. The process η will be assumed isotropic (and hence second-order stationary), requiring $C_Y(x_1, x_2)$ to depend only on $\| x_1 - x_2 \|$. A covariance model C_θ, known up to the value of a finite-dimensional parameter θ, to be estimated, is to be fitted to the observed data.

The traditional approaches to the modelling and estimation of the covariogram function are described in Cressie (1993). Typically, the empirical covariogram or variogram (the variogram of the process Y is defined as $\gamma(x_1, x_2) = (1/2)\,\text{var}(Y(x_1) - Y(x_2))$) is computed and a parametric model is selected from

a small set of commonly used curves. The parametric curve is then found by some visual or optimisation technique which makes it pass close to the points of the empirical (co)variogram. Usually the parametric curve to be fitted does not depend linearly on the parameters which need to be estimated, which leads to nonlinear optimisation problems. An alternative set of models was proposed by Shapiro and Botha (1991). The models they propose are additive, that is they are linear combinations of fully specified models, in this case Bessel functions. The method of estimation they use is the traditional approach of fitting the curve to the empirical (co)variogram.

This paper proposes a new, broad class of additive models motivated by the spectral representation of the covariance function of an isotropic random process, henceforth referred to as spectral component additive models. Furthermore, the estimation techniques described by Powojowski (1999a) are applied to this new class of models. The techniques are based on orthogonal projections of products of residuals onto a linear space spanned by a finite set of valid covariance models. To apply these techniques successfully, it is necessary to provide an adequate class of additive covariance models. This paper demonstrates that the proposed class of spectral component additive models is adequate under fairly broad circumstances, namely, in the situation where the process η is isotropic and possesses a piecewise continuous spectral density. In addition it is shown that projection estimators applied to spectral component additive models reveal a theoretical connexion with the spectral density of the covariance function of the process.

The paper is organised as follows: after defining the notation used throughout the paper, the spectral representation of the covariance function of an isotropic function is discussed to motivate the subsequent introduction of the spectral component additive model. Basic properties of the model are then established and

the application of projection estimation methods is discussed. Connexions with the spectral theory of covariance functions are explored and they are subsequently used to show that the class of spectral component additive models is dense in the set of valid covariance functions possessing a spectral density. Finally, the model is illustrated in a study comparing it to commonly used covariance models.

2.3. NOTATION

To avoid confusion which might arise due to the frequent occurrence of multiple subscripts, the following notation will be used throughout the paper: if A is a matrix, its entries will be denoted by $A(i,j)$, while $A_{i,j}$ may denote a matrix from some (doubly) indexed set of matrices. Similarly, if θ is a vector, its components will be denoted by $\theta(i)$, while θ_i may denote a vector from some indexed set of vectors.

In the most general setting, one considers a random process Y on the domain D, a subset of a d-dimensional Euclidean space. The process Y is observed at n locations $\{x_i\}_{i=1}^n$, $x_i \in D$. Let $Y_n = (Y(x_1), \ldots, Y(x_n))'$ and $Y_n(i) = Y(x_i), 1 \leq i \leq n$. It will be further assumed that

$$Y_n = X_n\beta + \eta_n \tag{2.3.1}$$

where $\eta_n = (\eta(x_1), \ldots, \eta(x_n))'$ and η is an isotropic random process with $\mathbf{E}[\eta] = 0$. It will be assumed that X_n has p columns corresponding to different regression terms. Thus $X_n(l,k) = r_k(x_l)$, $1 \leq k \leq p$, $1 \leq l \leq n$, where x_l is the l-th location in the sample and r_k is a continuous function defined on D and it is the k-th regression term in the mean of Y. If present, the term $r_1 \equiv 1$ corresponds to the (non-zero) constant term in the mean of Y. The term $r_k(x_l) = x_l(1)$, where $x_l(1)$ is the first component of the d-dimensional vector $x_l \in D$, would correspond to a linear trend in the mean of $Y(x)$ in the direction of the first component of

x. The matrix X_n will always be known, while the $p \times 1$ vector β may have to be estimated. The function $C_Y(x_1, x_2) = \text{cov}(Y(x_1), Y(x_2)) = \text{cov}(\eta(x_1), \eta(x_2))$ is called the covariance function of the process Y (and of the zero-mean process η). Let $K_{Y,n} = \text{var}(Y_n)$. Thus $K_{Y,n}$ is a symmetric matrix whose entries are $K_{Y,n}(i,j) = C_Y(x_i, x_j)$. If C_θ or C_i is a given covariance function model, one defines the symmetric matrix $K_{\theta,n}$ or $K_{i,n}$ in a similar way, by putting $K_{\theta,n}(k,j) = C_\theta(x_k, x_j)$ or $K_{i,n}(k,j) = C_i(x_k, x_j)$. Thus $K_{\theta,n}$ is a fixed matrix depending only on the model C_θ and on the set of locations $\{x_i\}_{i=1}^n$, $x_i \in D$.

The model C_θ will be called additive if it is of the form

$$C_\theta = \sum_{i=1}^{q} \theta(i) C_i, \tag{2.3.2}$$

where the components C_i are fully specified valid covariance functions and the only parameters to be estimated are the $\theta(i)$. Throughout the paper the components C_i as well as C_Y will be assumed continuous and isotropic. The process of interest will also be assumed isotropic in $\mathbf{R}^d$. An isotropic process is stationary and it will be convenient at times to switch between the covariance functions $C_i(x_1, x_2)$, $C_Y(x_1, x_2)$, $C_\theta(x_1, x_2)$ and the isotropic versions of these functions $\mathbf{C}_i(\rho)$, $\mathbf{C}_Y(\rho)$ and $\mathbf{C}_\theta(\rho)$ where $\rho = \| x_1 - x_2 \|$.

Two intervals $[a_1, b_1]$ and $[a_2, b_2]$ will be called non-overlapping if their intersection is at most one point.

2.4. The covariogram and the spectral density of an isotropic stochastic process

This section reviews standard results concerning the spectral representation of an isotropic covariance function. The covariance function of an isotropic process

in $\mathbf{R}^d$ has the form (Schoenberg, 1938)

$$\mathbf{C}_Y(\rho) = \rho^{(2-d)/2} \int_0^\infty \lambda^{(2-d)/2} J_{(d-2)/2}(\lambda\rho) dG_Y(\lambda) \tag{2.4.1}$$

where $G_Y(\lambda)$ is a bounded non-decreasing function and J_ν is the Bessel function of the first kind and ν-th order. It will be assumed that $G_Y(\lambda)$ is continuous and piecewise differentiable with the derivative $g_Y(\lambda)$. Thus $g_Y(\lambda)$ is nonnegative, piecewise continuous and summable. Putting $\Psi_Y(\lambda) = \lambda^{-d/2} g_Y(\lambda)$ one obtains

$$\mathbf{C}_Y(\rho) = \rho^{(2-d)/2} \int_0^\infty \lambda \Psi_Y(\lambda) J_{(d-2)/2}(\lambda\rho) d\lambda \tag{2.4.2}$$

with $\lambda^{d/2}\Psi_Y(\lambda)$ nonnegative and summable. The function Ψ_Y will be called the spectral density of the covariance function $\mathbf{C}_Y$. If $\lambda^{1/2}\Psi(\lambda)$ is piecewise continuous and absolutely summable, the function

$$\Lambda(\rho) = \mathcal{H}_\nu(\Psi)(\rho) = \int_0^\infty \lambda J_\nu(\lambda\rho)\Psi(\lambda) d\lambda \tag{2.4.3}$$

will be called the Hankel transform of the function Ψ of order ν. Details may be found in Sneddon (1972). It thus follows that $\Lambda(\rho) = \rho^{(d-2)/2}\mathbf{C}_Y(\rho)$ is the Hankel transform of order $(d-2)/2$ of the function Ψ_Y. The Hankel inversion theorem states that under those assumptions

$$\int_0^\infty \rho J_\nu(\lambda\rho)\Lambda(\rho) d\rho = \frac{1}{2}(\Psi(\lambda-) + \Psi(\lambda+)). \tag{2.4.4}$$

(The notation $\Psi(\lambda-)$ denotes the left limit of Ψ at λ. Similarly $\Psi(\lambda+)$ denotes the right limit.) Moreover, Parseval's formula for Hankel transforms states that if $\lambda^{1/2}\Psi_i(\lambda), i = 1, 2$ are piecewise continuous and absolutely summable functions, then

$$\int_0^\infty \rho\Lambda_1(\rho)\Lambda_2(\rho) d\rho = \int_0^\infty \lambda\Psi_1(\lambda)\Psi_2(\lambda) d\lambda \tag{2.4.5}$$

and both integrals exist.

This paper explores the possibility of modelling and estimating the covariance function based on the expression (2.4.2). In addition to estimating the covariance function, this approach will be seen to provide information about the spectral properties of the process.

Before moving on, the general representation of the covariance function of a stationary (not necessarily isotropic) process is briefly discussed. It follows from Bochner's theorem that a covariance function C whose spectral density exists may be expressed as

$$C(x,0) = (2\pi)^{-d/2} \int_{\mathbf{R}^d} \exp(\mathbf{i}\langle x, \xi\rangle) f(\xi) d\xi \tag{2.4.6}$$

where $\langle , \rangle$ denotes the standard inner product (dot product) in $\mathbf{R}^d$ and f is a non-negative, integrable function. It is a well known fact that if C is a function of $\| x \|$, the function f is radially symmetric, and therefore putting $\rho = \| x \|$ one may define a function $w(\rho) = f(x)$. With this notation, the following general identity is obtained through change of variable

$$(2\pi)^{-d/2} \int_{\mathbf{R}^d} \exp(\mathbf{i}\langle x, \xi\rangle) f(\xi) d\xi \;=\; \rho^{(2-d)/2} \int_0^\infty \lambda^{d/2} w(\lambda) J_{(d-2)/2}(\lambda\rho) d\lambda \tag{2.4.7}$$

(see, for example, Sneddon, 1972). Under isotropy (2.4.6) becomes (see, for example, Adler, 1980)

$$\begin{aligned} C(x,0) = \mathbf{C}(\rho) &= \rho^{(2-d)/2} \int_0^\infty \lambda^{d/2} w(\lambda) J_{(d-2)/2}(\lambda\rho) d\lambda \\ &= \rho^{(2-d)/2} \int_0^\infty \lambda \Psi(\lambda) J_{(d-2)/2}(\lambda\rho) d\lambda \end{aligned} \tag{2.4.8}$$

where $\Psi(\lambda) = \lambda^{(d-2)/2} w(\lambda)$, and Ψ is as in (2.4.2). The equation (2.4.8) connects the spectral representation of the covariance function of an isotropic process with its (more general) spectral representation resulting from the fact that it is also the covariance function of a second-order stationary process.

2.5. Model definition

The estimation of the covariance function requires a family of functions which can be parametrised with a finite-dimensional vector. This section defines the spectral component additive models, which will be the main focus in the remainder of the paper.

Shapiro and Botha (1991) suggested the following class of models

$$\mathbf{C}_\theta(\rho) = \sum_{i=1}^{q} \theta(i) J_{(d-2)/2}(\lambda_i \rho) \tag{2.5.1}$$

where the λ_i are fixed positive numbers and θ is a vector with non-negative entries to be estimated. This model, however, produces covariance functions which decay very slowly - on the order of $\rho^{-1/2}$ - which may be too slow for many applications. It is also seen in Powojowski(1999a) that slow decay rates complicate asymptotic results of the projection estimators considered in the next section.

The general expression (2.4.2) suggests the following class of models for the covariance function

$$\mathbf{C}_\theta(\rho) = \sum_{i=1}^{q} \theta(i) \mathbf{C}_i(\rho) \tag{2.5.2}$$

where

$$\mathbf{C}_i(\rho) = \rho^{(2-d)/2} \int_{a_i}^{b_i} \lambda J_{(d-2)/2}(\lambda\rho)\Psi(\lambda)d\lambda = \rho^{(2-d)/2} \int_0^{\infty} \lambda J_{(d-2)/2}(\lambda\rho)\Psi_i(\lambda)d\lambda \tag{2.5.3}$$

where Ψ is a fully specified, non-negative, bounded function and $\Psi_i(\lambda) = \Psi(\lambda) I_{[a_i,b_i]}(\lambda)$. Thus the components $\mathbf{C}_i(\rho)\rho^{(d-2)/2}$ have compactly supported spectral densities $\Psi(\lambda) I_{[a_i,b_i]}(\lambda)$.

Some basic properties of the component functions $\mathbf{C}_i$ will now be established.

Lemma 2.5.1. *If a_i and b_i are positive numbers with $a_i < b_i$, the function Ψ is continuous on the real line and $\mathbf{C}_i$ is defined by (2.5.3), then $\mathbf{C}_i$ is continuous, bounded and*

$$\int_0^\infty |\mathbf{C}_i(\rho)|^2 \rho^{d-1} d\rho < \infty. \tag{2.5.4}$$

Proof:

From the definition of the $\mathbf{C}_i$ one obtains

$$\int_0^\infty \Big(\mathbf{C}_i(\rho)\Big)^2 \rho^{d-1} d\rho = \int_0^\infty \rho \Big(\int_0^\infty \lambda J_{(d-2)/2}(\lambda\rho)\Psi_i(\lambda) d\lambda \Big)^2 d\rho$$
$$= \int_0^\infty \lambda (\Psi_i(\lambda))^2 d\lambda < \infty$$

since Parseval's identity may be used as the function $\lambda^{1/2}\Psi_i(\lambda)$ is piecewise continuous and absolutely summable.

If for $d > 1$ the function Ψ is taken to be

$$\Psi(\lambda) = \lambda^{(d-2)/2}, \tag{2.5.5}$$

the integral in (2.5.3) has a closed form, which results in the following covariance components

$$\mathbf{C}_i(\rho) = \rho^{(2-d)/2} \left(\frac{b_i^{d/2}}{\rho} J_{d/2}(b_i\rho) - \frac{a_i^{d/2}}{\rho} J_{d/2}(a_i\rho) \right). \tag{2.5.6}$$

Another reasonable choice would be $\Psi(\lambda) \equiv 1$, which coincides with (2.5.5) for $d = 2$. However, for $d > 2$ there may not be a convenient closed form for the resulting covariance components.

For the function Ψ defined by (2.5.5) it follows that

$$\mathbf{C}_i(\rho) = O(\rho^{-(d+1)/2}). \tag{2.5.7}$$

The formula (2.5.6) is a standard result and may be found in Sneddon (1972). To show (2.5.7) one applies (2.5.6) together with the well known fact (see for

example Sneddon, 1972) that $\lambda^{1/2}|J_\nu(\lambda)|$ is bounded on the positive real line. It follows from (2.5.7) that

$$\int_0^\infty \mathbf{C}_i(\rho)^2 \rho^{d-1} d\rho < \infty. \tag{2.5.8}$$

This last result plays a role in convergence arguments related to the projection estimator described in the next section. For details, see Powojowski (1999a).

In all situations considered in this paper the compact supports $[a_i, b_i]$ of the functions Ψ_i will be pairwise non-overlapping.

2.6. Spectral component estimation

The model components $\mathbf{C}_i$ defined in the previous section (with any choice of the Ψ_i) will be referred to as spectral covariance components, while the functions Ψ_i will be referred to as spectral components. The rationale behind these terms will soon become apparent.

2.6.1. Covariogram estimation through projections

The focus of this section is on the estimation of the parameters of the model (2.5.2). Given the class of models (2.5.2), with the components $\mathbf{C}_i$ fully specified and q fixed, the estimator described in Powojowski (1999a) may be used to find estimates of the $\theta(i)$. If the parameter β of (2.3.1) is known, it can be assumed to be zero without loss of generality. The $\hat{\theta}_n(i)$ are selected so as to minimise $\| Y_n Y_n' - \sum_{i=1}^q \hat{\theta}_n(i) K_{i,n} \|$ where the norm is the square root of the sum of squares of differences between the entries of the two matrices. The minimisation results in $\hat{\theta}_n$ given by

$$\hat{\theta}_n = [\mathrm{tr}(K_{i,n} K_{j,n})]^{-1} [Y_n' K_{i,n} Y_n] \tag{2.6.1}$$

where $\hat{\theta}_n = (\hat{\theta}_n(1), \ldots, \hat{\theta}_n(q))'$ and $K_{i,n}$ is the matrix whose (m,l) entry is $\mathbf{C}_i(\| x_m - x_l \|)$. (The notation $[\mathrm{tr}(K_{i,n}K_{j,n})]$ denotes a $q \times q$ matrix whose (i,j) - th entry is $\mathrm{tr}(K_{i,n}K_{j,n})$. Similarly, $[Y_n'K_{i,n}Y_n]$ denotes a $q \times 1$ vector.)

In the more general case of unknown β, one defines the residuals

$$e_n = (I_n - X_n(X_n'X_n)^{-1}X_n')Y_n = P_nY_n$$

where P_n is a projection matrix. The estimator (2.6.1) is then replaced by

$$\hat{\theta}_n = [\mathrm{tr}(U_{i,n}U_{j,n})]^{-1}[e_n'U_{i,n}e_n] \tag{2.6.2}$$

where

$$U_{i,n} = P_nK_{i,n}P_n.$$

In (2.6.2) the vector $\hat{\theta}_n$ minimises $\| e_ne_n' - \sum_{i=1}^q \hat{\theta}_n(i)U_{i,n} \|$.

Finally, the estimators (2.6.1) and (2.6.2) can be extended to the class of estimators

$$\hat{\theta}_{V,n} = [\mathrm{tr}((U_{i,n} * V_n)U_{j,n})]^{-1}[e_n'(U_{i,n} * V_n)e_n] \tag{2.6.3}$$

where the $n \times n$ symmetric matrix $V_n(l,m) = \nu(\| x_{n,l} - x_{n,m} \|)$, has entries given by a fully specified positive function ν, and $(A * B)(k,l) = A(k,l)B(k,l)$ is the Hadamard matrix product. The matrix V can be thought of as weight factors and the estimator (2.6.3) minimises a weighted sum of squares.

The properties of the estimators (2.6.1), (2.6.2) and (2.6.3) are described in Powojowski (1999a), along with relevant asymptotic settings. For the purpose of the current discussion, it will simply be assumed that for every positive integer n, the sampling configuration $\{x_{n,1}, \ldots, x_{n,n}\}$ is located on a compact domain D_n with Lebesgue measure $\mu(D_n) = r_n^d$, with $r_1 = 1$. It will be further assumed that the sequence $\{r_n\}_n^\infty$ is increasing and divergent. Under certain conditions

(discussed in detail in Powojowski, 1999a) concerning the sizes of the domains D_m and the sampling configuration $\{x_{n,1}, \ldots, x_{n,n}\}$, if

$$\int_0^\infty |\mathbf{C}_Y(\rho)| \rho^{d-1} d\rho < \infty, \text{ and} \tag{2.6.4}$$

$$\int_0^\infty |\mathbf{C}_i(\rho)|^2 \rho^{d-1} d\rho < \infty \tag{2.6.5}$$

the following holds:

Proposition 2.6.1. *The estimator (2.6.1) satisfies*

$$\hat{\theta}_n \to_p \theta = A^{-1}M \text{ as } n \to \infty \tag{2.6.6}$$

where A is a $q \times q$ matrix whose entries are given by

$$A(i,j) = \alpha_G \int_0^\infty \mathbf{C}_i(\rho)\mathbf{C}_j(\rho)\rho^{d-1} d\rho \tag{2.6.7}$$

and M is a $q \times 1$ matrix whose entries are given by

$$M(i) = \alpha_G \int_0^\infty \mathbf{C}_i(\rho)\mathbf{C}_Y(\rho)\rho^{d-1} d\rho \tag{2.6.8}$$

and the constant α_G equals $d\pi^{d/2}/\Gamma(d/2+1)$.

The condition in (2.6.4) will be assumed to hold. For the covariance components defined by (2.5.3) with (2.5.5), the condition in (2.6.5) holds by (2.5.8). (Incidentally, one observes that the components of the model described in (2.5.1) do not satisfy (2.6.5).) Moreover, one has

Proposition 2.6.2. *If the observed process Y is Gaussian, then*

$$\lim_{n\to\infty} r_n^d \operatorname{var}(\hat{\theta}_n) = 2A^{-1}BA^{-1} \tag{2.6.9}$$

where B is a $q \times q$ matrix whose entries are given by

$$B(i,j) = \alpha_G \int_0^\infty H_i(\rho)H_j(\rho)\rho^{d-1} d\rho \tag{2.6.10}$$

with α_G defined as in Proposition 2.6.1, and

$$H_i(\| \xi - \eta \|) = h_i(\xi, \eta) = \int_{\mathbf{R}^d} C_i(\xi, \lambda) C_Y(\lambda, \eta) d\lambda$$
$$= \int_{\mathbf{R}^d} \psi_i(\xi - \eta - \lambda) \psi_Y(\lambda) d\lambda = \zeta_i(\xi - \eta)$$

where ζ_i is the convolution of the functions ψ_i and ψ_Y defined by

$$\psi_i(\xi) = C_i(\xi, 0)$$

$$\psi_Y(\xi) = C_Y(\xi, 0).$$

The last proposition gives the rate at which the variance of the estimator converges to zero. A similar bound can be obtained for non-Gaussian process under additional assumptions. For details and for extensions to the more general estimators (2.6.2) and (2.6.3) the reader is referred to Powojowski (1999a).

2.6.2. Spectral additive components

This section examines the particular form assumed by the expressions in Propositions 2.6.1 and 2.6.2 for the model defined in (2.5.2).

Theorem 2.6.1. *Let $\mathbf{C}_\theta$ be as in (2.5.2) and let the intervals $[a_i, b_i]$ be non-overlapping. Then the matrices A, B of (2.6.7), (2.6.10) are diagonal and their entries are given by*

$$A(i,i) = \alpha_G \int_{a_i}^{b_i} \lambda (\Psi(\lambda))^2 d\lambda \tag{2.6.11}$$

and

$$B(i,i) = \alpha_G (2\pi)^d \int_{a_i}^{b_i} \lambda^{3-d} (\Psi(\lambda)\Psi_Y(\lambda))^2 d\lambda \tag{2.6.12}$$

while the entries of M of (2.6.8) are given by

$$M(i) = \alpha_G \int_{a_i}^{b_i} \lambda \Psi(\lambda) \Psi_Y(\lambda) d\lambda. \tag{2.6.13}$$

Thus the limiting expectation of the estimator is

$$\mathbf{E}[\hat{\theta}_n(i)] \to \frac{\int_{a_i}^{b_i} \lambda \Psi(\lambda) \Psi_Y(\lambda) d\lambda}{\int_{a_i}^{b_i} \lambda (\Psi(\lambda))^2 d\lambda}. \tag{2.6.14}$$

Since A and B are diagonal, the components of $\hat{\theta}_n$ are asymptotically uncorrelated if the process is Gaussian, by Proposition 2.6.2.

Proof:

The limit matrix A will be considered first. It follows that

$$\begin{aligned} A(i,j) &= \alpha_G \int_0^\infty \mathbf{C}_i(\rho) \mathbf{C}_j(\rho) \rho^{d-1} d\rho \\ &= \alpha_G \int_0^\infty \rho \int_0^\infty \lambda J_{(d-2)/2}(\lambda\rho) \Psi_i(\lambda) d\lambda \int_0^\infty \lambda J_{(d-2)/2}(\lambda\rho) \Psi_j(\lambda) d\lambda \, d\rho \\ &= \alpha_G \int_0^\infty \lambda \Psi_i(\lambda) \Psi_j(\lambda) d\lambda \end{aligned}$$

by Parseval's identity. If $i \neq j$ the supports of the functions Ψ_i and Ψ_j are non-overlapping and the integral equals zero. There is nothing substantially different about the entries of M.

For the limit matrix B one considers

$$B(i,j) = \alpha_G \int_0^\infty H_i(\rho) H_j(\rho) \rho^{d-1} d\rho$$

where

$$H_i(\| \xi - \eta \|) = \zeta_i(\xi - \eta) = (\psi_i * \psi_Y)(\xi - \eta)$$

and

$$\psi_i(\xi) = C_i(\xi, 0)$$

$$\psi_Y(\xi) = C_Y(\xi, 0).$$

The Fourier transform of a function f will be denoted by $\mathcal{F}(f)$. Let $f_i(x)$ and $f_Y(x)$ be such that $\psi_i = \mathcal{F}(f_i)$ and $\psi_Y = \mathcal{F}(f_Y)$. The existence of such representations, together with the fact that the functions $f_i(x)$ and $f_Y(x)$ are nonnegative for $x \in \mathbf{R}^d$, follows from (2.4.6). By the isotropy, the functions f_i and f_Y are also radially symmetric, that is they depend only on $\| x \|$. Moreover, by the convolution theorem,

$$(\psi_i * \psi_Y)(\xi) = (2\pi)^{d/2}\mathcal{F}(f_i(\xi)f_Y(\xi)) = (2\pi)^{d/2}\mathcal{F}(w_i(\| \xi \|)w_Y(\| \xi \|))$$

where $w_Y(\| \xi \|) \equiv f_Y(\xi)$ and $w_i(\| \xi \|) \equiv f_i(\xi)$. Therefore

$$\zeta_i(\xi) = \int_{\mathbf{R}^d} f_i(x)f_Y(x)\exp(\mathbf{i}\langle x, \xi\rangle)dx$$

and, by (2.4.7)

$$H_i(\rho) = (2\pi)^{d/2}\rho^{(2-d)/2}\int_0^\infty \lambda^{d/2}w_i(\lambda)w_Y(\lambda)J_{(d-2)/2}(\lambda\rho)d\lambda.$$

Therefore, by Parseval's identity

$$\begin{aligned}
&\int_0^\infty H_i(\rho)H_j(\rho)\rho^{d-1}d\rho \\
&\qquad = (2\pi)^d\int_0^\infty \rho\Bigg(\int_0^\infty \lambda(\lambda^{(d-2)/2}w_i(\lambda)w_Y(\lambda))J_{(d-2)/2}(\lambda\rho)d\lambda \\
&\qquad\qquad \int_0^\infty \lambda(\lambda^{(d-2)/2}w_j(\lambda)w_Y(\lambda))J_{(d-2)/2}(\lambda\rho)d\lambda\Bigg)d\rho \\
&\qquad\qquad = (2\pi)^d\int_0^\infty \lambda\lambda^{d-2}(w_Y(\lambda))^2w_i(\lambda)w_j(\lambda)d\lambda \\
&\qquad\qquad\qquad = \delta_{i,j}(2\pi)^d\int_0^\infty \lambda^{3-d}(\Psi_Y(\lambda)\Psi_i(\lambda))^2d\lambda
\end{aligned}$$

where $\delta_{i,j}$ equals one if $i = j$ and zero otherwise.

The requirement for the covariance function to be positive-definite imposes the constraint $\theta(i) \geq 0$ in (2.5.2). It follows from Theorem 2.6.1 that the limit $\theta = A^{-1}M$ in Proposition 2.6.1 has non-negative entries. This implies that from an asymptotic point of view the constraints $\theta(i) \geq 0$ do not pose any problem. It is, however, possible that in a finite sample context the constraints $\theta(i) \geq 0$ will have to be dealt with.

2.7. Approximation property of the spectral component models

In this section the relationship between the covariance functions of isotropic random fields and their spectral densities is applied to show that the models of the form (2.5.3) can be used to approximate an arbitrary covariance function. The following result shows that the class of models defined in (2.5.2) and (2.5.3) is dense in the set of isotropic covariance functions satisfying (2.6.4) and whose spectral density exists.

Theorem 2.7.1. *Let $\mathbf{C}_Y$ be an isotropic covariance function in $\mathbf{R}^d$ with a piecewise continuous spectral density and satisfying (2.6.4). Let Ψ be a non-negative, continuous, bounded function. Then for any $\epsilon > 0$ there exists an integer q, and positive numbers a_i, b_i, for $i = 1, \ldots, q$, such that the intervals $[a_i, b_i]$ are non-overlapping and the model defined by (2.5.2) and (2.5.3) satisfies*

$$\int_0^\infty \left(\mathbf{C}_Y(\rho) - \sum_{i=1}^{q} \theta(i) \mathbf{C}_i(\rho) \right)^2 \rho^{d-1} d\rho \leq \epsilon.$$

Proof:

By Parseval's formula the following holds

$$\int_0^\infty \left(\mathbf{C}_Y(\rho) - \sum_{i=1}^{q} \theta(i)\mathbf{C}_i(\rho) \right)^2 \rho^{d-1} d\rho$$

$$= \int_0^\infty \left(\Psi_Y(\lambda) - \sum_{i=1}^{q} \theta(i)\Psi_i(\lambda) \right)^2 \lambda d\lambda$$

$$= \sum_{i=1}^{q} \int_{a_i}^{b_i} (\Psi_Y(\lambda) - \theta(i)\Psi(\lambda))^2 \lambda d\lambda + \int_{S^c} (\Psi_Y(\lambda))^2 \lambda d\lambda. \quad (2.7.1)$$

The set S^c is the complement (with respect to the positive real line) of the maximal support S of the function $\sum_{i=1}^{q} \theta(i)\mathbf{C}_i$, which is the union of the intervals $[a_i, b_i]$. It is easily seen that the last integral can be made arbitrarily small by selecting $S = [0, R]$, since

$$\lim_{R \to \infty} \int_R^\infty (\Psi_Y(\lambda))^2 \lambda d\lambda = 0$$

by (2.6.4). It is also easy to see that the sum

$$\sum_{i=1}^{q} \int_{a_i}^{b_i} (\Psi_Y(\lambda) - \theta(i)\Psi(\lambda))^2 \lambda d\lambda$$

can be made arbitrarily small by making the intervals $[a_i, b_i]$ small.

Given an integer q, and non-overlapping intervals $[a_i, b_i]$, $i = 1, \ldots, q$, the quantity

$$\int_0^\infty \left(\mathbf{C}_Y(\rho) - \sum_{i=1}^{q} \theta(i)\mathbf{C}_i(\rho) \right)^2 \rho^{d-1} d\rho \quad (2.7.2)$$

is minimised by $\theta = A^{-1}M$, given in (2.6.6). This follows from elementary properties of orthogonal projections in inner product spaces. For details, the reader is referred to Powojowski (1999a). Hence, the asymptotic bias of the estimated covariance function $\sum_{i=1}^{q} \hat{\theta}(i)\mathbf{C}_i$ can be made arbitrarily small by selecting appropriate values of q and $[a_i, b_i], i = 1, \ldots, q$.

Moreover, (2.7.1) shows that the same vector $\theta = A^{-1}M$ minimises the quantity

$$\int_0^\infty \left(\Psi_Y(\lambda) - \sum_{i=1}^{q} \theta(i)\Psi_i(\lambda) \right)^2 \lambda d\lambda. \tag{2.7.3}$$

Hence, the best approximation of the true covariance function $\mathbf{C}_Y$ among covariance models of the form $\sum_{i=1}^{q} \theta(i)\mathbf{C}_i$ in the sense of the norm (2.7.2) is also the best approximation of the spectral density Ψ_Y among the models of the form $\sum_{i=1}^{q} \theta(i)\Psi_i$ in the sense of the norm (2.7.3).

2.8. Approximate lack of bias and the estimation of the spectral density

This section illustrates two aspects of covariogram estimation where the additive spectral component approach has a distinct advantage over traditional approaches. One is the model flexibility, or approximate lack of bias, the other is the estimation of the spectral density of the covariance function.

2.8.1. The bias of the spectral component additive models.

Traditionally, a parametric covariance model is selected from a small set of known positive-definite functions. Some estimation procedure is then used to determine the parameters. Among the most popular parametric families one finds the exponential, Gaussian, spherical and hole-effect models. These will be described shortly. It should be observed that each parametric family imposes rigid constraints on the form of the covariogram. Since often there are no good reasons for selecting one family over another, it seems likely that often models will be misspecified. For example, if the data were produced from the hole-effect model, but the exponential model is used for the covariogram, the resulting

estimated covariogram will not reflect the form of the true covariogram. In this numerical study, various traditional models are considered. A spectral component additive model is used to estimate the true covariogram. The parameter q and the intervals $[a_i, b_i], i = 1, \ldots, q$ are fixed. The spectral component model used here has seven components of the form (2.5.6). Four parametric families are considered: exponential, Gaussian, spherical and hole-effect. From each family, two models are specified. In each of the resulting eight cases it is assumed that the specified parametric model is the true covariogram $\mathbf{C}_Y$ and that the spectral component additive model $\mathbf{C}_\theta$ is used to estimate the covariogram. The domain where the process is observed is a square with side of length eight (thus the dimension d equals two). Eighty locations are initially drawn randomly on the domain and in each of the eight cases the process is observed at those locations. In each case, the mean of the covariogram estimate $\mathbf{C}_{\hat{\theta}}$ is computed from

$$\mathbf{E}[\mathbf{C}_{\hat{\theta}}] = \sum_{k=1}^{q} \mathbf{E}[\hat{\theta}(i)]\mathbf{C}_i$$

where from (2.6.1) one has

$$\mathbf{E}[\hat{\theta}] = [\mathrm{tr}(K_{i,n}K_{j,n})]^{-1}[K_{i,n}K_{Y,n}]. \tag{2.8.1}$$

In each case the mean of the estimated covariogram is compared visually to the true covariogram $\mathbf{C}_Y$. The departure between the two curves corresponds to the bias introduced due to model misspecification.

2.8.2. The spectral density of the covariance function

Expression (2.7.1) suggests that approximating the true covariance function $\mathbf{C}_Y$ through spectral components implies approximating the spectral density Ψ_Y of the covariance function through a sum of elements Ψ_i of (2.4.2) whose spectral supports are non-overlapping. In each of the eight cases considered, the true

density function Ψ_Y is visually compared to the mean $\mathbf{E}[\Psi_{\hat{\theta}}]$ of the spectral density of the estimate $\mathbf{C}_{\hat{\theta}}$ given by

$$\mathbf{E}[\Psi_{\hat{\theta}}] = \sum_{k=1}^{q} \mathbf{E}[\hat{\theta}(i)]\Psi_i$$

where $\mathbf{E}[\hat{\theta}]$ is given by (2.8.1).

2.8.3. The spectral component additive model

The additive spectral component model will be given by (2.5.2) with the components $\mathbf{C}_i$ given by

$$\mathbf{C}_i(\rho) = \rho^{(2-d)/2}\left(\frac{b_i^{d/2}}{\rho}J_{d/2}(b_i\rho) - \frac{a_i^{d/2}}{\rho}J_{d/2}(a_i\rho)\right) \tag{2.8.2}$$

with $d = 2$, $q = 7$ and $(a_1, b_1) = (0, 1), (a_2, b_2) = (1, 2), (a_3, b_3) = (2, 4), (a_4, b_4) = (4, 6), (a_5, b_5) = (6, 8), (a_6, b_6) = (8, 10), (a_7, b_7) = (10, 16)$. The components are scaled so that $\mathbf{C}_i(0) = 1, 1 \leq i \leq 7$ and they are shown in Figure 2.1.

The spectral density of the component $\mathbf{C}_i$ is $2/(b_i^2 - a_i^2)$ on the interval $[a_i, b_i]$ and zero elsewhere.

2.8.4. Parametric models

This section reviews the standard parametric models which will be compared to the mean of the estimated spectral component additive model given next.

2.8.4.1. *The exponential model*

The exponential model can be parametrised as follows (Cressie, 1993)

$$\mathbf{C}_{ex(c,a)}(\rho) = c\exp(-\rho/a) \tag{2.8.3}$$

and it is the Hankel transform of

$$\Psi_{ex(c,a)}(\lambda) = \frac{c}{a}\left(\lambda^2 + \frac{1}{a^2}\right)^{-3/2}. \tag{2.8.4}$$

Two exponential models are considered, $(c, a) = (1, 0.3)$ and $(c, a) = (1, 1)$. Comparisons of the true and modelled covariogram are shown in Figures 2.2 and 2.3, while comparisons of their spectral densities are shown in Figures 2.10 and 2.11.

2.8.4.2. *The Gaussian model*

The Gaussian model can be parametrised as

$$\mathbf{C}_{ga(c,a)}(\rho) = c\exp(-\rho^2/a) \tag{2.8.5}$$

and it is the Hankel transform of

$$\Psi_{ga(c,a)}(\lambda) = \frac{ca}{2}\exp(-a\lambda^2/4). \tag{2.8.6}$$

The Gaussian models considered are $(c, a) = (1, 0.2)$ and $(c, a) = (1, 2)$ and the resulting covariances are shown in Figures 2.4 and 2.5. Their spectral densities are compared in Figures 2.12 and 2.13.

2.8.4.3. *The spherical model*

The spherical model can be parametrised as

$$\mathbf{C}_{sp(c,a)}(\rho) = \begin{cases} c\left(1 + \frac{1}{2}(\frac{\rho}{a})^3 - \frac{3}{2}\frac{\rho}{a}\right) & \text{if } \rho \leq a \\ 0 & \text{otherwise.} \end{cases} \tag{2.8.7}$$

The spectral density has a rather complicated form and is evaluated from the definition through numerical integration instead. The spherical models considered are $(c, a) = (1, 1)$ and $(c, a) = (1, 3)$ and the results for the covariance functions are shown in Figures 2.6 and 2.7. Comparisons of the spectral densities are shown in Figures 2.14 and 2.15.

2.8.4.4. *The hole-effect model*

The hole-effect model can be parametrised as

$$\mathbf{C}_{ho(c,a)}(\rho) = ca\frac{\sin(\rho/a)}{\rho} \tag{2.8.8}$$

while the spectral density is given by

$$\Psi_{ho(c,a)}(\lambda) = ca\begin{cases} (\frac{1}{a^2} - \lambda^2)^{-1/2} & \text{if } 0 < \lambda < \frac{1}{a} \\ 0 & \text{if } \frac{1}{a} < \lambda. \end{cases} \tag{2.8.9}$$

It is seen that Ψ has a singularity at $1/a$. It is easy to show that this model does not satisfy the condition in (2.6.4), so Theorem 2.7.1 does not apply. It is, however included here because it is a well-known model. It will be seen that in the finite sample context this model can also be well approximated by the spectral component additive model. The hole-effect models considered are $(c, a) = (1, 0.1)$ and $(c, a) = (1, 0.5)$ and the resulting covariances are compared in Figures 2.8 and 2.9. Their spectral densities are compared in Figures 2.16 and 2.17.

2.8.5. Results

The real covariance functions are compared with the expectation of the estimator of the model defined in Section 2.8.3 in Figures 2.2 - 2.9. In all cases the bias seems minimal, which demonstrates the great flexibility of the spectral component model. It should be stressed that the model (2.8.2) may have too many components to give the best tradeoff between bias and variance when estimating the covariance function with eighty observations. The issues involved in model selection are addressed in detail in Powojowski (1999b). Here the emphasis is on illustrating the model flexibility.

The mean estimated spectral densities are shown in Figures 2.10 - 2.17, along with the spectral densities of the true models. It is seen that in all cases the mean of the estimate provides a fair approximation of the true spectral density.

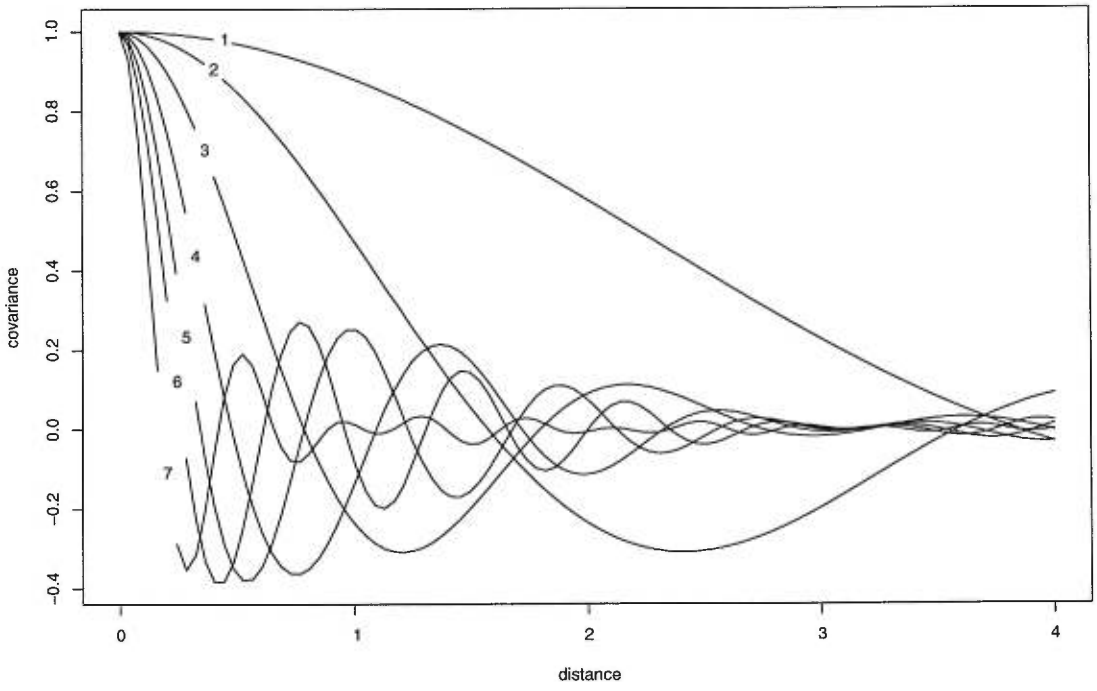

Figure 2.1: *Spectral components of the additive model.*

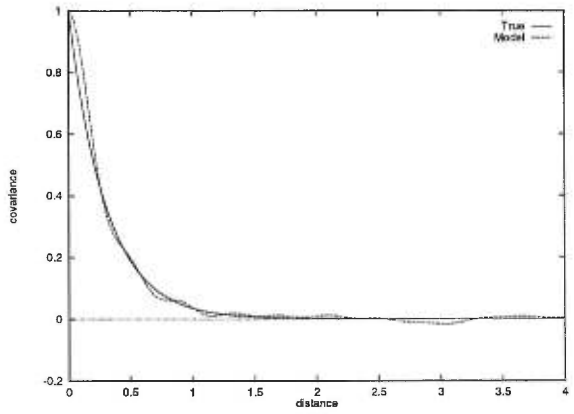

Figure 2.2: *Exponential model with* $(c, a) = (1, 0.3)$.

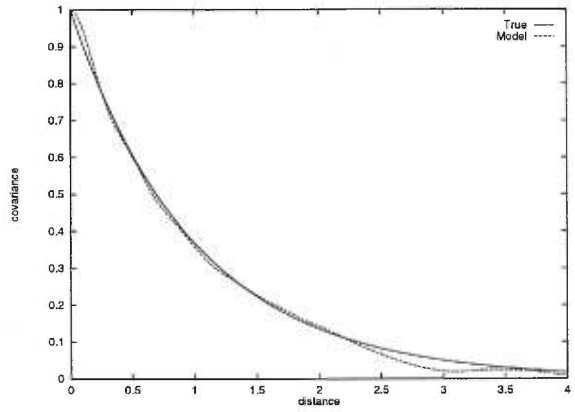

Figure 2.3: *Exponential model with* $(c, a) = (1, 1)$.

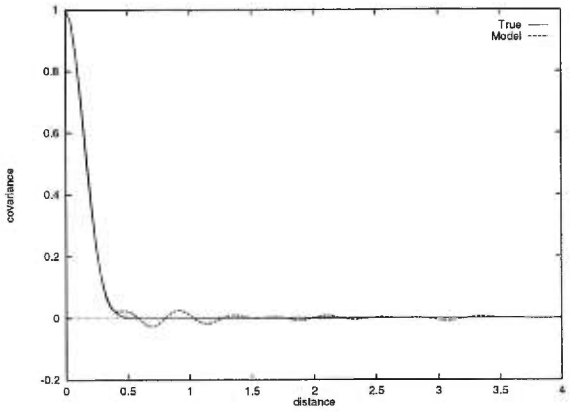

Figure 2.4: *Gaussian model with* $(c, a) = (1, 0.2)$.

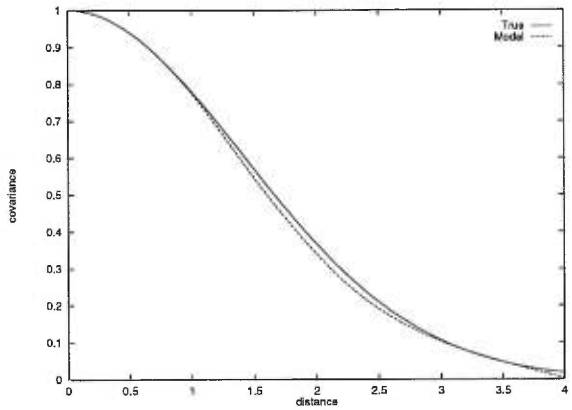

Figure 2.5: *Gaussian model with* $(c, a) = (1, 2)$.

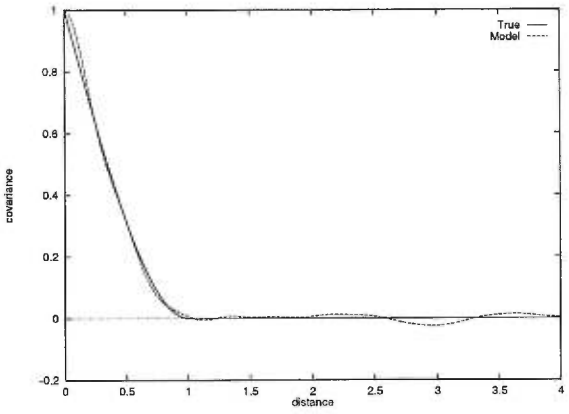

Figure 2.6: *Spherical model with* $(c, a) = (1, 1)$.

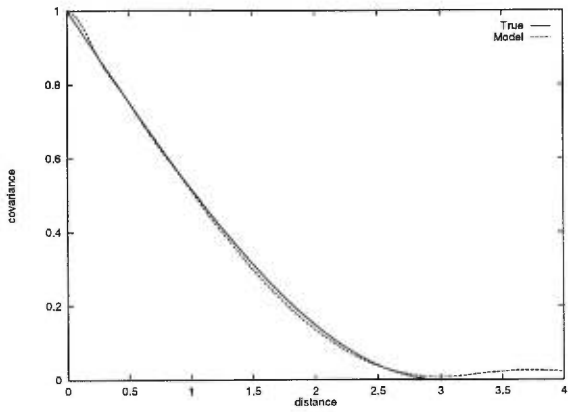

Figure 2.7: *Spherical model with* $(c, a) = (1, 3)$.

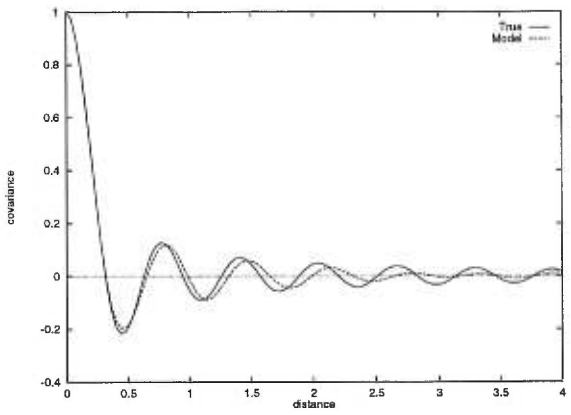

Figure 2.8: *Hole-effect model with* $(c, a) = (1, 0.1)$.

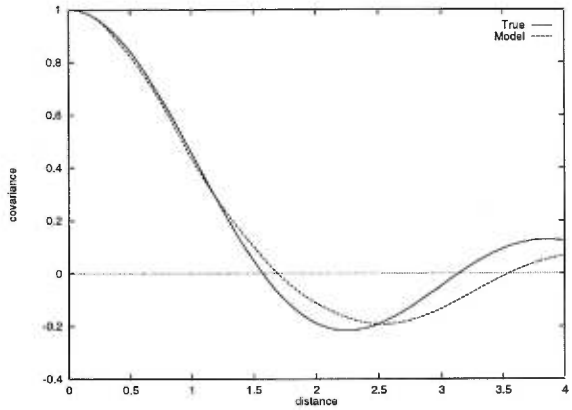

Figure 2.9: *Hole-effect model with* $(c, a) = (1, 0.5)$.

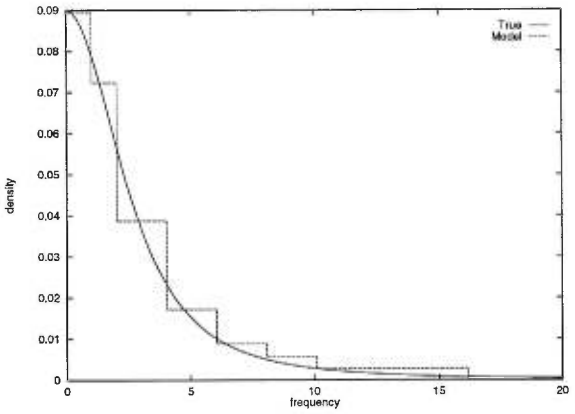

Figure 2.10: *Spectral density of the exponential model with* $(c, a) = (1, 0.3)$.

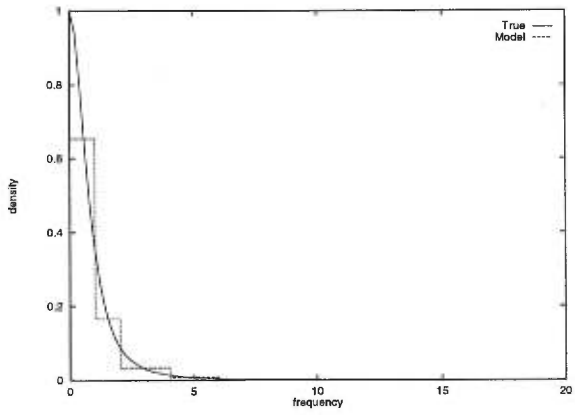

Figure 2.11: *Spectral density of the exponential model with* $(c, a) = (1, 1)$.

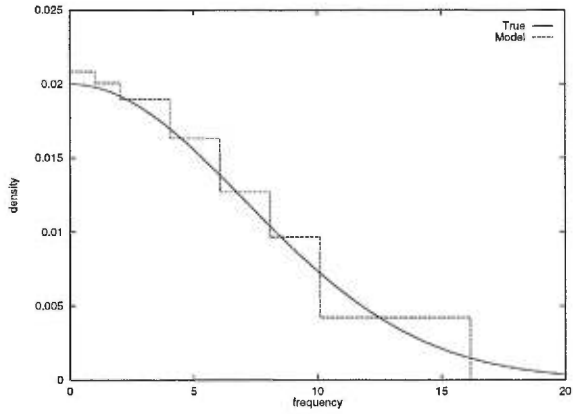

Figure 2.12: *Spectral density of the Gaussian model with* $(c, a) = (1, 0.2)$.

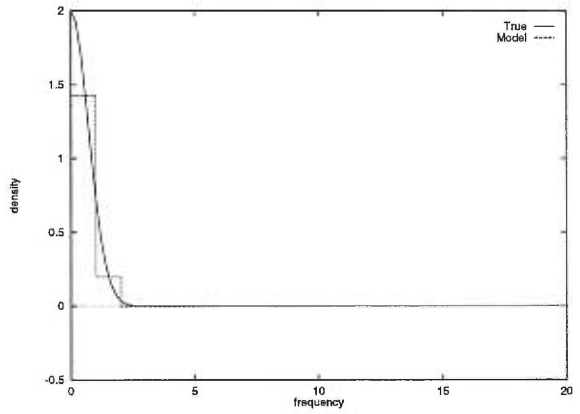

Figure 2.13: *Spectral density of the Gaussian model with* $(c, a) = (1, 2)$.

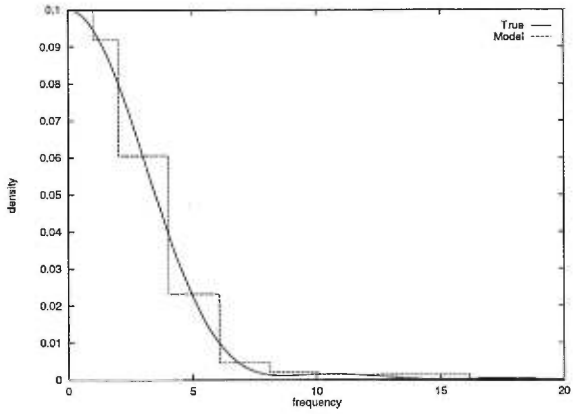

Figure 2.14: *Spectral density of the spherical model with* $(c, a) = (1, 1)$.

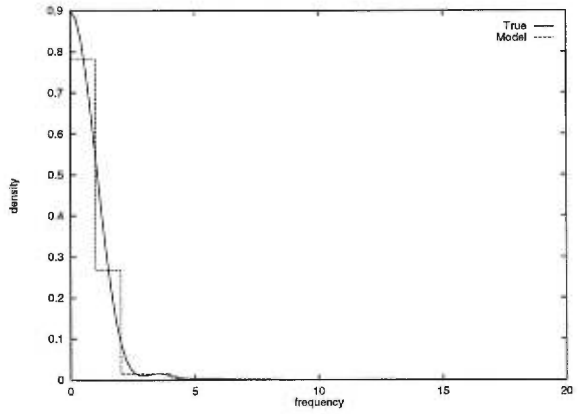

Figure 2.15: *Spectral density of the spherical model with* $(c, a) = (1, 3)$.

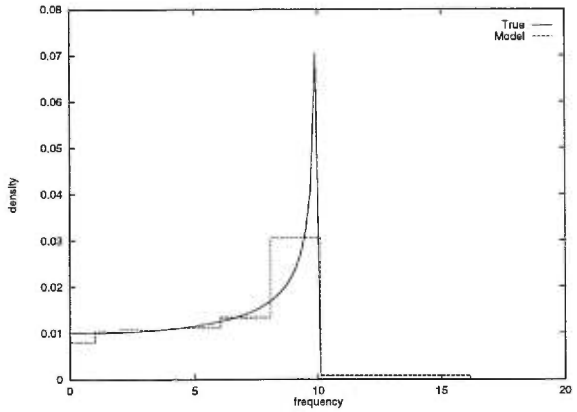

Figure 2.16: *Spectral density of the hole-effect model with* $(c, a) = (1, 0.1)$.

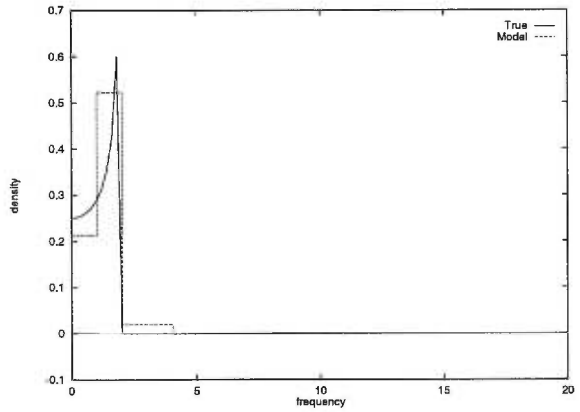

Figure 2.17: *Spectral density of the hole-effect model with* $(c, a) = (1, 0.5)$.

2.9. Conclusion

The paper proposes a new class of covariogram models for isotropic random processes, the spectral component additive models. The models have a natural motivation in the spectral representation of the covariance function. They generalise the ideas of Shapiro and Botha (1991), and possess some desirable properties, such as faster rate of decay, as described in the discussion following Proposition 2.6.1. The proposed class is very flexible, indeed, it contains elements which are arbitrarily close to any valid integrable covariance function possessing a spectral density.

The projection estimation methods proposed in Powojowski (1999a) are applicable to the class of spectral component additive models and lead to particularly simple asymptotic results. It is seen that the asymptotic mean of the projection estimator has a simple expression as a linear functional of the spectral density of the true process. It is also shown that, at least in the case of Gaussian process, the estimates of the individual covariance components are asymptotically uncorrelated.

A numerical study compares the standard parametric models with the means of the estimator obtained with a spectral component additive model of order seven (same fitted model for eight different true models). The experiment demonstrates that approximate lack of bias can indeed be achieved for a finite sample of eighty observations.

In addition, the study illustrates how the spectral density of the true covariance function may be estimated by the spectral density of the covariance estimate obtained using the spectral component additive model. While the spectral methods have not been applied extensively in geostatistics, it has been suggested that the reason for this is the requirement for the observations to be regularly spaced

on a lattice (Chiles and Delfiner, 1999). Spectral component additive models provide a way of estimating the spectral densities from irregularly spaced data.

REFERENCES

ADLER R.J. (1980). *The Geometry of Random Fields.* John Wiley & Sons.

CHILES, J.P., DELFINER, P. (1999). *Geostatistics: modeling spatial uncertainty.* John Wiley & Sons.

CRESSIE, N.A.C (1993). *Statistics for Spatial Data.* Revised edition. John Wiley & Sons.

POWOJOWSKI, M. (1999a). Additive Covariogram Models and Estimation through Projections. In preparation.

POWOJOWSKI, M. (1999b). Model Selection in Covariance Estimation. In preparation.

SCHOENBERG, I. J. (1938). Metric spaces and completely monotone functions. *The Annals of Mathematics* **39** 811-841.

SHAPIRO, A. and BOTHA, J.D. (1991). Variogram fitting with a general class of conditionally nonnegative definite functions. *Computational Statistics and Data Analysis* **11** 87-96.

SNEDDON, I.N (1972). *The Use of Integral Transforms.* McGraw-Hill.

Chapitre 3

MODEL SELECTION IN COVARIOGRAM ESTIMATION

3.1. Abstract

The modelling and estimation of the covariance function of a second-order stationary, isotropic random process involve a number of decisions in selecting the right model. Techniques are proposed, which address the model selection issues arising in applying the spectral component additive models described in Powojowski (1999b). A data-driven procedure for determining the spectral support is proposed. Another procedure is suggested for selecting the end-points of the component support intervals. Finally, a criterion is derived for deciding between competing models. The techniques are tested using synthetic data and are seen to produce good results. Finally, they are applied to a real data set of Davis (1973).

3.2. Introduction

For a random process $Y(x), x \in D$, where D is a subset of a d-dimensional Euclidean space, the covariogram is defined as $C_Y(x_1, x_2) = \mathrm{cov}(Y(x_1), Y(x_2))$. A common problem in geostatistics is one of estimating the function C_Y based on one realisation of the process Y observed at a finite number of locations $x_1, x_2, \ldots, x_n$

in D. It is important to note that the knowledge of function values $C(x_1, x_2)$ for arbitrary $(x_1, x_2) \in D^2$ is required, and not simply the covariances of Y at lags observed in the sample. The fact of observing only one realisation forces one to make certain assumptions about the process Y, which translate into restrictions on the form of C_Y. There also exist theoretical reasons for restricting the function families considered. The covariogram has to be a positive definite function. Further restrictions may be desirable. In a typical covariogram estimation problem it is supposed that the observed process Y follows the model

$$Y = X\beta + \eta.$$

The known regressor X contains terms corresponding to the mean of the process and any trend that is allowed for, while the parameter β is unknown and the random term η is assumed to have zero mean and an unknown covariogram C_Y. Throughout the paper, the process η will be assumed isotropic, requiring $C_Y(x_1, x_2)$ to depend only on $\| x_1 - x_2 \|$. Putting $\rho = \| x_1 - x_2 \|$, the covariogram of an isotropic process Y will be denoted by $\mathbf{C}_Y(\rho) = C_Y(x_1, x_2)$. A covariance model $\mathbf{C}_\theta(\rho)$, known up to the value of a finite-dimensional parameter θ, to be estimated, is to be fitted to the observed data. Throughout the paper, the model $\mathbf{C}_\theta(\rho)$ will be additive, that is, it will have the form

$$\mathbf{C}_\theta = \sum_{i=1}^{q} \theta(i)\mathbf{C}_i,$$

where the functions $\mathbf{C}_i$ are fully specified valid isotropic covariance models. The class of spectral component additive models for the covariance function of an isotropic random process described in Powojowski (1999b) contains elements arbitrarily close to the covariogram function of any isotropic process with absolutely continuous (with respect to the Lebesgue measure) spectral measure. The projection estimator described in Powojowski (1999a) is well suited for parameter

estimation in spectral component additive models. When attempting to use such a model, one must decide on the number and shape of the additive components in the model. This paper suggests methods of model selection based on the number of observations, their locations and the observed values.

The paper is organised as follows: Firstly, notation is established, the spectral component additive models are introduced and the methods of projection estimation are reviewed. Next, a technique for determining the effective spectral support of a covariance function is derived and tested with simulated data. The issue of end-point selection for the spectral components is addressed next. Subsequently, a criterion for model comparison is derived and tested in a simulation study. Finally, all of the proposed techniques are applied to a data set of Davis (1973).

3.3. Spectral component additive models and their estimation

This section defines the notation used throughout the paper, introduces spectral component additive models and the projection estimation methods. For details on projection estimators the reader is referred to Powojowski (1999a). Spectral component additive models are described in detail in Powojowski (1999b).

3.3.1. Notation

To avoid confusion which might arise due to the frequent occurrence of multiple subscripts, the following notation will be used throughout the paper. If A is a matrix, its entries will be denoted by $A(i,j)$, while $A_{i,j}$ may denote a matrix from some (doubly) indexed set of matrices. Similarly, if θ is a vector, its components

will be denoted by $\theta(i)$, while θ_i may denote a vector from some indexed set of vectors.

In the most general setting, one considers a random process Y on the domain D, a subset of a d-dimensional Euclidean space. The process Y is observed at n locations $\{x_i\}_{i=1}^n$, $x_i \in D$. Let $Y_n = (Y(x_1), \ldots, Y(x_n))'$ and $Y_n(i) = Y(x_i), 1 \leq i \leq n$. It will be further assumed that

$$Y_n = X_n\beta + \eta_n \tag{3.3.1}$$

where $\eta_n = (\eta(x_1), \ldots, \eta(x_n))'$ and η is an isotropic (and hence also second-order stationary) random process with $\mathbf{E}[\eta] = 0$. It will be assumed that X_n has p columns corresponding to different regression terms. Thus $X_n(l,k) = r_k(x_l)$, $1 \leq k \leq p$, $1 \leq l \leq n$, where x_l is the l-th location in the sample and r_k is a continuous function defined on D and it is the k-th regression term in the mean of Y. If present, the term $r_1 \equiv 1$ corresponds to the (non-zero) constant term in the mean of Y. The term $r_k(x_l) = x_l(1)$, where $x_l(1)$ is the first component of the d-dimensional vector $x_l \in D$, would correspond to a linear trend in the mean of $Y(x)$ in the direction of the first component of x. The matrix X_n will always be known, while the $p \times 1$ vector β may have to be estimated. If $\mathbf{C}_Y(\rho)$ denotes the covariance function of the process Y (and of the zero-mean process η), the matrix $K_{Y,n} = \mathrm{var}(Y_n)$ is a symmetric matrix with entries $K_{Y,n}(i,j) = \mathbf{C}_Y(\| x_i - x_j \|)$. If $\mathbf{C}_\theta$ is a given covariance function model, one defines the symmetric matrix $K_{\theta,n}$ in a similar way, by putting $K_{\theta,n}(i,j) = \mathbf{C}_\theta(\| x_i - x_j \|)$. Thus $K_{\theta,n}$ is a fixed matrix depending only on the model $\mathbf{C}_\theta$ and on the set of locations $\{x_i\}_{i=1}^n$, $x_i \in D$.

If $\mathbf{C}_Y$ is the covariogram of an isotropic random process Y on $\mathbf{R}^d$, whose spectral measure is absolutely continuous, it has the spectral representation

$$\mathbf{C}_Y(\rho) = \rho^{(2-d)/2} \int_0^\infty \lambda \Psi_Y(\lambda) J_{(d-2)/2}(\lambda\rho) d\lambda \tag{3.3.2}$$

where J_ν is the Bessel function of the first kind and ν-th order and $\Psi_Y(\lambda)$ is a non-negative function such that $\lambda^{d/2}\Psi_Y(\lambda)$ is summable on the positive real line (Schoenberg, 1938). The function Ψ_Y will be referred to as the spectral density of the covariance function $\mathbf{C}_Y$. The spectral density Ψ_θ of the covariance model $\mathbf{C}_\theta$ is defined by an expression similar to (3.3.2) with $\mathbf{C}_Y$ replaced by $\mathbf{C}_\theta$.

Two intervals $[a_1, b_1]$ and $[a_2, b_2]$ will be called non-overlapping if their intersection is at most one point.

3.3.2. Spectral component additive models

For simplicity, it will initially be assumed that the isotropic random process of interest, Y, has zero mean. Following the approach in Powojowski (1999b) the class of models considered will be parametrised as

$$\mathbf{C}_\theta(\rho) = \sum_{i=1}^{q} \theta(i) \mathbf{C}_i(\rho) \tag{3.3.3}$$

where

$$\mathbf{C}_i(\rho) = \rho^{(2-d)/2} \int_{a_i}^{b_i} \lambda \Psi(\lambda) J_{(d-2)/2}(\lambda\rho) d\lambda. \tag{3.3.4}$$

The functions $\mathbf{C}_i$ will be referred to as the spectral components (as will their spectral densities, the functions $\Psi_i = \Psi I_{[a_i,b_i]}$ when no ambiguity exists). The intervals $[a_i, b_i]$ are finite and non-overlapping. The function $\Psi(\lambda)$ is a completely specified, non-negative, piecewise continuous function. If $d > 1$, one convenient

choice is

$$\Psi(\lambda) = \lambda^{(d-2)/2} \tag{3.3.5}$$

since it results in the covariance functions with closed forms

$$\mathbf{C}_i(\rho) = \rho^{(2-d)/2}\left(\frac{b_i^{d/2}}{\rho} J_{d/2}(b_i\rho) - \frac{a_i^{d/2}}{\rho} J_{d/2}(a_i\rho)\right). \tag{3.3.6}$$

Another choice would be

$$\Psi(\lambda) \equiv 1. \tag{3.3.7}$$

For $d = 2$ the two are identical, otherwise the latter choice may result in expressions for the spectral components without convenient forms. In any case, it can be easily shown that

$$\mathbf{C}_i(0) = \frac{2^{(d-2)/2}}{\Gamma(d/2)} \int_{a_i}^{b_i} \Psi(\lambda)\lambda d\lambda.$$

3.3.3. Projection estimators for additive covariance models

Assuming that the parameter q and the intervals $(a_i, b_i), 1 \leq i \leq q$ are given, and that the process Y is observed at the locations $(x_1, x_2, \ldots, x_n)'$, yielding the sample $Y_n = (Y(x_1), Y(x_2), \ldots, Y(x_n))'$, the coefficients $\theta(i)$ can be estimated via

$$\hat{\theta}_n = [\mathrm{tr}(K_{i,n}K_{j,n})]^{-1}[Y_n' K_{i,n} Y_n] \tag{3.3.8}$$

where $\hat{\theta}_n = (\hat{\theta}_{1,n}, \ldots, \hat{\theta}_{q,n})'$ and $K_{i,n}$ is the matrix whose (m,l)-th entry is $\mathbf{C}_i(\| x_m - x_l \|)$. (The notation $[\mathrm{tr}(K_{i,n}K_{j,n})]$ denotes a $q \times q$ matrix whose (i,j) - th entry is $\mathrm{tr}(K_{i,n}K_{j,n})$. Similarly, $[Y_n' K_{i,n} Y_n]$ denotes a $q \times 1$ vector.) The estimator $\hat{\theta}_n$ in (3.3.8) may equivalently be defined as the vector θ which

minimises the objective function

$$\| Y_n Y_n' - \sum_{i=1}^{q} \theta(i) K_{i,n} \|^2 = \mathrm{tr}\left(\left(Y_n Y_n' - \sum_{i=1}^{q} \theta(i) K_{i,n} \right)^2 \right). \tag{3.3.9}$$

The properties of the estimator (3.3.8) and its extensions are described in Powojowski (1999a) and Powojowski (1999b).

As is shown in Powojowski (1999b) if the process is observed on an increasingly large domain at a sufficient number of uniformly placed locations, the function

$$\mathbf{E}[\mathbf{C}_{\hat{\theta}}(\rho)] = \sum_{i=1}^{q} \mathbf{E}[\hat{\theta}_i] \mathbf{C}_i(\rho) \tag{3.3.10}$$

converges to the orthogonal projection of the true covariance function $\mathbf{C}_Y$ on the linear space spanned by the functions $\mathbf{C}_i$, where the inner product is defined as

$$\langle \phi_1, \phi_2 \rangle = \int_0^{\infty} \phi_1(\rho) \phi_2(\rho) \rho^{d-1} d\rho. \tag{3.3.11}$$

The norm resulting from the inner product is

$$\| \phi_1 - \phi_2 \| = \langle \phi_1 - \phi_2, \phi_1 - \phi_2 \rangle^{1/2}. \tag{3.3.12}$$

Applying the definitions and Parseval's formula (Sneddon, 1972), one obtains the following expression for the norm of the squared bias

$$\begin{aligned} \| \mathbf{C}_Y - \mathbf{E}[\mathbf{C}_{\hat{\theta}}] \|^2 &= \int_0^{\infty} \left(\mathbf{C}_Y(\rho) - \sum_{i=1}^{q} \mathbf{E}[\hat{\theta}_i] \mathbf{C}_i(\rho) \right)^2 \rho^{d-1} d\rho \\ &= \int_0^{\infty} \left(\Psi_Y(\lambda) - \sum_{i=1}^{q} \mathbf{E}[\hat{\theta}_i] \Psi_i(\lambda) \right)^2 \lambda d\lambda \\ &= \sum_{i=1}^{q} \int_{a_i}^{b_i} (\Psi_Y(\lambda) - \mathbf{E}[\hat{\theta}_i] \Psi(\lambda))^2 \lambda d\lambda + \int_{S^c} (\Psi_Y(\lambda))^2 \lambda d\lambda \end{aligned} \tag{3.3.13}$$

where S^c is the complement (with respect to the positive real line) of the maximal support S of the function $\mathbf{C}_\theta$, which is the union of the intervals $[a_i, b_i]$. From

the last expression, one gains the qualitative intuition that the bias is reduced by making the support S cover more of the support of Ψ_Y (or at least the portions of it where Ψ_Y takes high values) or by making the intervals $[a_i, b_i]$ smaller, so that the piecewise approximations of $\Psi_Y(\lambda)$ by the functions $\Psi_i(\lambda)$ are better.

A common complication in working with real data is the presence of unknown mean in the process. More precisely, the observed process Y is of the form (3.3.1), that is

$$Y_n = X_n\beta + \eta_n$$

where X_n is a fully specified regression matrix, β is an unknown vector parameter, and η_n is a zero-mean, isotropic process. In such a case, one may compute the regression residuals e_n given by

$$e_n = (I_n - X_n(X_n'X_n)^{-1}X_n')Y_n = P_nY_n, \tag{3.3.14}$$

where P_n is a projection matrix. The estimator (3.3.8) extends to the case of a process with unknown mean to take the form:

$$\hat{\theta}_n = [\mathrm{tr}(U_{i,n}U_{j,n})]^{-1}[e_n'U_{i,n}e_n] \tag{3.3.15}$$

where

$$U_{i,n} = P_nK_{i,n}P_n.$$

The estimator $\hat{\theta}_n$ in (3.3.15) may equivalently be defined as the vector θ which minimises the objective function

$$\| e_ne_n' - \sum_{i=1}^{q}\theta(i)U_{i,n} \|^2 = \mathrm{tr}\left(\left(e_ne_n' - \sum_{i=1}^{q}\theta(i)U_{i,n}\right)^2\right). \tag{3.3.16}$$

Details may be found in Powojowski (1999a).

3.4. SPECTRAL SUPPORT SELECTION

The model (3.3.3) has a compactly supported spectral density, whereas the spectral density Ψ_Y of the process may have infinite support. The expression (3.3.13) points out the potential bias that may be introduced by truncating the spectral density — namely the last term corresponds to the contribution of the frequencies which are not accounted for in the model $\mathbf{C}_\theta$. Since the true spectral density Ψ_Y satisfies the requirement that $\lambda^{d/2}\Psi_Y(\lambda)$ be summable, if in addition Ψ_Y is bounded, the term

$$\int_{S^c} (\Psi_Y(\lambda))^2 \lambda d\lambda$$

can be made arbitrarily small by selecting an appropriate compact support S. Thus the issue is to select the support S sufficiently large to avoid introducing a large bias. This S will be referred to as the effective support of the spectral density function of the process. On the other hand, selecting a support larger than necessary forces one to use larger intervals $[a_i, b_i]$ for the fixed value of q, or increasing the number of components q to keep the size of the intervals down.

A data-driven procedure will now be proposed to determine the effective support of Ψ_Y. Given $\nu > 0$, the model

$$\mathbf{C}_{0,\nu}(\rho) = \theta_{0,\nu} \frac{2}{\nu^2} \rho^{(2-d)/2} \int_0^\nu \lambda J_{(d-2)/2}(\lambda\rho) d\lambda \tag{3.4.1}$$

is a special case of (3.3.3), with $\Psi(\lambda) \equiv 2/\nu^2$, $q = 1$ and $[a_1, b_1] = [0, \nu]$. It is shown in Powojowski (1999a) that under certain assumptions concerning the sampling configuration, as the number of observations n increases, one has

$$\lim_{n\to\infty} \mathbf{E}[\hat{\theta}_{0,\nu,n}] = \frac{\langle \mathbf{C}_{0,\nu}, \mathbf{C}_Y \rangle}{\langle \mathbf{C}_{0,\nu}, \mathbf{C}_{0,\nu} \rangle}$$

where $\hat{\theta}_{0,\nu,n}$ is the estimator given by (3.3.8) for the one-component model (3.4.1), computed from the n available observations of the process Y. The inner product

is that given by (3.3.11). Using the spectral representation of the last result, as presented in Powojowski (1999b), one obtains

$$\frac{\langle \mathbf{C}_{0,\nu}, \mathbf{C}_Y \rangle}{\langle \mathbf{C}_{0,\nu}, \mathbf{C}_{0,\nu} \rangle} = \frac{\int_0^\nu \Psi_Y(\lambda)\Psi(\lambda)\lambda d\lambda}{\int_0^\nu (\Psi(\lambda))^2 \lambda d\lambda} = \int_0^\nu \Psi_Y(\lambda)\lambda d\lambda$$

since the function Ψ is a constant chosen to satisfy the last equality. Thus for a sufficiently large sampling domain, sampled sufficiently densely (in the sense made precise in Powojowski, 1999a), one obtains

$$\mathbf{E}[\hat{\theta}_{0,\nu}] \approx \frac{\langle \mathbf{C}_{0,\nu}, \mathbf{C}_Y \rangle}{\langle \mathbf{C}_{0,\nu}, \mathbf{C}_{0,\nu} \rangle} = \int_0^\nu \Psi_Y(\lambda)\lambda d\lambda. \tag{3.4.2}$$

Hence, assuming that $\hat{\theta}_{0,\nu}$ is a continuously differentiable function of ν (which is easily verified), one obtains

$$\mathbf{E}[\frac{\partial}{\partial \nu}\hat{\theta}_{0,\nu}] = \frac{\partial}{\partial \nu}\mathbf{E}[\hat{\theta}_{0,\nu}] \approx \nu\, \Psi_Y(\nu). \tag{3.4.3}$$

The procedure consists of constructing a large number k of models of the form (3.4.1) where the values ν are taken to be

$$\nu_j = \frac{j}{k}\,\nu_{max}, \qquad 1 \leq j \leq k \tag{3.4.4}$$

where ν_{max} is a fixed value, believed to be high enough so that higher frequencies must be irrelevant. (Some heuristics for the selection of ν_{max} will be given later.) The estimator $\hat{\theta}_{0,\nu_j}$ of (3.3.8) is computed for $1 \leq j \leq k$. Plotting the values $(\hat{\theta}_{0,\nu_j} - \hat{\theta}_{0,\nu_{j-1}})/\nu_j$ against the values ν_j for $1 < j \leq k$ provides an estimate $\hat{\Psi}_Y(\nu_j)$ of $\Psi_Y(\nu_j)$. The estimate is a function of ν which can be examined to determine whether there seems to be a point ν_t beyond which $\Psi_Y(\nu)$ is essentially zero. The interval $[0, \nu_t]$ will be used as the effective support of Ψ_Y.

The selection of ν_{max} may be based on the availability of sampling locations. The basic idea is that if the sampled locations $(x_1, \ldots, x_n)$ provide minimal non-zero distance of ρ_{min} between sampled locations, one is unlikely to be very successful at fitting a model of type (3.3.3) containing components whose behaviour on the interval $[0, \rho_{min}]$ varies considerably, since there simply are no data for it. Examining the representation (3.3.2) of the covariance function, one observes that on any compact set $\mathbf{C}_Y(\rho)$ may be approximated arbitrarily closely by a sum of the form

$$\sum_i c_i \rho^{(2-d)/2} J_{(d-2)/2}(\nu_i \rho). \tag{3.4.5}$$

The function $\rho^{(2-d)/2} J_{(d-2)/2}(\rho)$ has the value $2^{(d-2)/2}/\Gamma(d/2)$ at the origin and as ρ increases, eventually hits its first zero at some value $\rho_{d,1}$. It would seem futile to include components in (3.4.5) which attain zero for values of ρ less than ρ_{min}. Since the first positive root of $J_{(d-2)/2}(\nu_i \rho)$ is $\rho_{d,1}/\nu_i$, one may take the value of $\nu_{max} = \rho_{d,1}/\rho_{min}$ as a rough heuristic guess. While it is entirely possible that the contribution of frequencies higher than $\rho_{d,1}/\rho_{min}$ to the spectral density Ψ_Y are significant, the prospects for their adequate modelling and estimation given the available data are slim. For $d = 2$ one has approximately $\rho_{2,1} = 2.4048$, the first positive root of the Bessel function J_0.

The procedure is illustrated by a simulation study. A square with side of length two is sampled uniformly to select twenty locations which are fixed throughout the study. Two covariance models are used to simulate zero-mean Gaussian random processes on the square (hence the domain is two-dimensional). Both models have the same parametric form:

$$\mathbf{C}_Y(\rho) = c \exp(-\rho/a) \tag{3.4.6}$$

and their spectral densities are given by

$$\Psi_Y(\lambda) = \frac{c}{a}\left(\lambda^2 + \frac{1}{a^2}\right)^{-3/2}. \tag{3.4.7}$$

In both cases $c = 1$, with $a = 0.3$ for model A and $a = 1$ for model B. In each case the process is observed at the fixed twenty locations and the spectral density estimate $\hat{\Psi}_Y(\nu)$ is computed as described above. The number of discrete ν_j considered is $k = 200$ and the minimal (non-zero) distance found in the sample is 0.0450, thus producing $\nu_{max} = 53.44$. Thus for each realisation a curve is obtained by plotting the values $(\hat{\theta}_{0,\nu_j} - \hat{\theta}_{0,\nu_{j-1}})/\nu_j$ against the values ν_j for $1 < j \leq k$. To avoid excessive cluttering, only twenty simulations were performed. The curves obtained are plotted, along with the true spectral density, in Figures 3.1 and 3.2. Since in both cases $\hat{\Psi}_Y(\nu)$ becomes very nearly zero for values much smaller than $\nu_{max} = 53.44$, the plots shown in Figures 3.1 and 3.2 are truncated at $\nu_{max} = 25$ and $\nu_{max} = 12$, respectively. While the procedure admittedly has a subjective component (of having to choose the first point beyond which the estimated spectral density becomes zero), it appears that the effective support of the true density function is captured rather well by the estimates. It is seen from Figure 3.1 that based on any one of the plotted realisations, the effective support would likely be selected to be the interval starting at zero and ending somewhere between 7 and 15, which seems reasonable given the shape of the true spectral density function. Figure 3.2 implies that the effective support would likely be selected to be the interval starting at zero and ending somewhere between 3 and 9, which again seems reasonable.

In practice the mean of the process Y is often not known. In this case one attempts to model the mean of the process Y as in (3.3.1) and apply the estimator (3.3.15) to estimate the covariance function of the process. To estimate

the effective spectral support of the true covariance function, one modifies the approach described above by using the estimator in (3.3.15), rather than that in (3.3.8) to obtain $\hat{\theta}_{0,\nu_j}$. This approach is tried for models A and B described above. Two choices of the mean model X_n are considered. In one case the regressor matrix X_n is assumed to contain just one column of ones (thus Y_n has non-zero, but constant mean). Hence in this case $X_n = X_{const}$ has one column with $X_{const}(k,1) = 1, 1 \leq k \leq n$. In the other, X_n contains three columns which allow for any linear trend on a two-dimensional domain. In this case $X_n = X_{lin}$ has three columns with the k-th row of the matrix X_{lin} being $(1, x_k(1), x_k(2))$, where $(x_k(1), x_k(2)), 1 \leq k \leq n$ are the coordinates of the k-th location in the sample. The results of estimating $\hat{\Psi}_Y(\nu)$ are presented in Figures 3.3 and 3.4 for the constant mean X_{const}, and in 3.5 and 3.6 for the linear trend X_{lin}. It appears that the spectral density is underestimated when the residuals are used, particularly for low frequencies. This is intuitively correct, since low frequency dependence in the data will be indistinguishable from a linear trend on a sufficiently small domain. Computing the residuals after fitting a linear trend will effectively filter the low-frequency components out. The encouraging message from the figures is that the effective support of the spectral density can still be inferred with the described method.

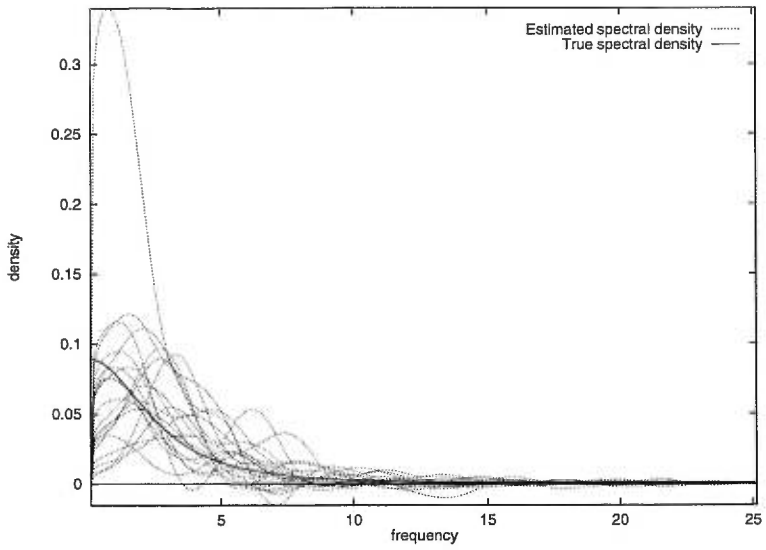

Figure 3.1: *Estimates of the spectral density of model A.*

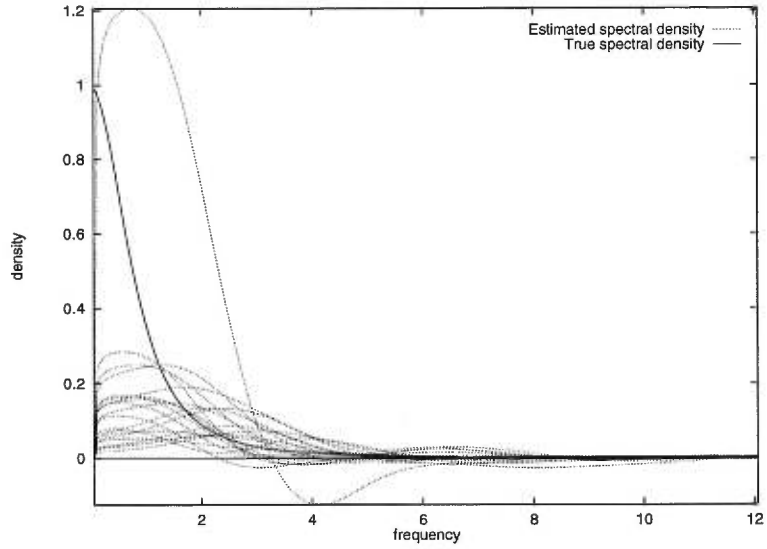

Figure 3.2: *Estimates of the spectral density of model B.*

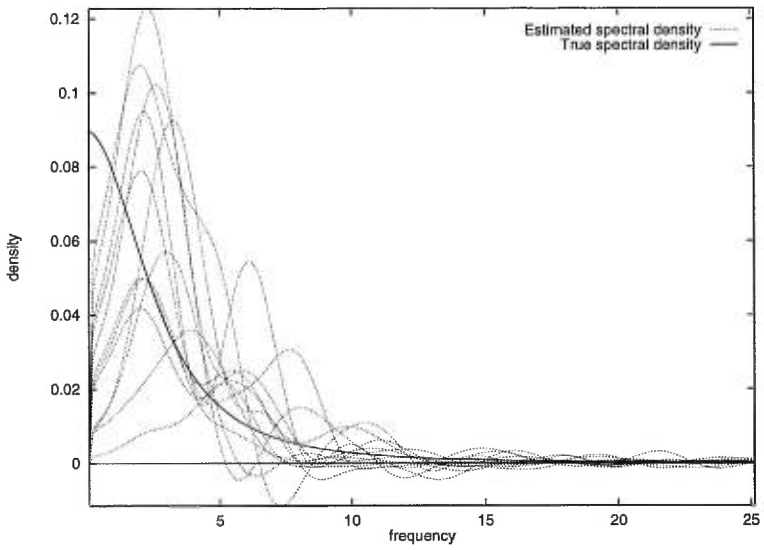

Figure 3.3: *Estimates of the spectral density of model A from a process with unknown constant mean.*

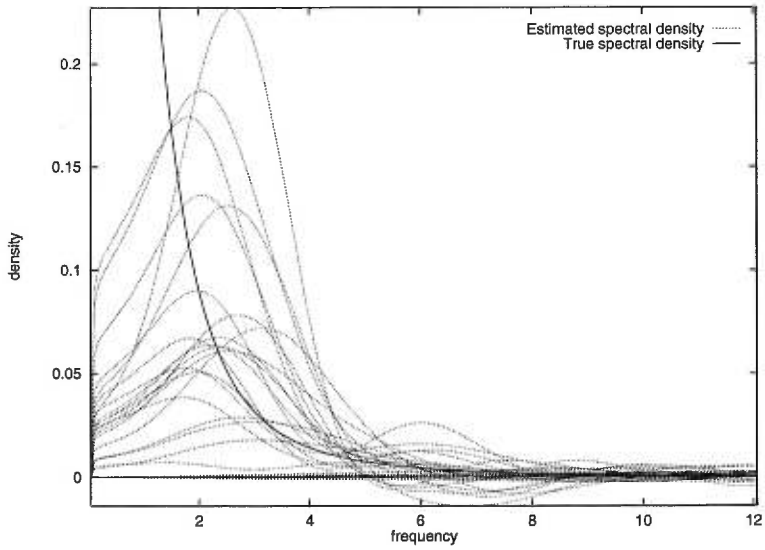

Figure 3.4: *Estimates of the spectral density of model B from a process with unknown constant mean.*

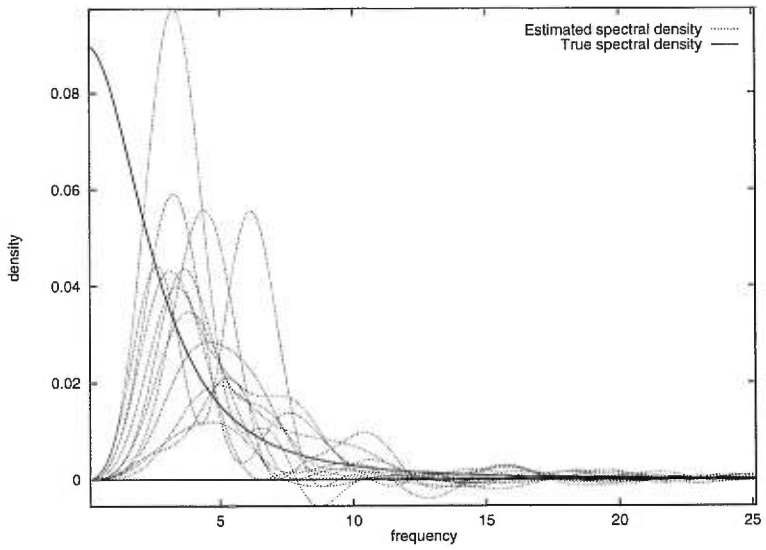

Figure 3.5: *Estimates of the spectral density of model A from a process with unknown linear trend.*

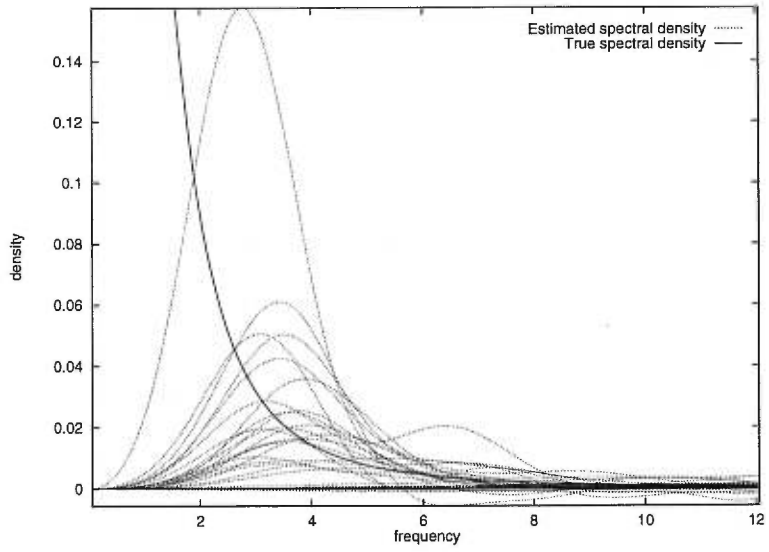

Figure 3.6: *Estimates of the spectral density of model B from a process with unknown linear trend.*

3.5. END-POINT SELECTION

Having selected the support for Ψ_θ (possibly by the methods described in the previous section), one is faced with the problem of selecting the number of components q and the interval ends a_i and b_i, for $1 \leq i \leq q$. This section deals with selecting the end-points, assuming that both the support $[0, \nu_t]$ and the order q have been decided upon. For simplicity the class of models with $\Psi(\lambda)$ given by (3.3.7) is considered, though the principle can be applied generally. Optimal selection of the end-points depends on the ultimate goals one wants to achieve. Often the covariogram estimation is an intermediate step in a larger study. This section discusses some possible approaches to end-point selection, while recognising that different criteria of optimality may be dictated by the broader objective of covariogram estimation.

As is seen in expression (3.3.13), the contribution to the squared bias from the spectral component $\mathbf{C}_i$ is

$$h_i = \min_{\theta(i)} \left\{ \int_{a_i}^{b_i} (\Psi_Y(\lambda) - \theta(i))^2 \lambda d\lambda \right\} \tag{3.5.1}$$

and it is caused by the departure of the function $\Psi_Y(\lambda)$ from being constant on the interval $[a_i, b_i]$. To simplify the analysis, it will be assumed that the function $\Psi_Y(\lambda)$ is approximately linear in that interval, hence

$$\Psi_Y(\lambda) \approx \alpha_i + \beta_i \lambda, \qquad \lambda \in [a_i, b_i]. \tag{3.5.2}$$

If (3.5.2) holds exactly, it can easily be shown that

$$h_i = \frac{\beta_i^2}{36} \left(9(b_i^4 - a_i^4) - 8 \frac{(b_i^3 - a_i^3)^2}{b_i^2 - a_i^2} \right). \tag{3.5.3}$$

Not surprisingly, higher absolute values of β_i cause larger contribution to bias. If some prior knowledge of the steepness of the function $\Psi_Y(\lambda)$ is available, it could be used to ensure that regions where $\Psi_Y(\lambda)$ changes rapidly are divided

into smaller intervals than flat regions. The estimated spectral density curve of the previous section may give some indication of the derivative of $\Psi_Y(\lambda)$.

Assuming that nothing is known about the slope of $\Psi_Y(\lambda)$, one may try to select end-points in such a way that for equal values of $|\beta_i|$, for all $1 \leq i \leq q$, the contributions h_i to bias from components $\mathbf{C}_i$ are equal.

This may be compared to choosing equal length bins for probability density estimation via histogram. Equal length bins are the optimal choice only if the true density has constant slope over its entire support, which generally is not the case. However, in the absence of knowledge about the slope of the true density one usually chooses the equal bin sizes, which contribute equal error terms due to approximation by a step function under the assumption that the slopes of the density function are equal in each bin.

Assuming that the $|\beta_i|$ are equal, it can easily be shown that in order to construct q adjacent intervals with equal values h_i, together with $a_1 = 0$ and $b_1 = b > 0$, one puts

$$b_i = b_1 \prod_{j=1}^{i} \gamma_j, \qquad 1 \leq i \leq q \tag{3.5.4}$$

$$a_{i+1} = b_i, \qquad 1 \leq i < q \tag{3.5.5}$$

where the γ_j are defined recursively by

$$\gamma_1 = 1,$$

and where γ_{j+1} is the smallest root of the equation

$$9(\gamma_{j+1}^4 - 1) - 8\frac{(\gamma_{j+1}^3 - 1)^2}{\gamma_{j+1}^2 - 1} - \left(\prod_{l=1}^{j} \gamma_l\right)^{-4} = 0$$

which exceeds 1. The recursive definition is used to compute the initial ten values of γ_j, shown in Table 3.1. In order to construct q intervals whose joint

j	1	2	3	4	5	6	7	8	9	10
γ_j	1	1.637	1.342	1.234	1.178	1.144	1.121	1.104	1.091	1.081
$\prod_{l=1}^{j} \gamma_l$	1	1.637	2.196	2.711	3.194	3.655	4.096	4.522	4.936	5.338

Table 3.1: *Solutions of the recursive equation for optimal end-points.*

support is $[0, \nu_t]$ it suffices to put $b = \nu_t/(\prod_{l=1}^{q} \gamma_l)$. It is apparent that the above construction of intervals $[a_i, b_i]$ results in shorter intervals at higher frequencies. This is a consequence of the weighting by λ in the integral in (3.5.1). While this is consistent with the definition of the inner product in (3.3.11), it downplays the importance of good fit for low frequencies λ and sometimes leads to covariance models whose visual fit may seem inadequate. In geostatistical practice, the short-range behaviour of the covariance model is usually most important, hence the emphasis on high frequencies may be desirable.

However, for some applications one may be more concerned with accurate estimation of low frequencies. This might be the case, for example, if one were interested in the range over which the covariance function is considerably different from zero. In such applications, it is possible to define an alternative inner product to (3.3.11), giving rise to a norm (defined by (3.3.12) with the new inner product) which is more sensitive to differences at lower frequencies. For example, on the space of functions of the form

$$\phi_i(\rho) = \rho^{(2-d)/2} \int_0^\infty \lambda \Psi_i(\lambda) J_{(d-2)/2}(\lambda\rho) d\lambda \tag{3.5.6}$$

where the functions $\lambda^{d/2}\Psi_i(\lambda)$ are summable, one may define the inner product

$$\langle \phi_1, \phi_2 \rangle_w = \int_0^\infty \Psi_1(\lambda)\Psi_2(\lambda) w(\lambda) d\rho \tag{3.5.7}$$

where $w(\lambda)$ is any fixed, piecewise continuous, bounded function. The set of valid covariance functions possessing spectral density is a subset of this inner product space.

The natural interpretation of $\| \mathbf{C}_Y - \mathbf{E}[\mathbf{C}_{\hat{\theta}}] \|_w$ as the squared bias of $\mathbf{C}_{\hat{\theta}}$ is not preserved, since Parseval's formula no longer applies. Setting $w(\lambda) \equiv 1$ results in the following expression for the measure of departure due to the i-th spectral component

$$h_{1,i} = \min_{\theta(i)} \left\{ \int_{a_i}^{b_i} (\Psi_Y(\lambda) - \theta(i))^2 d\lambda \right\}. \tag{3.5.8}$$

In order to make these contributions equal under the assumptions of linearity of Ψ_Y as in (3.5.2) and the equality of the $|\beta_i|$, the lengths $b_i - a_i$ of the intervals must be equal.

Thus the optimal choice of the end-points may depend on one's broader objectives. In subsequent sections the first method described in this section will be applied.

3.6. Model comparison

So far the discussion of appropriate model selection has been concerned with reducing the bias only. As is often the case, there is a tradeoff between constructing richer models to reduce bias and keeping the number of parameters to be estimated small in order to reduce the overall variance. The standard approach is to minimise the mean squared error of the estimate. Let the process Y with the true covariance function $\mathbf{C}_Y$ be observed at the locations $(x_1, x_2, \ldots, x_n)$ and let K_i be the matrix whose (m,l)-th entry is $\mathbf{C}_i(\| x_m - x_l \|)$, with the matrices K_Y and K_θ defined in an analogous fashion. Thus $K_\theta = \sum_{k=1}^{q} \theta(i) K_i$ for any

vector $\theta = (\theta(1), \theta(2), \ldots, \theta(q))'$. The quantity

$$\mathbf{E}[\mathrm{tr}((K_Y - K_{\hat{\theta}})^2)] \tag{3.6.1}$$

is the total mean squared error between the true and the estimated values of the covariance functions. The vector $\hat{\theta}$ minimising this quantity is the estimator in (3.3.8). In particular, it follows that

$$\mathrm{tr}(K_i(K_Y - K_{\mathbf{E}[\hat{\theta}]})) = \mathrm{tr}(K_i(K_Y - \sum_{j=1}^{q} \mathbf{E}[\hat{\theta}_j]K_j)) = 0. \tag{3.6.2}$$

Putting $K_R = K_Y - \sum_{k=1}^{q} \mathbf{E}[\hat{\theta}_i]K_i$, one obtains

$$\begin{aligned}
\mathbf{E}[\mathrm{tr}((K_Y - K_{\hat{\theta}})^2)] &= \mathbf{E}\left[\mathrm{tr}\left(\left(\sum_{i=1}^{q}(\mathbf{E}[\hat{\theta}_i] - \hat{\theta}_i)K_i + K_R\right)^2\right)\right] \\
&= \sum_{i=1}^{q}\sum_{j=1}^{q} cov(\hat{\theta}_i, \hat{\theta}_j)\,\mathrm{tr}(K_iK_j) + \mathrm{tr}(K_R^2) \\
&= \sum_{i=1}^{q}\sum_{j=1}^{q} cov(\hat{\theta}_i, \hat{\theta}_j)\,\mathrm{tr}(K_iK_j) + \mathrm{tr}\left(\left(K_Y - \sum_{i=1}^{q}\mathbf{E}[\hat{\theta}_i]K_i\right)^2\right) \\
&= \sum_{i=1}^{q}\sum_{j=1}^{q} cov(\hat{\theta}_i, \hat{\theta}_j)\,\mathrm{tr}(K_iK_j) - \sum_{i=1}^{q}\sum_{j=1}^{q}\mathbf{E}[\hat{\theta}_i]\mathbf{E}[\hat{\theta}_j]\,\mathrm{tr}(K_iK_j) + \mathrm{tr}(K_Y^2).
\end{aligned}$$

The last term in the last expression does not depend on the model, so it can be dropped for the purpose of comparing different models. Therefore, the goal is to find the model (defined by the parameters q, a_i and b_i for $1 \leq i \leq q$) which

minimises the criterion

$$S = \sum_{i=1}^{q}\sum_{j=1}^{q}(cov(\hat{\theta}_i,\hat{\theta}_j) - \mathbf{E}[\hat{\theta}_i]\mathbf{E}[\hat{\theta}_j])\,\mathrm{tr}(K_iK_j)$$

$$= \mathrm{tr}\left(\left[cov(\hat{\theta}_i,\hat{\theta}_j) - \mathbf{E}[\hat{\theta}_i]\mathbf{E}[\hat{\theta}_j]\right][\mathrm{tr}(K_iK_j)]\right)$$

$$= \sum_{i=1}^{q}\sum_{j=1}^{q}(2cov(\hat{\theta}_i,\hat{\theta}_j) - \mathbf{E}[\hat{\theta}_i\hat{\theta}_j])\,\mathrm{tr}(K_iK_j)$$

$$= \mathrm{tr}\left(\left[2cov(\hat{\theta}_i,\hat{\theta}_j) - \mathbf{E}[\hat{\theta}_i\hat{\theta}_j]\right][\mathrm{tr}(K_iK_j)]\right). \quad (3.6.3)$$

In practice the values $cov(\hat{\theta}_i,\hat{\theta}_j), \mathbf{E}[\hat{\theta}_j]$, and $\mathbf{E}[\hat{\theta}_i\hat{\theta}_j]$ are not known and must be estimated. Clearly, $\mathbf{E}[\hat{\theta}_i\hat{\theta}_j]$ can be estimated without bias by $\hat{\theta}_i\hat{\theta}_j$. Constructing an estimator of $cov(\hat{\theta}_i,\hat{\theta}_j)$ is harder. In the remainder of this paper, the process Y will be assumed Gaussian. In that case one has

$$var(\hat{\theta}_n) = 2[\mathrm{tr}(K_{i,n}K_{j,n})]^{-1}[\mathrm{tr}(K_{i,n}K_{Y,n}K_{j,n}K_{Y,n})][\mathrm{tr}(K_{i,n}K_{j,n})]^{-1} \quad (3.6.4)$$

which leads to the plug-in estimate

$$\widehat{var(\hat{\theta}_n)} = 2[\mathrm{tr}(K_{i,n}K_{j,n})]^{-1}[\mathrm{tr}(K_{i,n}K_{\hat{\theta}_n,n}K_{j,n}K_{\hat{\theta}_n,n})][\mathrm{tr}(K_{i,n}K_{j,n})]^{-1}. \quad (3.6.5)$$

The plug-in estimate, though biased, leads to a computable approximate criterion, given by

$$\hat{S} = \mathrm{tr}\left(4[\mathrm{tr}(K_{i,n}K_{j,n})]^{-1}[\mathrm{tr}(K_{i,n}K_{\hat{\theta}_n,n}K_{j,n}K_{\hat{\theta}_n,n})] - [\hat{\theta}_i\hat{\theta}_j][\mathrm{tr}(K_iK_j)]\right). \quad (3.6.6)$$

The use of the criterion is now illustrated in a simulation study. A random Gaussian process in two dimensions is simulated on a square with side of length two. The simulated process will have known zero mean and the covariance function given by

$$\mathbf{C}_Y(\rho) = c\exp(-\rho^2/a) \quad (3.6.7)$$

model	(q)	(a_1, b_1)	(a_2, b_2)	(a_3, b_3)	S_{20}	S_{80}
A	2	(0, 6.111)	(6.111, 10)	-	-15.453	-115.935
B	3	(0, 6.830)	(6.830, 11.178)	(11.178, 15)	-14.651	-126.633
C	2	(0, 9.166)	(9.166, 15)	-	-17.561	-133.55

Table 3.2: *Three models to be fitted.*

with spectral density given by

$$\Psi_Y(\lambda) = \frac{ca}{2} \exp(-a\lambda^2/4). \tag{3.6.8}$$

The parameters a and c equal 0.2 and 1.0 respectively. The true model is assumed unknown, and three different spectral component additive models are fitted to the simulated data. The function $\Psi(\lambda)$ in (3.3.4) is assumed constant. Model orders considered are $q = 2$ and $q = 3$. The model parameters are given in Table 3.2. Model A was constructed to have two components and the spectral support $[0, 10]$, while models B and C were constructed to contain three and two components, respectively, and to have spectral supports $[0, 15]$ and $[0, 10]$, respectively. In all cases the end-points for the spectral components were selected using the method described in Section 3.5. The entire procedure is repeated twice, once with a sample of size twenty, once with a sample of size eighty. In both cases the sampling locations are initially obtained by a random sampling on the square and thereafter remain fixed throughout the study. Knowing the true covariance function, it is possible to compute the exact criterion (3.6.3) for all three models. Their values are given in the last two columns of Table 3.2, with S_n denoting the criterion computed for the sample of size n. The model which minimises the criterion would be considered optimal.

In practice, the true covariance is unknown and one can only compute the estimated criterion (3.6.6). To assess how well the estimated criterion performs, for each model one hundred samples of size twenty were simulated and (3.6.6) was computed. The same was repeated for samples of size eighty. In each simulation the three models were then compared pairwise (A to B, B to C and A to C) using the approximate criterion (3.6.6). Since the true model is known, the models A, B and C can also be compared pairwise based on the exact criterion (3.6.3). The frequency with which the pairwise comparisons based on the approximate criterion agreed with the comparison based on the exact criterion was assessed. The results are summarised in Table 3.3. The entries below the diagonal give the number of simulations (out of a hundred), where model selection based on $\hat{S}_{20}$ would give the same result as selection based on the (unknown in practice) exact criterion S_{20}, while the entries above the diagonal contain analogous information for $\hat{S}_{80}$ versus S_{80}. The agreement seems remarkably good.

For both sample sizes, model C is better than the other two. However, for $n = 80$, model B is better than model A. Figures 3.7 - 3.9 compare the true spectral density of the process with the mean spectral densities of the models A, B and C, computed from

$$\mathbf{E}[\Psi_{\hat{\theta}}(\lambda)] = \sum_{i=1}^{q} \mathbf{E}[\hat{\theta}(i)]\Psi_i(\lambda) \tag{3.6.9}$$

where

$$\mathbf{E}[\hat{\theta}] = [\mathrm{tr}(K_{i,n}K_{j,n})]^{-1}[\mathrm{tr}(K_{i,n}K_{Y,n})], \tag{3.6.10}$$

as follows from (3.3.8). As can be observed from Figures 3.7 - 3.9, a good portion of the spectral density of the true covariance function extends beyond $\lambda = 10$. Truncating the spectral density to $\lambda \leq 10$ introduces a large bias. For a large sample, this bias contributes more to the mean squared error than the increase

	model A	model B	model C
model A	-	93	98
model B	76	-	92
model C	97	100	-

Table 3.3: *Percentages of accurate pairwise comparisons based on the estimated criterion (comparison based on the frequency with which the approximate criterion agreed with that based on the exact criterion). Entries below the diagonal are for* $n = 20$, *above the diagonal for* $n = 80$.

in variance caused by the addition of another component to the model. Hence model B becomes better than model A. However, model C has the same spectral support as model B and one fewer component. Its lower variance makes it better than model B. Figures 3.10 - 3.12 compare the true covariance function of the process with the mean of the estimates of covariance functions from models A, B and C, computed from

$$\mathbf{E}[\mathbf{C}_{\hat{\theta}}(\rho)] = \sum_{i=1}^{q} \mathbf{E}[\hat{\theta}(i)]\mathbf{C}_i(\rho) \tag{3.6.11}$$

where $\mathbf{E}[\hat{\theta}]$ is given by (3.6.10). A visual inspection of Figures 3.10 - 3.12 reveals that the mean of model B indeed seems closer to the true model than that of model C, but the latter seems acceptable, particularly at small distances.

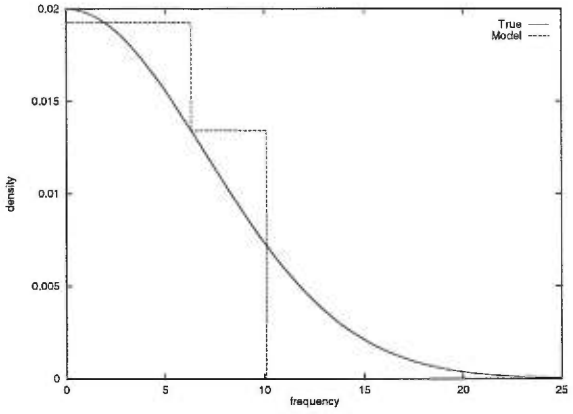

Figure 3.7: *Mean spectral density of model A fit versus that of the true model (n=80).*

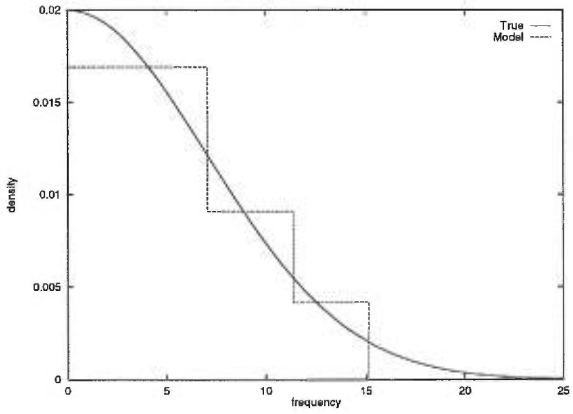

Figure 3.8: *Mean spectral density of model B fit versus that of the true model (n=80).*

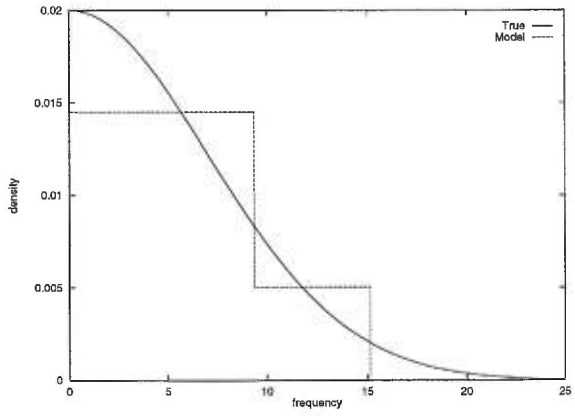

Figure 3.9: *Mean spectral density of model C fit versus that of the true model (n=80).*

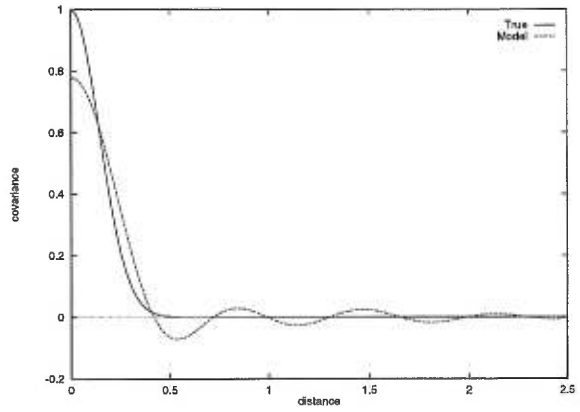

Figure 3.10: *Mean estimated covariance function with model A versus the true model (n=80).*

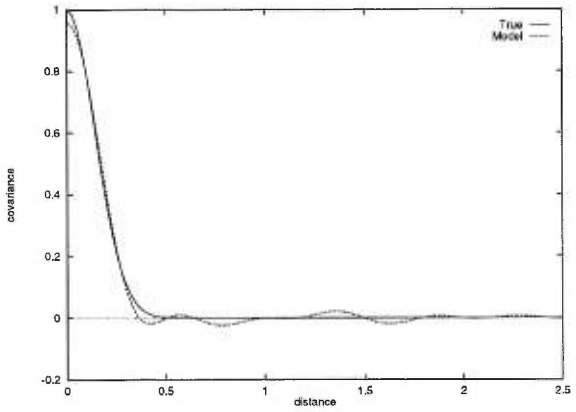

Figure 3.11: *Mean estimated covariance function with model B versus the true model (n=80).*

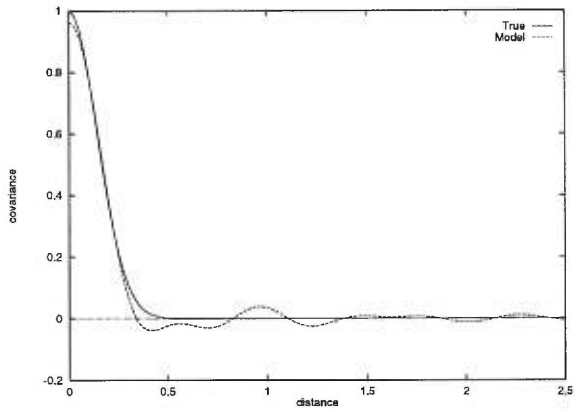

Figure 3.12: *Mean estimated covariance function with model C versus the true model (n=80).*

3.7. CASE STUDY

The methods of the previous sections are now applied to the data set of Davis (1973), representing 52 surface elevation measurements. The first step is recognising the need for modelling the process mean. Two different models for the mean will be used. In the first one, the mean of the process will be assumed to be an unknown constant, which leads to the regressor matrix X_{const} in (3.3.1). In the other, the mean of the process is assumed to be an arbitrary unknown linear trend, which leads to the appropriate regressor matrix X_{lin}. The matrices X_{const} and X_{lin} are obtained in the same way as described in Section 3.4.

3.7.1. Spectral support determination

The next step involves determining the support of the spectral density of the process, as described in Section 3.4. The procedure is performed with both sets of residuals — that resulting from X_{const}, as well as that resulting from using X_{lin}. A grid of 200 values was used to estimate the spectral density and the smallest non-zero distance found in the sample is $\rho_{min} = 0.2$, hence one obtains $\nu_{max} = 12.024$. The results are shown in Figures 3.13 and 3.14. It turns out that the estimated spectral density is essentially zero for ν much smaller than $\nu_{max} = 12.024$, and therefore the plots in Figures 3.13 and 3.14 are truncated at $\nu = 6$. It appears that the spectral support for the residuals from fitting a linear trend may be larger than that for the residuals from the constant mean model. The former will be assumed to be adequately covered by the interval $[0, 5]$, while the interval $[0, 2.5]$ will be used for the latter.

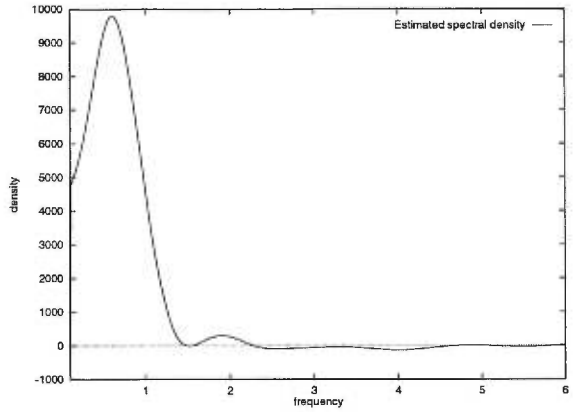

Figure 3.13: *Estimated support of spectral density with the constant mean model.*

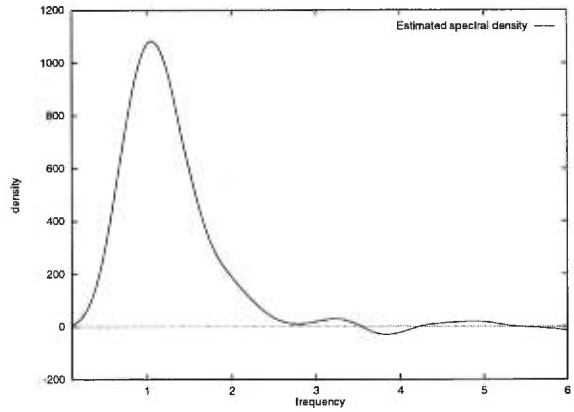

Figure 3.14: *Estimated support of spectral density with the linear trend mean model.*

model	(q)	(a_1, b_1)	(a_2, b_2)	(a_3, b_3)	(a_4, b_4)	W
A1	4	(0, 0.922)	(0.922, 1.509)	(1.509, 2.025)	(2.025, 2.5)	no
A2	3	(0, 1.138)	(1.138, 1.863)	(1.863, 2.5)	-	no
A3	2	(0, 1.528)	(1.528, 2.5)	-	-	no
A4	3	(0, 1.528)	(1.528, 2.5)	-	-	yes

Table 3.4: *Four covariance models to be fitted together with the constant model for the mean.*

3.7.2. Model building

The next step is building a covariance model to be fitted. Four models are considered for each model of the process mean. Table 3.4 describes the models used together with the constant mean model, while Table 3.5 shows the models used together with the linear trend mean model. The endpoint selection was performed based on the procedure described in Section 3.5. The last columns in Tables 3.4 and 3.5 indicate whether the nugget effect component is included in the model. The nugget effect is a common modification in geostatistics (see, for example Cressie, 1993), where a component of the form

$$W(\rho) = \begin{cases} 1 & \text{if } \rho = 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.7.1}$$

is added, leading to a model of order $q + 1$:

$$\mathbf{C}_\theta(\rho) = \gamma W(\rho) + \sum_{i=1}^{q} \theta(i)\mathbf{C}_i(\rho). \tag{3.7.2}$$

For details on projection estimation in models including the nugget effect component, the reader is referred to Powojowski (1999a).

model	(q)	(a_1, b_1)	(a_2, b_2)	(a_3, b_3)	(a_4, b_4)	W
B1	4	(0, 1.845)	(1.845, 3.019)	(3.019, 4.051)	(4.051, 5)	no
B2	3	(0, 2.277)	(2.277, 3.726)	(3.726, 5)	-	no
B3	2	(0, 3.055)	(3.055, 5)	-	-	no
B4	3	(0, 3.055)	(3.055, 5)	-	-	yes

Table 3.5: *Four covariance models to be fitted together with the linear trend model for the mean.*

3.7.3. Estimation and model comparison

Since the mean of the process is not known and has to be estimated, the estimator (3.3.15) will be used. Two different models for the mean of the process Y will be applied. They will be the same models as in Section 3.4, that is X_{const} which models a constant mean of Y and X_{lin}, which models any linear trend.

The estimates $\hat{\theta}_n(i)$ will also be constrained to be nonnegative. This is necessary for the estimated model to be a valid covariance function, that is a positive-definite function. Thus in practice the estimator (3.3.15) will be computed for all submodels (by excluding certain additive components) and the model minimising the objective function in (3.3.16), while satisfying $\hat{\theta}_n(i) \geq 0$ for all its components, will be retained. Thus the estimator $\hat{\theta}$ is the vector that minimises the objective function

$$\mathrm{tr}((e_n e_n' - \sum_{i=1}^{q} \theta(i) U_{i,n})^2)$$

subject to the constraints $\theta(i) \geq 0$. In other words, the optimisation of the function in (3.3.16) will be performed over the positive quadrant $\mathbf{R}_+^d$, rather than over the entire space $\mathbf{R}^d$.

model	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	W	$\hat{S}$
A1	6130.81	0	358.062	0	-	7.366e+09
A2	5941.79	0	158.11	-	-	1.878e+09
A3	5604.31	0	-	-	-	-1.871e+09
A4	5604.31	0	-	-	0	-1.780e+09

Table 3.6: *Estimated parameters for the four covariance models fitted together with the constant model for the mean.*

model	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	W	$\hat{S}$
B1	1587.29	0	66.378	227.463	-	-1.571e+08
B2	1602.08	0	174.776	-	-	-1.771e+08
B3	1494.11	70.776	-	-	-	-1.335e+08
B4	1494.11	70.776	-	-	0	-1.215e+08

Table 3.7: *Estimated parameters for the four covariance models fitted together with the linear trend model for the mean.*

The results are shown in Tables 3.6 and 3.7. The resulting estimated covariance functions for models A1 - A3 and B1 - B3 are shown in Figures 3.15 - 3.20. In those figures, the products of residuals are marked as individual dots with coordinates $(\| x_k - x_l \|, e(k)e(l))$, where $1 \leq k \leq 52$, $1 \leq j \leq 52$ and $e(k) = Y(k) - \hat{Y}(k)$ is the residual computed as in (3.3.14). The curves plotted in the figures are $(\rho, \mathbf{C}_{\hat{\theta}}(\rho))$, where the estimated values of $\hat{\theta}$ are given in Tables 3.6 and 3.7. Table 3.6 indicates that for the constant mean model the optimal covariance model is A3. However, only the first additive component has a nonzero coefficient, which makes the best fit a single-component model. From Table 3.7

one observes that the optimal model is B2, which also has one of its three terms eliminated. In neither case did the inclusion of the nugget effect lead to a better model. (The plots of the fits with nugget effect are not shown, since they are identical to models A3 and A4.) It would seem that the curve obtained with model A1 may be a little too high. This may be due to the fact that this model contains a very low-frequency component, which may be poorly estimated.

While the plots seem to indicate that the estimated model follows the data fairly well, Figures 3.15 - 3.20 may be misleading. Let P denote the projection in (3.3.14). The mean of $e(k)e(l)$ is $(PK_YP)(k,l)$, rather than $K_Y(k,l)$, which may be estimated by $(PK_{\hat{\theta}}P)(k,l)$ rather than by $K_{\hat{\theta}}(k,l)$. It may therefore be informative to plot the points $(\| x_k - x_l \|, (PK_{\hat{\theta}}P)(k,l))$ along with the products of residuals. This is carried out for models A3 and B2 and the resulting plots are shown in Figures 3.21 and 3.22. One observes that in those two cases the plots of fitted covariograms (Figures 3.17 and 3.19) and the plots of estimated covariances of residuals (Figures 3.21 and 3.22) are similar.

Comparing the results of this section with models previously fitted to the same data (e.g. Ripley, 1988, Wackernagel, 1995), one observes two main differences. Firstly, the models for the mean of the process used by those authors are linear trends or quadratic surfaces. It seems that the constant mean model was not attempted. No reasons are given for this omission, but it appears that the products of the residuals from the constant mean model are not very well estimated by a covariance model which is positive everywhere, such as the exponential or the Gaussian model used by the authors. This leads to the other major difference, namely the presence of negative covariances in the estimated model. Visual examination of Figures 3.15 - 3.20 suggests that negative covariances are plausible. Spectral component additive models are capable of capturing negative

covariances, while the exponential or the Gaussian model is not. The flexibility of spectral component additive models is illustrated in Powojowski (1999b), where comparisons are made to a number of standard parametric models. The results of this section further indicate that spectral component additive models can capture detailed features of the covariogram, which may be difficult to reproduce with standard parametric models.

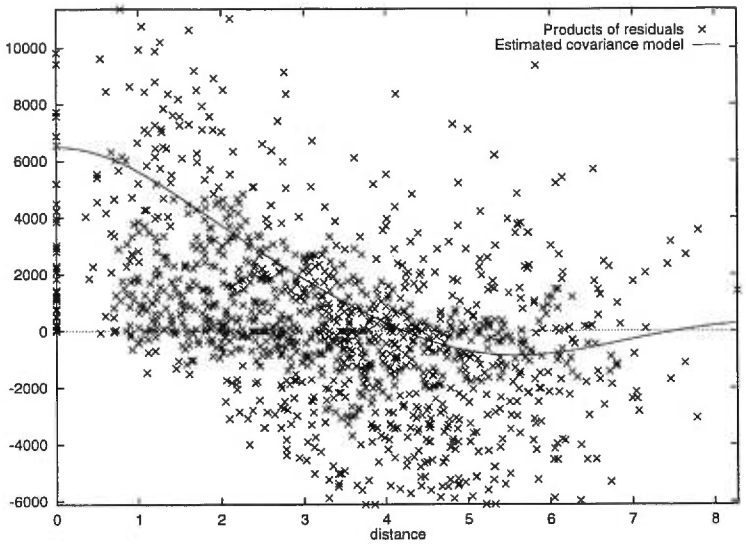

Figure 3.15: *Model A1.*

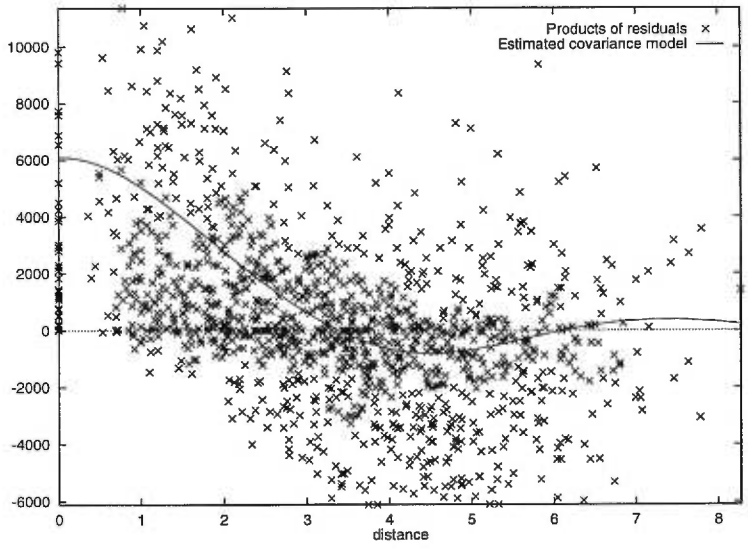

Figure 3.16: *Model A2.*

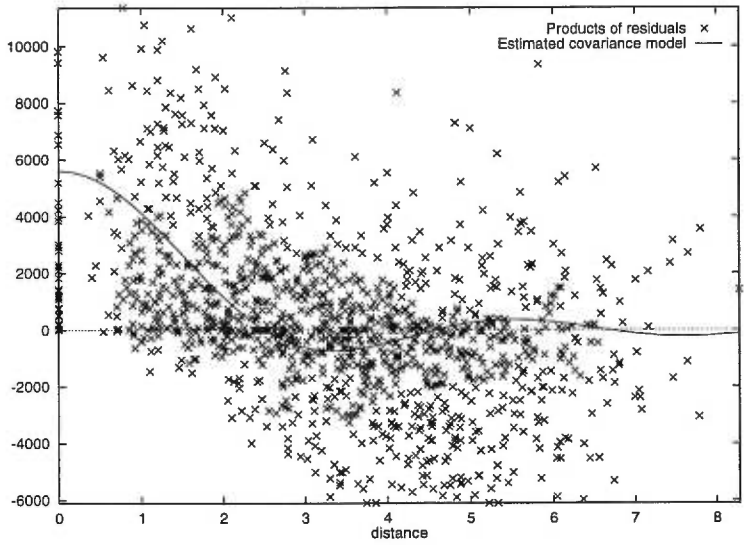

Figure 3.17: *Model A3.*

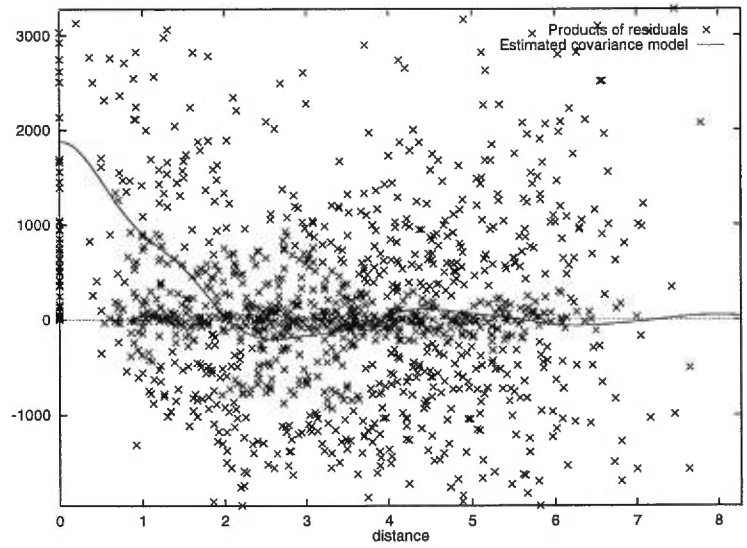

Figure 3.18: *Model B1.*

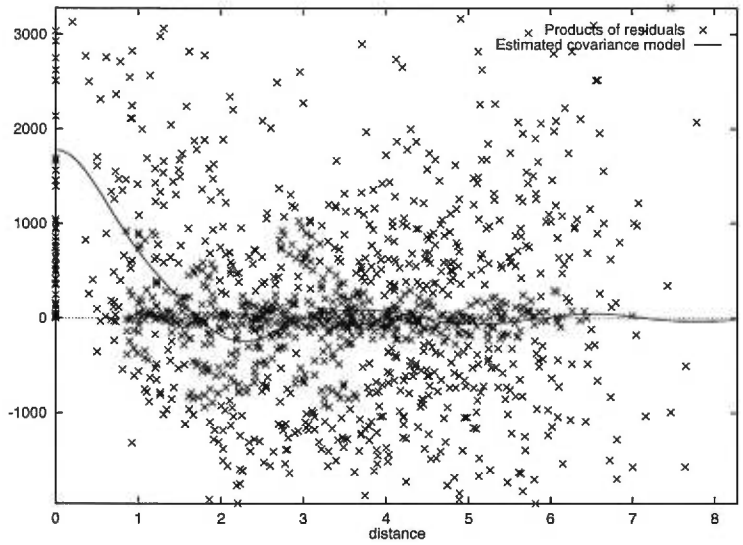

Figure 3.19: *Model B2.*

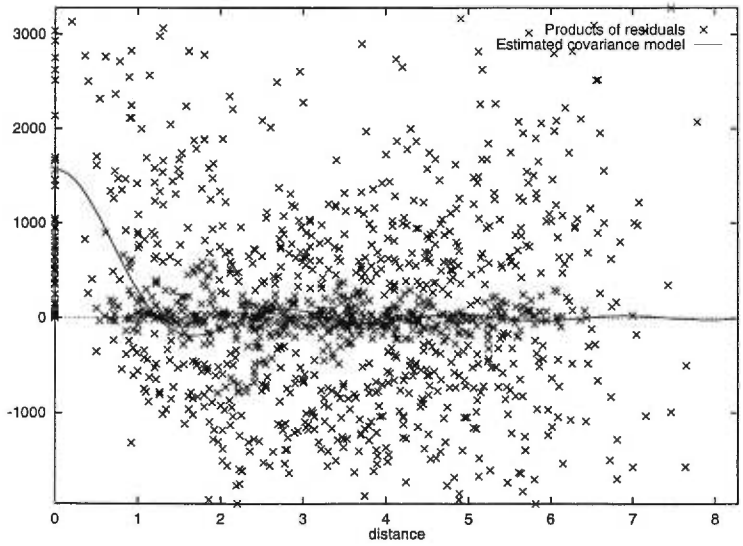

Figure 3.20: *Model B3.*

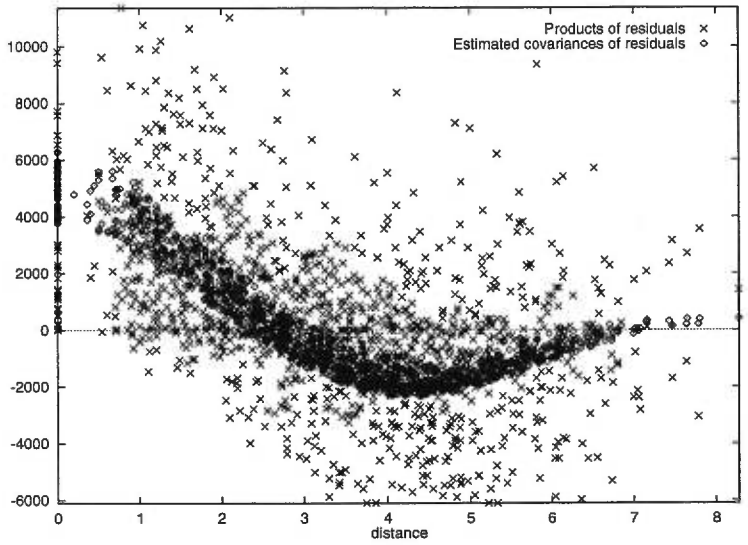

Figure 3.21: *Products of residuals and estimated covariances of residuals (model A3).*

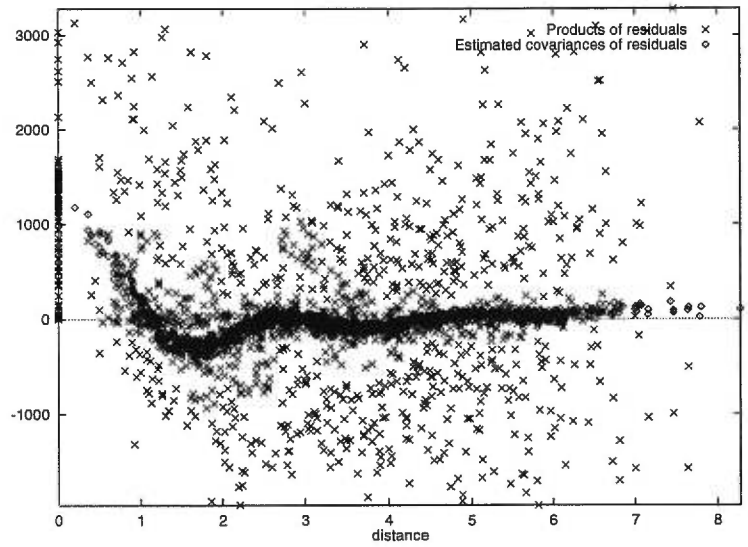

Figure 3.22: *Products of residuals and estimated covariances of residuals (model B2).*

3.8. Conclusion

The paper examines model selection issues involved in the modelling and estimation of the covariance functions of isotropic random processes. The class of spectral component additive models, introduced in Powojowski (1999b), is considered in detail. A data-driven procedure for determining the spectral support is proposed. Another procedure is suggested for selecting the end-points of the component support intervals. Finally, a criterion is derived for deciding between competing models. These techniques are tested using synthetic data and are seen to produce good results.

The proposed techniques are applied to a real data set of Davis (1973). The data set has been examined by a number of researchers, among others Ripley (1988) and Wackernagel (1995). The spectral component additive models appear to bring a few new elements to the analysis of the data set. One is their ability to capture fine features of the covariogram, such as regions of negative covariance. Another is the information about the spectral density of the covariogram, such as its effective support and relative contributions of different frequencies. While the spectral methods have not been applied extensively in geostatistics, it has been suggested that the reason for this is the requirement for the observations to be regularly spaced on a lattice (Chiles and Delfiner, 1999). Spectral component additive models provide a way of estimating the spectral densities from irregularly spaced data.

In addition, the implementation of the estimation and model selection procedures is very straightforward and usually requires only linear algebra tools. The model selection techniques presented here seem highly practical in working with

real data. Additional work is required in establishing the impact of estimating the mean of the process on the estimation of low-frequency covariance components. In addition, criteria for selection of the mean model would be very useful.

REFERENCES

ADLER R.J. (1980). *The Geometry of Random Fields.* John Wiley & Sons.

CHILES, J.P., DELFINER, P. (1999). *Geostatistics: modeling spatial uncertainty.* John Wiley & Sons.

CRESSIE, N.A.C (1993). *Statistics for Spatial Data.* Revised edition. John Wiley & Sons.

DAVIS, J.C. (1973). *Statistics and Data Analysis in Geology.* John Wiley & Sons.

POWOJOWSKI, M. (1999a). Additive Covariogram Models and Estimation through Projections. In preparation.

POWOJOWSKI, M. (1999b). Estimation of the Covariogram of an Isotropic Process through Spectral Component Additive Models. In preparation.

RIPLEY, B.D. (1988). *Statistical Inference for Spatial Processes.* Cambridge University Press.

SCHOENBERG, I. J. (1938). Metric spaces and completely monotone functions. *The Annals of Mathematics* **39** 811-841.

SNEDDON, I.N (1972). *The Use of Integral Transforms.* McGraw-Hill.

WACKERNAGEL, H. (1995). *Multivariate geostatistics.* Springer-Verlag.

CONCLUSION

Cette thèse présente une nouvelle approche à l'estimation de la fonction de covariance. Elle propose une nouvelle classe de modèles de covariance pour des processus isotropes, ainsi qu'un nouvel estimateur pour des modèles additifs. Les trois articles traitent de la construction du modèle, la sélection du meilleur modèle et l'estimation des paramètres. Des propriétés théoriques de l'estimateur sont obtenues. La mise en oeuvre de l'approche est illustrée par une application aux données de Davis (1973). Cette nouvelle approche présente de nombreux avantages, déjà mentionnés dans le texte. En conclusion, on mentionne des généralisations possibles de l'approche.

Une généralisation permettant l'estimation du variogramme plutôt que du covariogramme est en principe possible. Le variogramme est défini par

$$\gamma(x_1, x_2) = \text{var}(Y(x_1) - Y(x_2)). \tag{3.8.1}$$

Au lieu de considérer l'ensemble de matrices dont les éléments sont $K(k,l) = \text{cov}(Y(x_k), Y(x_l))$, on peut considérer les matrices définies par $L(k,l) = \gamma(Y(x_k), Y(x_l))$. En définissant un produit scalaire dans l'espace de matrices symétriques (possiblement le même que dans le premier article) on peut obtenir un estimateur par projection de paramètres d'un modèle de la forme

$$\gamma_\theta = \sum_{i=1}^{q} \theta(i)\gamma_i \tag{3.8.2}$$

où les γ_i sont des modèles de variogramme déterminés.

Il n'est pas clair que les propriétés asymptotiques peuvent être reproduites pour cet estimateur. La difficulté provient du fait qu'un variogramme ne converge pas vers zéro lorsque la distance augmente.

Une autre généralisation facile est l'application de l'estimateur par projection à des observations qui ne sont pas les valeurs du processus Y même, mais plutôt une transformation linéaire $Z = GY$ de Y. Si la matrice G n'est pas inversible, les valeurs de Y ne peuvent pas être récupérées à partir de Z. De plus, Z peut être non stationnaire même si Y est stationnaire. Si le covariogramme de Y peut être modélisé par

$$K_Y = \sum_{i=1}^{q} \theta(i) K_i \tag{3.8.3}$$

il s'ensuit que

$$K_Z = \sum_{i=1}^{q} \theta(i) G K_i G' \tag{3.8.4}$$

L'estimateur par projection peut être employé avec des observations de Z afin d'estimer les paramètres $\theta(i)$ du modèle (3.8.4). Cela produit un estimé $\sum_{i=1}^{q} \hat{\theta}(i) K_i$ de K_Y.

Finalement, une généralisation aux modèles non additifs (non linéaires) est possible. L'estimateur par projection peut être défini comme la valeur θ qui minimise la norme

$$\| YY' - K_\theta \| \tag{3.8.5}$$

où YY' est la matrice de produits des observations de Y ou de leurs résidus, tandis que K_θ est une matrice de covariance obtenue par le modèle. Si K_θ dépend de θ de façon linéaire et la norme est définie par un produit scalaire, on obtient le cas considéré dans cette thèse. Cependant, si K_θ n'est pas additif, on peut toujours définir un estimateur $\hat{\theta}$ comme le paramètre qui minimise (3.8.5).

Certains avantages de l'estimateur par projection sont préservés, par exemple le calcul du variogramme ou covariogramme empirique n'est pas nécessaire et l'estimateur n'exige pas que le processus soit stationnaire. D'autres propriétés ne sont pas facilement retrouvées car des expressions simples ne sont pas généralement disponibles pour la moyenne ou la variance de l'estimateur.

Annexe A

PROOFS OF THE RESULTS

Proof of Lemma 1.5.1:

Let $\epsilon > 0$. For any positive integers n and k, let $\{A_1^k, \ldots, A_k^k\}$ denote the partition of D as in Definition 1.5.2. Then

$$\sum_{l=1}^{n} \phi_1(x, x_l)\phi_2(y, x_l) = \sum_{i=1}^{k} P_i^{n,k} \phi_1(x, \xi_i^{n,k})\phi_2(y, \xi_i^{n,k})$$

where $P_i^{n,k} = |\{l : 1 \leq l \leq n \text{ and } x_j \in A_i^k\}|$ and $\xi_i^{n,k}$ is a point in A_i^k. This is a simple consequence of the continuity of ϕ_1 and ϕ_2 and the connectivity of A_i^k, since there is a point $\xi_i^{n,k} \in A_i^k$ such that

$$\phi_1(x, \xi_i^{n,k})\phi_2(y, \xi_i^{n,k}) = \frac{1}{P_i^{n,k}} \sum_{\{l:1\leq l\leq n \text{ and } x_j \in A_i^k\}} \phi_1(x, x_l)\phi_2(y, x_l).$$

Hence

$$\begin{aligned} |R_k^n| &= \left| \sum_{l=1}^{n} \frac{1}{n} \phi_1(x, x_l)\phi_2(y, x_l) - \int_D \phi_1(x, \xi)\phi_2(y, \xi) f(\xi) d\xi \right| \\ &= \left| \sum_{i=1}^{k} \frac{1}{n} \frac{k}{\mu(D)} \mu(A_i^k) P_i^{n,k} \phi_1(x, \xi_i^{n,k})\phi_2(y, \xi_i^{n,k}) \right. \\ &\qquad \left. - \sum_{i=1}^{k} \int_{A_i^k} \phi_1(x, \xi)\phi_2(y, \xi) f(\xi) d\xi \right| \end{aligned}$$

since $\mu(A_i^k) = \mu(D)/k$. Putting $\psi(x,y,\xi) = \phi_1(x,\xi)\phi_2(y,\xi)$ one obtains

$$
\begin{aligned}
|\, R_k^n \,| &= \left| \sum_{i=1}^{k} \int_{A_i^k} \left(\frac{k}{\mu(D)} \frac{P_i^{k,n}}{n} \psi(x,y,\xi_i^k) - \psi(x,y,\xi) f(\xi) \right) d\xi \right| \\
&= \left| \sum_{i=1}^{k} \psi(x,y,\xi_i^k) \int_{A_i^k} \left(\frac{k}{\mu(D)} \frac{P_i^{k,n}}{n} - f(\xi) \right) d\xi \right. \\
&\qquad \left. + \sum_{i=1}^{k} \int_{A_i^k} f(\xi)(\psi(x,y,\xi_i^k) - \psi(x,y,\xi)) d\xi \right|.
\end{aligned}
$$

Since the function ψ is continuous on the compact domain D^3, it is uniformly continuous on D^3. Thus there exists $\delta > 0$ such that

$$
\| \, (x_1,y_1,z_1) - (x_2,y_2,z_2) \, \| < \delta \Rightarrow |\psi(x_1,y_1,z_1) - \psi(x_2,y_2,z_2)| < \epsilon
$$

for all (x_1,y_1,z_1) and (x_2,y_2,z_2) in D^3. Since the diameters of the sets A_i^k converge uniformly to 0 as $k \to \infty$, for any $\delta > 0$ it is possible to find K such that

$$
k > K \Rightarrow \operatorname{diam}(A_i^k) < \delta, \qquad i = 1, \ldots, k.
$$

Hence it follows that for $k > K$

$$
\begin{aligned}
&\left| \sum_{i=1}^{k} \int_{A_i^k} f(\xi)(\psi(x,y,\xi_i^k) - \psi(x,y,\xi)) d\xi \right| \\
&\qquad \leq \sum_{i=1}^{k} \int_{A_i^k} f(\xi) \left| \, (\psi(x,y,\xi_i^k) - \psi(x,y,\xi)) \, \right| d\xi \\
&\qquad\qquad \leq \epsilon \sum_{i=1}^{k} \int_{A_i^k} f(\xi) d\xi = \epsilon \int_D f(\xi) d\xi = \epsilon.
\end{aligned}
$$

It is important to note that the choice of K is not dependent on x, y or n, but solely on ϵ. This leaves the term

$$\left|\sum_{i=1}^{k}\psi(x,y,\xi_i^k)\int_{A_i^k}\left(\frac{k}{\mu(D)}\frac{P_i^{k,n}}{n}-f(\xi)\right)d\xi\right|$$
$$\leq\sum_{i=1}^{k}\left|\psi(x,y,\xi_i^k)\left(\frac{P_i^{k,n}}{n}-\int_{A_i^k}f(\xi)d\xi\right)\right|\leq M\sum_{i=1}^{k}\left|\left(\frac{P_i^{k,n}}{n}-\int_{A_i^k}f(\xi)d\xi\right)\right|$$

where M is a bound on $|\psi(x,y,z)|$ over D^3, which exists by the continuity of ψ and the compactness of D^3. Thus the choice of ϵ determines K, and given any $k > K$, by the definition of an in-fill sampling configuration, there is a positive integer N such that $n > N$ guarantees, for all $i = 1, \ldots, k$

$$\left|\frac{|\{l : 1 \leq l \leq n \text{ and } x_l \in A_i^k\}|}{n}-\int_{A_i^k}f(u)du\right|<\frac{\epsilon}{Mk}.$$

It follows that for any $\epsilon > 0$, K and N can be found such that

$$n > N,\ k > K \Rightarrow |R_k^n| < 2\epsilon$$

which completes the proof.

Proof of Lemma 1.5.2:

The notation from the proof of the previous result will be used, except for $\psi(x,y)$, which will denote $\phi_1(x,y)\phi_2(x,y)$ and will be defined on D^2. Let $B_{i_1,i_2}^k = A_{i_1}^k \times A_{i_2}^k$. It follows that $B_{i_1,i_2}^k \subseteq D^2$ is connected, $\mu(B_{i_1,i_2}^k) = \mu(D)/k^2$, $\text{diam}(B_{i_1,i_2}^k) \to 0$ uniformly (in i_1 and i_2) as $k \to \infty$. Moreover, from elementary properties of limits it follows that for any positive integer k, $\forall\epsilon > 0\ \exists N$ such that

$\forall\ 1 \leq i_1 \leq k, 1 \leq i_2 \leq k$

$$n > N \Rightarrow \left| \frac{|\{(l_1, l_2) : 1 \leq l_1 \leq n\ , 1 \leq l_2 \leq n \text{ and } (x_{l_1}, x_{l_2}) \in B^k_{i_1,i_2}\}|}{n} - \int_{A^k_{i_1}} f(u)du \int_{A^k_{i_2}} f(u)du \right| < \epsilon. \quad \text{(A.0.6)}$$

Putting $P^{k,n}_{i_1,i_2} = |\{(l_1, l_2) : 1 \leq l_1 \leq n\ , 1 \leq l_2 \leq n \text{ and } (x_{l_1}, x_{l_2}) \in B^k_{i_1,i_2}\}|$ one obtains $P^{k,n}_{i_1,i_2} = P^{k,n}_{i_1} P^{k,n}_{i_2}$. Again, the following holds

$$\sum_{l_1=1}^{n}\sum_{l_2=1}^{n} \phi_1(x_{l_1}, x_{l_2})\phi_2(x_{l_1}, x_{l_2}) = \sum_{i_1=1}^{k}\sum_{i_2=1}^{k} P^{n,k}_{i_1,i_2}\phi_1(\xi^{n,k}_{i_1,i_2}, \eta^{n,k}_{i_1,i_2})\phi_2(\xi^{n,k}_{i_1,i_2}, \eta^{n,k}_{i_1,i_2})$$

where $(\xi^{n,k}_{i_1,i_2}, \eta^{n,k}_{i_1,i_2}) \in A^k_{i_1} \times A^k_{i_2}$.

Let $\epsilon > 0$. One has

$$
\begin{aligned}
|R_k^n| = \Bigg| & \sum_{l_1=1}^{n}\sum_{l_2=1}^{n}\frac{1}{n^2}\phi_1(x_{l_1},x_{l_2})\phi_2(x_{l_1},x_{l_2}) \\
& - \int_D\int_D \phi_1(\xi,\eta)\phi_2(\xi,\eta)f(\xi)f(\eta)d\xi d\eta \Bigg| \\
= & \Bigg| \sum_{i_1=1}^{k}\sum_{i_2=1}^{k}\frac{1}{n^2}\frac{k^2}{\mu(D)}\mu(B_{i_1,i_2}^k)P_{i_1,i_2}^{n,k}\psi(\xi_{i_1,i_2}^{n,k},\eta_{i_1,i_2}^{n,k}) \\
& - \sum_{i_1=1}^{k}\sum_{i_2=1}^{k}\int_{A_{i_1}^k}\int_{A_{i_2}^k}\psi(\xi,\eta)f(\xi)f(\eta)d\xi\, d\eta \Bigg| \\
= & \Bigg| \sum_{i_1=1}^{k}\sum_{i_2=1}^{k}\int_{A_{i_1}^k}\int_{A_{i_2}^k}\left(\frac{k^2}{\mu(D)}\frac{P_{i_1,i_2}^{k,n}}{n^2}\psi(\xi_{i_1,i_2}^{n,k},\eta_{i_1,i_2}^{n,k}) - \psi(\xi,\eta)f(\xi)f(\xi)\right)d\xi\, d\eta \Bigg| \\
\leq & \Bigg| \sum_{i_1=1}^{k}\sum_{i_2=1}^{k}\psi(\xi_{i_1,i_2}^{n,k},\eta_{i_1,i_2}^{n,k})\left(\frac{P_{i_1,i_2}^{k,n}}{n^2} - \int_{A_{i_1}^k}\int_{A_{i_2}^k} f(\xi)f(\eta)d\xi\, d\eta\right)\Bigg| \\
& + \Bigg| \sum_{i_1=1}^{k}\sum_{i_2=1}^{k}\int_{A_{i_1}^k}\int_{A_{i_2}^k} f(\xi)f(\eta)(\psi(\xi_{i_1,i_2}^{n,k},\eta_{i_1,i_2}^{n,k}) - \psi(\xi,\eta))d\xi\, d\eta \Bigg|.
\end{aligned}
$$

Again, the continuity of the function ψ on the compact domain D^2, guarantees uniform continuity on D^2. Thus there exists $\delta > 0$ such that

$$\| (x_1,y_1) - (x_2,y_2) \| < \delta \Rightarrow |\psi(x_1,y_1) - \psi(x_2,y_2)| < \epsilon$$

for all (x_1,y_1) and (x_2,y_2) in D^2. Since the diameters of the sets B_{i_1,i_2}^k converge uniformly to 0 as $k \to \infty$, for any $\delta > 0$ it is possible to find K such that

$$k > K \Rightarrow \operatorname{diam}(B_{i_1,i_2}^k) < \delta, \qquad i_1 = 1,\ldots,k,\ i_1 = 1,\ldots,k.$$

Hence it follows that for $k > K$

$$\left| \sum_{i_1=1}^{k} \sum_{i_2=1}^{k} \int_{A_{i_1}^k} \int_{A_{i_2}^k} f(\xi) f(\eta) (\psi(\xi_{i_1,i_2}^{n,k}, \eta_{i_1,i_2}^{n,k}) - \psi(\xi,\eta)) d\xi \, d\eta \right|$$

$$\leq \sum_{i_1=1}^{k} \sum_{i_2=1}^{k} \int_{A_{i_1}^k} \int_{A_{i_2}^k} |f(\xi) f(\eta)| \left(\psi(\xi_{i_1,i_2}^{n,k}, \eta_{i_1,i_2}^{n,k}) - \psi(\xi,\eta)\right) | d\xi \, d\eta$$

$$\leq \epsilon \sum_{i_1=1}^{k} \sum_{i_2=1}^{k} \int_{A_{i_1}^k} \int_{A_{i_2}^k} f(\xi) f(\eta) d\xi \, d\eta = \epsilon.$$

Again, one observes that K is not dependent on n, but solely on ϵ. This remaining term is bounded as follows:

$$\left| \sum_{i_1=1}^{k} \sum_{i_2=1}^{k} \psi(\xi_{i_1,i_2}^{n,k}, \eta_{i_1,i_2}^{n,k}) \left(\frac{P_{i_1,i_2}^{k,n}}{n^2} - \int_{A_{i_1}^k} \int_{A_{i_2}^k} f(\xi) f(\eta) d\xi \, d\eta \right) \right|$$

$$\leq M \sum_{i_1=1}^{k} \sum_{i_2=1}^{k} \left| \left(\frac{P_{i_1,i_2}^{k,n}}{n^2} - \int_{A_{i_1}^k} \int_{A_{i_2}^k} f(\xi) f(\eta) d\xi \, d\eta \right) \right|.$$

Where M is an upper bound on $|\psi(x,y)|, (x,y) \in D^2$. Since M and k are now fixed, it is possible to find N such that

$$n > N \Rightarrow \left| \left(\frac{P_{i_1,i_2}^{k,n}}{n^2} - \int_{A_{i_1}^k} \int_{A_{i_2}^k} f(\xi) f(\eta) d\xi \, d\eta \right) \right| < \frac{\epsilon}{Mk^2}$$

which guarantees $|R_k^n| < 2\epsilon$ as required.

Proof of Lemma 1.5.3:

It suffices to show that

$$|R_n| = \left| \sum_{k=1}^{n} \sum_{l=1}^{n} \frac{1}{n^2} \phi_{1,n}(x_k, x_l) \phi_{2,n}(x_k, x_l) - \phi_1(x_k, x_l) \phi_2(x_k, x_l) \right| \to 0$$

as $n \to \infty$. One observes:

$$|R_n| \leq \sum_{k=1}^{n}\sum_{l=1}^{n}\frac{1}{n^2}|\,\phi_{1,n}(x_k,x_l)(\phi_{2,n}(x_k,x_l)-\phi_2(x_k,x_l)) + \phi_2(x_k,x_l)(\phi_{1,n}(x_k,x_l)-\phi_1(x_k,x_l))\,|$$

$$\leq M\frac{1}{n^2}\left(\sum_{k=1}^{n}\sum_{l=1}^{n}|\phi_{2,n}(x_k,x_l)-\phi_2(x_k,x_l)|+\sum_{k=1}^{n}\sum_{l=1}^{n}|\phi_{1,n}(x_k,x_l)-\phi_1(x_k,x_l))|\right)$$

where M is a uniform bound on $|\phi_2|$ and $|\phi_{1,n}|$ on D^2, which exists by the compactness of D^2, the uniform convergence of $\{\phi_{1,n}\}$ and the continuity of all the functions involved. By the uniform convergence it follows that given an $\epsilon > 0$ a number N can be found such that

$$n > N \Rightarrow |\,\phi_{i,n}(x,y)-\phi_i(x,y))\,| < \frac{\epsilon}{M}$$

for $i = 1, 2$ and any $(x, y) \in D^2$. This leads to $|R_n| < 2\epsilon$ as required.

Proof of Lemma 1.5.4:

One has

$$\sum_{k_1=1}^{n}\sum_{k_2=1}^{n}\sum_{k_3=1}^{n}\sum_{k_4=1}^{n}\frac{1}{n^4}\phi_1(x_{k_1},x_{k_2})\phi_2(x_{k_2},x_{k_3})\phi_3(x_{k_3},x_{k_4})\phi_4(x_{k_4},x_{k_1})$$

$$=\sum_{k_1=1}^{n}\sum_{k_3=1}^{n}\frac{1}{n^2}\left(\sum_{k_2=1}^{n}\frac{1}{n}\phi_1(x_{k_1},x_{k_2})\phi_2(x_{k_2},x_{k_3})\right)\left(\sum_{k_4=1}^{n}\frac{1}{n}\phi_3(x_{k_3},x_{k_4})\phi_4(x_{k_4},x_{k_1})\right).$$

By Lemma 1.5.1

$$\sum_{k_2=1}^{n}\frac{1}{n}\phi_1(x_{k_1},x_{k_2})\phi_2(x_{k_2},x_{k_3}) \to \int_D \phi_1(x_{k_1},\lambda)\phi_2(\lambda,x_{k_3})f(\lambda)d\lambda$$

uniformly in (x_{k_1}, x_{k_3}) and

$$\sum_{k_4=1}^{n} \frac{1}{n}\phi_1(x_{k_3}, x_{k_4})\phi_2(x_{k_4}, x_{k_1}) \to \int_D \phi_1(x_{k_3}, \lambda)\phi_2(\lambda, x_{k_1})f(\lambda)d\lambda$$

uniformly in (x_{k_3}, x_{k_1}). Hence the result follows by Lemma 1.5.3.

Proof of Lemma 1.5.5:

Firstly, one observes that if A is a $q \times q$ symmetric non-negative definite matrix, then q diag$(A) - A$ is non-negative definite. Indeed, there exists a $q \times q$ matrix R such that $R'R = A$. Hence, for any q-dimensional vector x one has $(Rx)(j) = \sum_{i=1}^{q} R(j,i)x(i)$ and

$$\begin{aligned}
x'(q \text{ diag}(A) - A)x &= q\, x' \text{diag}(R'R)x - (Rx)'Rx \\
&= q \sum_{j=1}^{q}\sum_{i=1}^{q}(R(j,i))^2(x(i))^2 - \sum_{j=1}^{q}\left(\sum_{i=1}^{q} R(j,i)x(i)\right)^2 \\
&= \sum_{j=1}^{q}\left(q \sum_{i=1}^{q}(R(j,i))^2(x(i))^2 - \left(\sum_{i=1}^{q} R(j,i)x(i)\right)^2\right) \\
&= \sum_{j=1}^{q}\left(q \sum_{i=1}^{q}(R(j,i))^2(x(i))^2 - \sum_{i=1}^{q}\sum_{k=1}^{q} R(j,i)x(i)R(j,k)x(k)\right) \\
&\geq \sum_{j=1}^{q}\left(q \sum_{i=1}^{q}(R(j,i))^2(x(i))^2 \right. \\
&\quad \left. - \frac{1}{2}\sum_{i=1}^{q}\sum_{k=1}^{q}\left((R(j,i))^2(x(i))^2 + (R(j,k))^2(x(k))^2\right)\right) = 0.
\end{aligned}$$

Hence

$$q \text{ diag}\left([\text{cov}(Y_n'K_{i,n}Y_n, Y_n'K_{j,n}Y_n)]\right) - [\text{cov}(Y_n'K_{i,n}Y_n, Y_n'K_{j,n}Y_n)] \geq 0.$$

Moreover, since $\mathrm{var}(Y_n' K_{i,n} Y) \leq c\ \mathrm{tr}(K_{i,n} K_{Y,n} K_{i,n} K_{Y,n})$ one has

$$c\ \mathrm{diag}\Big([\mathrm{tr}(K_{i,n} K_{Y,n} K_{i,n} K_{Y,n})]\Big) - \mathrm{diag}\Big([\mathrm{cov}(Y_n' K_{i,n} Y_n, Y_n' K_{j,n} Y_n)]\Big) \geq 0.$$

It follows that

$$cq\ \mathrm{diag}\Big([\mathrm{tr}(K_{i,n} K_{Y,n} K_{i,n} K_{Y,n})]\Big) - [\mathrm{cov}(Y_n' K_{i,n} Y_n, Y_n' K_{j,n} Y_n)] \geq 0. \qquad \text{(A.0.7)}$$

One obtains

$$\begin{aligned}\mathrm{var}(\hat{\theta}_n(i)) &= \Big(A_n^{-1}(1/n^4)[\mathrm{cov}(Y_n' K_{i,n} Y_n, Y_n' K_{j,n} Y_n)] A_n^{-1}\Big)(i,i) \\ &= a'\ (1/n^4)[\mathrm{cov}(Y_n' K_{i,n} Y_n, Y_n' K_{j,n} Y_n)]\ a\end{aligned}$$

where a is the i-th column of A^{-1}. Hence by (A.0.7) one has

$$\mathrm{var}(\hat{\theta}_n(i)) \leq cq\ a'\ (1/n^4)\ \mathrm{diag}([\mathrm{tr}(K_{i,n} K_{Y,n} K_{i,n} K_{Y,n})])\ a$$

and finally

$$\mathrm{var}(\hat{\theta}_n(i)) \leq cq\ \Big(A^{-1}(1/n^4)\ \mathrm{diag}([\mathrm{tr}(K_{i,n} K_{Y,n} K_{i,n} K_{Y,n})]) A^{-1}\Big)(i,i) = cq\ E_n(i,i)$$

which concludes the proof.

Proof of Theorem 1.5.3:

Let $Q_n = X_n S X_n'$ and $P_n = X_n(X_n'X_n)^{-1}X_n'$. Then

$$
\begin{aligned}
&|Q_n(l_1,l_2) - nP_n(l_1,l_2)| \\
&\quad = |\sum_{k_1=1}^{p}\sum_{k_2=1}^{p} X_n(l_1,k_1)(S(k_1,k_2) - S_n(k_1,k_2))X_n(k_2,l_2)| \\
&\quad = |\sum_{k_1=1}^{p}\sum_{k_2=1}^{p} r_{k_1}(x_{l_1})(S(k_1,k_2) - S_n(k_1,k_2))r_{k_2}(x_{l_2})| \\
&\quad \leq \sum_{k_1=1}^{p}\sum_{k_2=1}^{p} |r_{k_1}(x_{l_1})r_{k_2}(x_{l_2})||S(k_1,k_2) - S_n(k_1,k_2)| \to 0 \quad \text{(A.0.8)}
\end{aligned}
$$

for $1 \leq k_1, k_2 \leq p$, $1 \leq l_1, l_2 \leq n$ as $n \to \infty$, since the functions r_k are continuous and therefore bounded on D. The convergence is uniform in x_i and x_j. Let

$$
Q(x_{l_1}, x_{l_2}) = Q_n(l_1,l_2) = \sum_{k_1=1}^{p}\sum_{k_2=1}^{p} r_{k_1}(x_{l_1})r_{k_2}(x_{l_2})S(k_1,k_2). \quad \text{(A.0.9)}
$$

Since the entries of S are constants, the function Q above can be defined for all $(x,y) \in D^2$. It is seen that $Q(x,y)$ is a continuous function on D^2 and it is independent of n. The variance of e is

$$
\begin{aligned}
U_{i,n}(l_1,l_2) = C_i(x_{l_1},x_{l_2}) &- \sum_{t_1=1}^{n} P_n(l_1,t_1)C_i(x_{t_1},x_{l_2}) - \sum_{t_2=1}^{n} P_n(l_2,t_2)C_i(x_{l_1},x_{t_2}) \\
&+ \sum_{t_1=1}^{n}\sum_{t_2=1}^{n} P_n(l_1,t_1)P_n(l_2,t_2)C_i(x_{t_1},x_{t_2}). \quad \text{(A.0.10)}
\end{aligned}
$$

Let

$$\phi_{i,n}(x_{l_1}, x_{l_2}) = C_i(x_{l_1}, x_{l_2}) - \frac{1}{n}\sum_{t_1=1}^{n} Q(x_{l_1}, x_{t_1}) C_i(x_{t_1}, x_{l_2})$$
$$- \frac{1}{n}\sum_{t_2=1}^{n} Q(x_{l_2}, x_{t_2}) C_i(x_{l_1}, x_{t_2})$$
$$+ \frac{1}{n^2}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n} Q(x_{l_1}, x_{t_1}) Q(x_{l_2}, x_{t_2}) C_i(x_{t_1}, x_{t_2}). \quad \text{(A.0.11)}$$

One obtains

$$|\phi_{i,n}(x_{l_1}, x_{l_2}) - U_{i,n}(l_1, l_2)|$$
$$\leq \frac{1}{n}\sum_{t_1=1}^{n} \Big| Q(x_{l_1}, x_{t_1}) - nP_n(l_1, t_1) \Big| \Big| C_i(x_{t_1}, x_{l_2}) \Big|$$
$$+ \frac{1}{n}\sum_{t_2=1}^{n} \Big| Q(x_{l_2}, x_{t_2}) - nP_n(l_2, t_2) \Big| \Big| C_i(x_{l_1}, x_{t_2}) \Big|$$
$$+ \frac{1}{n^2}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n} \Big| Q(x_{l_1}, x_{t_1}) Q(x_{l_2}, x_{t_2}) - n^2 P_n(l_1, t_1) P_n(l_2, t_2) \Big| \Big| C_i(x_{t_1}, x_{t_2}) \Big| \to 0$$

as $n \to \infty$ by (A.0.8) and by the boundedness of the function C_i and the convergence is uniform in (x_{l_1}, x_{l_2}). Also, by Lemmas 1.5.1 and 1.5.2

$$\phi_i(x_{l_1}, x_{l_2}) = \lim_{n\to\infty} \phi_{i,n}(x_{l_1}, x_{l_2})$$
$$= C_i(x_{l_1}, x_{l_2}) - \int_D Q(x_{l_1}, \xi) C_i(\xi, x_{l_2}) f(\xi) d\xi - \int_D Q(x_{l_2}, \xi) C_i(x_{l_1}, \xi) f(\xi) d\xi$$
$$+ \int_D \int_D Q(x_{l_1}, \xi) Q(x_{l_2}, \eta) C_i(\xi, \eta) f(\xi) f(\eta) d\xi d\eta \quad \text{(A.0.12)}$$

uniformly in $(x_{l_1}, x_{l_2}) \in D^2$, $1 \leq l_1, l_2 \leq n$. One defines ϕ_Y by replacing C_i by C_Y in the equation above.

Proof of Theorem 1.5.4:

Clearly one has $K_{W,n}(l_1, l_2) = W(x_{l_1}, x_{l_2}) = I_n(l_1, l_2)$. Let $U_{W,n} = P_n I_n P_n$. If in (A.0.11) C_i is replaced by W, the uniform limit in (A.0.12) is simply $W(x_{l_1}, x_{l_2})$. Hence and from the previous proof it follows that for $1 \leq j \leq q$

$$\lim_{n\to\infty} \frac{1}{n} \operatorname{tr}(U_{W,n} U_{j,n}) = \lim_{n\to\infty} \frac{1}{n} \operatorname{tr}(I_n U_{j,n}) = \lim_{n\to\infty} \sum_{l=1}^{n} \frac{1}{n} \phi_j(x_l, x_l)$$
$$= \int_D \phi_j(\xi, \xi) f(\xi) d\xi. \quad \text{(A.0.13)}$$

Moreover, if $U_{j,n} = U_{W,n}$ this last limit equals one. In a similar fashion

$$U_{(Y,\epsilon),n} = P_n K_{(Y,\epsilon),n} P_n = P_n(K_{Y,n} + \gamma I_n) P_n = U_{Y,n} + \gamma U_{W,n}.$$

One obtains

$$\lim_{n\to\infty} \frac{1}{n^2} \operatorname{tr}(U_{W,n} U_{(Y,\epsilon),n} U_{W,n} U_{(Y,\epsilon),n}) = \lim_{n\to\infty} \frac{1}{n^2} \operatorname{tr}(U_{Y,n} U_{Y,n})$$
$$= \lim_{n\to\infty} \sum_{k=1}^{n} \sum_{l=1}^{n} \frac{1}{n^2} \phi_Y(x_k, x_l) \phi_Y(x_l, x_k)$$
$$= \int_D \int_D \phi_Y(\xi, \eta) \phi_Y(\eta, \xi) f(\xi) f(\eta) d\xi d\eta = b_0 \quad \text{(A.0.14)}$$

and for $1 \leq j \leq q$, the following holds

$$\lim_{n\to\infty} \frac{1}{n^3} \operatorname{tr}(U_{W,n} U_{(Y,\epsilon),n} U_{j,n} U_{(Y,\epsilon),n})$$
$$= \lim_{n\to\infty} \frac{1}{n^3} \operatorname{tr}(I_n U_{Y,n} U_{j,n} U_{Y,n})$$
$$= \lim_{n\to\infty} \sum_{k=1}^{n} \sum_{l=1}^{n} \sum_{m=1}^{n} \frac{1}{n^3} \phi_Y(x_k, x_l) \phi_j(x_l, x_m) \phi_Y(x_m, x_k)$$
$$= \int_D \int_D h_{1,j}(\xi, \eta) \phi_Y(\eta, \xi) f(\xi) f(\eta) d\xi d\eta = b(j) \quad \text{(A.0.15)}$$

where

$$h_{1,j}(\xi,\eta) = \int_D \phi_Y(\xi,\lambda)\phi_j(\lambda,\eta)f(\lambda)d\lambda.$$

Finally, one has

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n^4}\operatorname{tr}(U_{i,n}U_{(Y,\epsilon),n}U_{j,n}U_{(Y,\epsilon),n}) \\ = \lim_{n\to\infty} \frac{1}{n^4}\operatorname{tr}(U_{i,n}U_{Y,n}U_{j,n}U_{Y,n}) \\ = \int_D\int_D h_{1,j}(\xi,\eta)h_{1,i}(\eta,\xi)f(\xi)f(\eta)d\xi d\eta = B_1(i,j). \quad \text{(A.0.16)}\end{aligned}$$

From the above remarks it follows that

$$\begin{bmatrix} \frac{1}{n} & a' \\ 0 & \frac{1}{n^2}I_q \end{bmatrix} \begin{bmatrix} n-p & [\operatorname{tr}(U_{i,n})]' \\ [\operatorname{tr}(U_{i,n})] & [\operatorname{tr}(U_{i,n}U_{j,n})] \end{bmatrix} \to \begin{bmatrix} 1 & 0 \\ 0 & A_1 \end{bmatrix} = A_{ne}. \quad \text{(A.0.17)}$$

As before, it will be assumed that A_1 is invertible. Similarly,

$$\begin{bmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{n^2}I_q \end{bmatrix} \begin{bmatrix} \operatorname{tr}(U_{W,n}U_{(Y,\epsilon),n}U_{W,n}U_{(Y,\epsilon),n}) & [\operatorname{tr}(U_{i,n}U_{(Y,\epsilon),n}U_{W,n}U_{(Y,\epsilon),n})]' \\ [\operatorname{tr}(U_{i,n}U_{(Y,\epsilon),n}U_{W,n}U_{(Y,\epsilon),n})] & [\operatorname{tr}(U_{i,n}U_{(Y,\epsilon),n}U_{j,n}U_{(Y,\epsilon),n})] \end{bmatrix} \begin{bmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{n^2}I_q \end{bmatrix}$$

$$\to \begin{bmatrix} b_0 & b' \\ b & B_1 \end{bmatrix} = B_{ne}. \quad \text{(A.0.18)}$$

Moreover,

$$\begin{bmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{n^2}I_q \end{bmatrix} \begin{bmatrix} \operatorname{tr}(U_{W,n}U_{(Y,\epsilon),n} \\ [\operatorname{tr}(U_{i,n}U_{(Y,\epsilon),n}] \end{bmatrix} \to \begin{bmatrix} m_0 \\ M_1 \end{bmatrix} = M_{ne} \quad \text{(A.0.19)}$$

where the $q \times q$ matrix M and m_0 are defined as follows

$$M_1(i) = \int_D\int_D \phi_Y(\xi,\eta)\phi_i(\eta,\xi)f(\xi)f(\eta)d\xi d\eta$$

$$m_0 = \int_D C_{Y,\epsilon}(\xi,\xi)f(\xi)d\xi.$$

It follows that

$$\lim_{n\to\infty} \mathbf{E}[\hat{\theta}] = \lim_{n\to\infty} \begin{bmatrix} n-p & [\mathrm{tr}(U_{i,n})]' \\ [\mathrm{tr}(U_{i,n})] & [\mathrm{tr}(U_{i,n}U_{j,n})] \end{bmatrix}^{-1} \begin{bmatrix} \mathrm{tr}(U_{W,n}U_{(Y,\epsilon),n} \\ [\mathrm{tr}(U_{i,n}U_{(Y,\epsilon),n}] \end{bmatrix}$$

$$= \lim_{n\to\infty} \begin{bmatrix} n-p & [\mathrm{tr}(U_{i,n})]' \\ [\mathrm{tr}(U_{i,n})] & [\mathrm{tr}(U_{i,n}U_{j,n})] \end{bmatrix}^{-1} \begin{bmatrix} n & 0 \\ 0 & n^2 I_q \end{bmatrix} \begin{bmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{n^2} I_q \end{bmatrix} \begin{bmatrix} \mathrm{tr}(U_{W,n}U_{(Y,\epsilon),n} \\ [\mathrm{tr}(U_{i,n}U_{(Y,\epsilon),n}] \end{bmatrix}$$

$$= \begin{bmatrix} 1 & a' \\ 0 & A_1^{-1} \end{bmatrix} \begin{bmatrix} m_0 \\ M_1 \end{bmatrix} = A_{ne}^{-1} M_{ne}. \quad \text{(A.0.20)}$$

Furthermore,

$$\lim_{n\to\infty} \begin{bmatrix} n-p & [\mathrm{tr}(U_{i,n})]' \\ [\mathrm{tr}(U_{i,n})] & [\mathrm{tr}(U_{i,n}U_{j,n})] \end{bmatrix}^{-1}$$

$$\begin{bmatrix} \mathrm{tr}(U_{W,n}U_{(Y,\epsilon),n}U_{W,n}U_{(Y,\epsilon),n}) & [\mathrm{tr}(U_{i,n}U_{(Y,\epsilon),n}U_{W,n}U_{(Y,\epsilon),n})]' \\ [\mathrm{tr}(U_{i,n}U_{(Y,\epsilon),n}U_{W,n}U_{(Y,\epsilon),n})] & [\mathrm{tr}(U_{i,n}U_{(Y,\epsilon),n}U_{j,n}U_{(Y,\epsilon),n})] \end{bmatrix}$$

$$\begin{bmatrix} n-p & [\mathrm{tr}(U_{i,n})]' \\ [\mathrm{tr}(U_{i,n})] & [\mathrm{tr}(U_{i,n}U_{j,n})] \end{bmatrix}^{-1}$$

$$= \lim_{n\to\infty} \begin{bmatrix} n-p & [\mathrm{tr}(U_{i,n})]' \\ [\mathrm{tr}(U_{i,n})] & [\mathrm{tr}(U_{i,n}U_{j,n})] \end{bmatrix}^{-1} \begin{bmatrix} n & 0 \\ 0 & n^2 I_q \end{bmatrix} \begin{bmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{n^2} I_q \end{bmatrix}$$

$$\begin{bmatrix} \mathrm{tr}(U_{W,n}U_{(Y,\epsilon),n}U_{W,n}U_{(Y,\epsilon),n}) & [\mathrm{tr}(U_{i,n}U_{(Y,\epsilon),n}U_{W,n}U_{(Y,\epsilon),n})]' \\ [\mathrm{tr}(U_{i,n}U_{(Y,\epsilon),n}U_{W,n}U_{(Y,\epsilon),n})] & [\mathrm{tr}(U_{i,n}U_{(Y,\epsilon),n}U_{j,n}U_{(Y,\epsilon),n})] \end{bmatrix}$$

$$\begin{bmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{n^2} I_q \end{bmatrix} \begin{bmatrix} n & 0 \\ 0 & n^2 I_q \end{bmatrix} \begin{bmatrix} n-p & [\mathrm{tr}(U_{i,n})]' \\ [\mathrm{tr}(U_{i,n})] & [\mathrm{tr}(U_{i,n}U_{j,n})] \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 1 & a' \\ 0 & A_1^{-1} \end{bmatrix} \begin{bmatrix} b_0 & b' \\ b & B_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ a & A_1^{-1} \end{bmatrix} = A_{ne}^{-1} B_{ne} A_{ne}^{-1}.$$

Proof of Lemma 1.5.6:

Putting

$$U_n = \frac{2}{n(n-1)} \sum_{k=1}^{n} \sum_{l=1}^{k-1} \phi_1(X_k, X_l)\phi_2(X_l, X_k)$$

one has $Z_n = U_n + \frac{2}{n(n-1)} \sum_{k=1}^{n} \phi_1(X_k, X_k)\phi_2(X_k, X_k) = U_n + V_n$ and since the functions ϕ_i are bounded the V_n is $O(n^{-1})$. It is easily seen that U_n is a U-statistic (Serfling 1980) whose kernel is $h(X_1, X_2) = \phi_1(X_1, X_2)\phi_2(X_1, X_2)$. The it follows directly from Theorem A of Section 5.5.1 in Serfling (1980) that

$$n^{1/2}\left(\frac{U_n - \kappa(D)}{2\varsigma_1^{1/2}}\right) \rightarrow_d \mathbf{N}(0,1).$$

The corresponding result for Z_n follows from Slutzky's theorem.

Proof of Theorem 1.5.5:

Let the vector random variable $E_n(X)$ consist of all entries of $A_n(X)$ lying on or above the diagonal, followed by the entries of $M_n(X)$, i.e.

$$E_n(X) = \Big(A_n(X)(1,1), A_n(X)(1,2), \ldots, A_n(X)(1,q),$$
$$A_n(X)(2,2), A_n(X)(2,3), \ldots, A_n(X)(2,q),$$
$$\ldots, A_n(X)(q,q), M_n(X)(1), \ldots, M_n(X)(q)\Big)'. \quad \text{(A.0.21)}$$

From the discussion following Theorem A of Section 5.5.1 in Serfling (1980) and Lemma 1.5.6 it follows that the vector $n^{1/2}(E_n(X) - E)$ is asymptotically multinormal with zero mean where E is defined by

$$E = \Big(A(1,1), A(1,2), \ldots, A(1,q), A(2,2), A(2,3), \ldots, A(2,q), \ldots, A(q,q),$$
$$M(1), \ldots, M(q)\Big)'. \quad \text{(A.0.22)}$$

For a non-singular square, $q \times q$ matrix B, it follows from the Cayley-Hamilton theorem that

$$B^{-1} = -(B^{q-1} + c_1 B^{q-2} + \ldots + c_{q-1})/c_q \tag{A.0.23}$$

where the c_i are the coefficients of the characteristic polynomial

$$\det(\lambda I - B) = \lambda^q + \ldots + c_q.$$

(By non-singularity of B, $c_q = \det(B) \neq 0$.) The coefficients c_i are polynomials in the entries of the matrix B. It follows that the $i - th$ entry of the vector $(A_n(X))^{-1} M_n(X)$ can be expressed as $H_i(E_n(X)) = P_i(E_n(X))/P(E_n(X))$, $1 \leq i \leq q$, where P_i and P are polynomials and $P(E_n(X)) = \det(A_n(X))$. Since A_n is positive definite, $P(E_n) = \det(A_n) \neq 0$, and H is differentiable at E_n. Hence the result follows.

Proof of Lemma 1.5.7: One seeks to show that $\mu(A) = 0$ implies $\mathbf{G}(A) = 0$ for any Lebesgue-measurable set A, where μ is the Lebesgue measure. It suffices to consider $A = [0, a]$ for some positive real number a. One observes that

$$\mathbf{G}(A) = \int_A d\, h(a)$$

where

$$h(a) = \int_{aB} \int_D f(x) f(x + \xi) I_D(x + \xi) dx\, d\xi$$

and where aB is the unit ball scaled by the factor a. Hence

$$\mathbf{G}([b, a]) = \int_{aB \setminus bB} \int_D f(x) f(x + \xi) I_D(x + \xi) dx\, d\xi$$

Since the function $f(x)f(x + \xi)I_D(x + \xi)$ is bounded on D, the result follows.

Proof of Lemma 1.5.8:

For each m, let $D_m = T_m(D)$. It is easily seen that

$$T_m(B_{r_m^{-1}\rho}(D)) = B_\rho(T_m(D))$$

hence $\mu(D_m \setminus B_\rho(D_m)) \leq c\rho r_m^{d-1}$. Let $\mathbf{F}_{1,m}(A) = \int_A f_m(\xi)d\xi$ for any measurable subset A of D. Let $s(d) = \pi^{d/2}/\Gamma(d/2+1)$. Therefore $\mu(\{x \in \mathbf{R}^d : \| x \| \leq \rho\}) = s(d)\rho^d$ and let $D_m = T_m(D)$. By the hypotheses of Lemma 1.5.8 $\mathbf{F}_{1,m}(A) = \mu(A)/\mu(D_m)$ one observes

$$\begin{aligned}
\mathbf{F}_{1,m}(D_m)\,\mathbf{F}_{1,m}(\{x \in \mathbf{R}^d : \| x \| \leq \rho\}) &\geq G_m(\rho) \\
&= \mathbf{F}_{2,m}(\{(\xi,\eta) \in D_m^2 : \| \xi - \eta \| \leq \rho\}) \\
&\geq \mathbf{F}_{1,m}(B_\rho(D_m))\,\mathbf{F}_{1,m}(\{x \in \mathbf{R}^d : \| x \| \leq \rho\}) \\
&= \mu(D_m)^{-2}\,\mu(B_\rho(D_m))\,\mu(\{x \in \mathbf{R}^d : \| x \| \leq \rho\}) \\
&\geq \mu(D_m)^{-2}\,(\mu(D_m) - c\rho r_m^{d-1})s(d)\rho^d \\
= \mu(D)^{-2}\,r_m^{-2d}(\mu(D)\,r_m^d - c\rho r_m^{d-1})s(d)\rho^d &= s(d)\,\mu(D)\,r_m^{-d}\rho^d - s(d)cr_m^{-d-1}\rho^{d+1}
\end{aligned}$$

while

$$\mathbf{F}_{1,m}(D_m)\,\mathbf{F}_{1,m}(\{x \in \mathbf{R}^d : \| x \| \leq \rho\}) = s(d)\,\mu(D)\,r_m^{-d}\rho^d.$$

Hence

$$s(d)\,\mu(D)\,r_m^{-d}\rho^d \geq G_m(\rho) \geq s(d)\,\mu(D)\,r_m^{-d}\rho^d - s(d)cr_m^{-d-1}\rho^{d+1}$$

and putting $\alpha = c\pi^{d/2}/\Gamma(d/2+1)$ one obtains the required result.

Proof of Lemma 1.5.9:

By Lemma 1.5.8

$$\begin{aligned}
r_m^d\left|\int_0^{\mathrm{diam}(D_m)} \phi(\rho)dG_m(\rho) - \alpha_G \int_0^{\mathrm{diam}(D_m)} \phi(\rho)\rho^{d-1}d\rho\right| & \\
\leq r_m^{-1}\alpha_2 \int_0^{\mathrm{diam}(D_m)} |\phi(\rho)|\rho^d d\rho \to 0 &
\end{aligned}$$

by the hypothesis, whence the lemma follows.

Two other elementary results will be useful in the further development:

Lemma A.0.1. *For any real nonnegative function* $\phi \in L^1(\mathbf{R}_+)$

$$\lim_{r_m \to \infty} \frac{1}{r_m} \int_0^{r_m} \phi(\rho) \rho d\rho = 0.$$

Lemma A.0.2. *If* ϕ *is a real function satisfying*

$$\int_0^{\infty} |\phi(\rho)|^2 \rho^{d-1} \rho < \infty$$

for an integer $d \geq 2$, *then*

$$\lim_{r_m \to \infty} r_m^{-d/2} \int_0^{r_m} |\phi(\rho)| \rho^{d-1} d\rho. < \infty$$

Proof of Lemma A.0.1:

Let $\phi_m(\rho) = \frac{1}{r_m} I_{[0,r_m]}(\rho) \phi(\rho) \rho$. One observes that

$$\frac{1}{r_m} \int_0^{r_m} \phi(\rho) \rho d\rho = \int_0^{\infty} \phi_m(\rho) d\rho$$

and

$$\lim_{r_m \to \infty} \phi_m(\rho) = 0 \, \forall \rho \geq 0,$$

moreover, $\phi_m \leq \phi$. Since ϕ is summable, by the dominated convergence theorem

$$\lim_{r_m \to \infty} \int_0^{\infty} \phi_m(\rho) d\rho = \int_0^{\infty} (\lim_{r_m \to \infty} \phi_m(\rho)) d\rho = 0$$

Proof of Lemma A.0.2:

Without loss of generality, ϕ may be assumed nonnegative. The first step in the proof is the case $d = 2$. Thus it is assumed that

$$\int_0^{\infty} (\phi(\rho))^2 \rho d\rho < \infty.$$

One defines the function $b(\rho) = \phi(\rho)\rho$ and the following measurable sets

$$S_+ = \{\rho : b(\rho) > 1\}$$

$$S_- = \{\rho : b(\rho) \leq 1\}.$$

It is seen that

$$\int_0^{r_m} \phi(\rho)\rho d\rho = \int_{S_+\cap[0,r_m]} \phi(\rho)\rho d\rho + \int_{S_-\cap[0,r_m]} \phi(\rho)\rho d\rho.$$

The second term will be dealt with first:

$$\lim_{r_m\to\infty} r_m^{-1} \int_{S_-\cap[0,r_m]} \phi(\rho)\rho d\rho \le \lim_{r_m\to\infty} r_m^{-1} \int_{S_-\cap[0,r_m]} d\rho$$
$$\le \lim_{r_m\to\infty} r_m^{-1} \int_0^{r_m} d\rho = 1.$$

To find a bound on the first term, one observes that from the hypothesis

$$\int_0^{\infty} (\phi(\rho))^2 \rho d\rho = \int_0^{\infty} (b(\rho))^2 \rho^{-1} d\rho < \infty,$$

hence

$$\int_{S_+} b(\rho)\rho^{-1} d\rho \le \int_{S_+} (b(\rho))^2 \rho^{-1} d\rho < \infty$$

which means that $I_{S_+}(\rho)b(\rho)\rho^{-1} \in L_1(\mathbf{R}_+)$, Lemma A.0.1 applies and thus

$$\lim_{r_m\to\infty} r_m^{-1} \int_0^{r_m} I_{S_+}(\rho)b(\rho)\rho^{-1}\rho d\rho = 0$$

thus

$$\lim_{r_m\to\infty} r_m^{-1} \int_{S_+\cap[0,r_m]} \phi(\rho)\rho d\rho = 0$$

which proves the case $d = 2$.

To show the result for an integer $d > 2$, one puts $\psi(\rho) = \phi(\rho)\rho^{(d-2)/2}$. It is seen that ψ satisfies

$$\int_0^{\infty} (\psi(\rho))^2 \rho d\rho < \infty$$

and thus the case shown above ($d = 2$) applies. Thus

$$\lim_{r_m\to\infty} r_m^{-1} \int_0^{r_m} \phi(\rho)\rho^{d/2} d\rho = \lim_{r_m\to\infty} r_m^{-1} \int_0^{r_m} \psi(\rho)\rho d\rho < \infty.$$

However,

$$r_m^{-1}\int_0^{r_m}\phi(\rho)\rho^{d/2}d\rho = r_m^{-1}\int_0^{r_m}\phi(\rho)\rho^{d/2}\rho^{d/2-1}\rho^{1-d/2}d\rho$$
$$\leq r_m^{-1}\int_0^{r_m}\phi(\rho)\rho^{d/2}\rho^{d/2-1}r_m^{1-d/2}d\rho = r_m^{-d/2}\int_0^{r_m}\phi(\rho)\rho^{d-1}d\rho$$

and Lemma A.0.2 is proved.

Proof of Theorem 1.5.6:

Initially it will be assumed that $v \equiv 1$ and that the covariance components $\mathbf{C}_i$ satisfy Condition 1.5.3 with $v \equiv 1$, which is equivalent to Condition 1.5.4. The matrix A_m can be expressed as

$$A_m(i,j) = \int_0^{\text{diam}(D_m)} \mathbf{C}_i(\rho)\mathbf{C}_j(\rho)dG_m(\rho)$$
$$= \alpha_G r_m^{-d}\int_0^{\text{diam}(D_m)} \mathbf{C}_i(\rho)\mathbf{C}_j(\rho)\rho^{d-1}d\rho - \int_0^{\text{diam}(D_m)} \mathbf{C}_i(\rho)\mathbf{C}_j(\rho)dR_m(\rho) \quad \text{(A.0.24)}$$

where

$$\left|\int_0^{\text{diam}(D_m)} \mathbf{C}_i(\rho)\mathbf{C}_j(\rho)dR_m(\rho)\right| \leq \int_0^{\text{diam}(D_m)} |\mathbf{C}_i(\rho)\mathbf{C}_j(\rho)|dR_m(\rho)$$
$$\leq \alpha_2(d+1)r_m^{-d-1}\int_0^{\text{diam}(D_m)} |\mathbf{C}_i(\rho)\mathbf{C}_j(\rho)|\rho^d d\rho.$$

Since Cauchy-Schwarz and the hypothesis of the result imply

$$\int_0^{\infty} |\mathbf{C}_i(\rho)\mathbf{C}_j(\rho)|\rho^{d-1}d\rho < \infty \quad \text{(A.0.25)}$$

then by Lemma A.0.1

$$r_m^{-1}\int_0^{r_m\,\text{diam}(D)} |\mathbf{C}_i(\rho)\mathbf{C}_j(\rho)|\rho^d d\rho \to 0 \quad \text{(A.0.26)}$$

as $r_m \to \infty$, the second term in (A.0.24) becomes insignificant, and $A_m(i,j)$ is proportional to r_m^{-d}.

If $\nu(\rho)$ is not constant and Condition 1.5.3 holds, but Condition 1.5.4 does not, it is sufficient to redefine $\mathbf{C}_i$ to be $\mathbf{C}_i(\rho)\nu(\rho)^{1/2}$ and the proof remains unchanged. While the functions $\mathbf{C}_i(\rho)\nu(\rho)^{1/2}$ may not be valid covariance functions, the only important elements in the proof are convergence arguments, and those are unaffected by this redefinition.

Next the entries of the matrix B_m are considered. Again, initially it will be assumed that $v \equiv 1$ and that the covariance components $\mathbf{C}_i$ satisfy Condition 1.5.3 with $v \equiv 1$, which is equivalent to Condition 1.5.4. One has

$$
\begin{aligned}
&|B_m(i,j)| \\
&\qquad = \left| \int_{D_m} \int_{D_m} \left(\int_{D_m} C_i(\xi,\lambda) C_Y(\lambda,\eta) f_m(\lambda) d\lambda \right) \right. \\
&\qquad \left. \left(\int_{D_m} C_j(\xi,\lambda) C_Y(\lambda,\eta) f_m(\lambda) d\lambda \right) f(\xi) f(\eta) d\xi d\eta \right| \\
&\qquad \leq \int_{D_m} \int_{D_m} \left(\int_{D_m} |C_i(\xi,\lambda) C_Y(\lambda,\eta)| f_m(\lambda) d\lambda \right) \\
&\qquad \left(\int_{D_m} |C_j(\xi,\lambda) C_Y(\lambda,\eta)| f_m(\lambda) d\lambda \right) f_m(\xi) f_m(\eta) d\xi d\eta. \quad \text{(A.0.27)}
\end{aligned}
$$

Supposing that $f_m \equiv \mu(D_m)^{-1}$ one considers the integral

$$
\int_{\mathbf{R}^d} |C_i(\xi,\lambda) C_Y(\lambda,\eta)| d\lambda. \quad \text{(A.0.28)}
$$

Let $\psi_i(x) = C_i(x,0)$ and $\psi_Y(x) = C_Y(x,0)$. Then by the stationarity of the process

$$
\begin{aligned}
\int_{\mathbf{R}^d} |C_i(\xi,\lambda) C_Y(\lambda,\eta)| d\lambda &= \int_{\mathbf{R}^d} |C_i(\xi-\eta-\lambda,0) C_Y(\lambda,0)| d\lambda \\
&= \int_{\mathbf{R}^d} |\psi_i(\xi-\eta-\lambda) \psi_Y(\lambda)| d\lambda = (|\psi_i| * |\psi_Y|)(\xi-\eta) = \vartheta_i(\xi-\eta)
\end{aligned}
$$

where $\psi_i * \psi_Y$ denotes the convolution of the two functions. If $\psi_i \in L^2$ and $\psi_Y \in L^1$ (and by its boundedness, also $\psi_Y \in L^2$) then the same applies to $|\psi_i|$

and $|\psi_Y|$. Firstly, the convolution integral converges:

$$|\psi_i| * |\psi_Y| = \int_{\mathbf{R}^d} |\psi_i(\xi - \eta - \lambda)|\,|\psi_Y(\lambda)| d\lambda$$
$$\leq \frac{1}{2} \int_{\mathbf{R}^d} \Big(|\psi_i(\xi - \eta - \lambda)|^2 + |\psi_Y(\lambda)|^2 \Big) d\lambda < \infty.$$

Let $\mathcal{F}(\psi)$ be the Fourier transform of ψ. It follows that $\mathcal{F}(|\psi_i|)$ and $\mathcal{F}(|\psi_Y|)$ exist and are all in L^2. Moreover, since $|\psi_Y| \in L^1$, it follows that

$$|\mathcal{F}(|\psi_Y|)| < M < \infty. \tag{A.0.29}$$

Furthermore

$$|\mathcal{F}(|\psi_i| * |\psi_Y|)| = |\mathcal{F}(|\psi_i|)\mathcal{F}(|\psi_Y|)| < M|\mathcal{F}(|\psi_i|)| \in L^2.$$

Since for L^2 functions the Fourier transform is fully reciprocal, it follows that

$$(|\psi_i| * |\psi_Y|)(\xi) = \vartheta_i(\xi) \in L^2. \tag{A.0.30}$$

It is also clear that the ϑ_i are isotropic since ψ_i and ψ_Y are. Let κ_i be a function defined on the positive real line such that

$$\vartheta_i(\xi) = \kappa_i(\| \xi \|). \tag{A.0.31}$$

Thus

$$\int_{\mathbf{R}^d} |\vartheta_i(\xi)\vartheta_j(\xi)| d\xi < \infty$$

and so

$$\int_0^\infty |\kappa_i(\rho)\kappa_j(\rho)| \rho^{d-1} d\rho < \infty \tag{A.0.32}$$

and hence the expression (A.0.27) takes the form

$$
\begin{aligned}
|B_m(i,j)| \leq \mu(D_m)^{-2} \int_{D_m}\int_{D_m}\left(\int_{\mathbf{R}^d}|C_i(\xi,\lambda)C_Y(\lambda,\eta)|d\lambda\right) \\
\left(\int_{\mathbf{R}^d}|C_j(\xi,\lambda)C_Y(\lambda,\eta)|d\lambda\right)f(\xi)f(\eta)d\xi d\eta \\
= r_m^{-2d}\int_0^{\mathrm{diam}(D_m)}\kappa_i(\rho)\kappa_j(\rho)dG_m(\rho) \\
\leq \alpha_G r_m^{-3d}\int_0^{\mathrm{diam}(D_m)}\kappa_i(\rho)\kappa_j(\rho)\rho^{d-1}d\rho - r_m^{-2d}\int_0^{\mathrm{diam}(D_m)}\kappa_i(\rho)\kappa_j(\rho)dR_m(\rho)
\end{aligned}
\tag{A.0.33}
$$

where the last term can be bounded by

$$
\begin{aligned}
\left|dr_m^{-2d}\int_0^{\mathrm{diam}(D_m)}\kappa_i(\rho)\kappa_j(\rho)dR_m(\rho)\right| \\
\leq \alpha_2(d+1)r_m^{-3d-1}\int_0^{\mathrm{diam}(D_m)}|\kappa_i(\rho)\kappa_j(\rho)|\rho^d d\rho.
\end{aligned}
$$

From (A.0.32) it follows that the first term of the expression is bounded by a multiple of r_m^{-3d}. Furthermore, the term

$$
r_m^{-1}\int_0^{\mathrm{diam}(D_m)}|\kappa_i(\rho)\kappa_j(\rho)|\rho^d
$$

converges to zero by Lemma A.0.1 and the last term in (A.0.33) is negligible.

If $\nu(\rho)$ is not constant and Condition 1.5.3 holds, but Condition 1.5.4 does not, it is sufficient to redefine $\mathbf{C}_i$ to be $\mathbf{C}_i(\rho)\nu(\rho)$ and to redefine $\nu(\rho) \equiv 1$ and the argument above remains unchanged. While the functions $\mathbf{C}_i(\rho)\nu(\rho)$ may not be valid covariance functions, the only important elements in the proof are convergence arguments, and those are unaffected by this redefinition.

If A is invertible, then

$$\lim_{r_m\to\infty} A_m^{-1}B_mA_m^{-1} = \lim_{r_m\to\infty} r_m^{-d}(r_m^dA_m)^{-1}(r_m^{3d}B_m)(r_m^dA_m)^{-1}$$
$$= \lim_{r_m\to\infty} r_m^{-d}A^{-1}BA^{-1} = 0$$

where B is such that

$$\lim_{r_m\to\infty} r_m^{3d}B_m(i,j) \leq B(i,j)$$

(the existence follows from earlier discussion). The expressions for the entries of A follow directly from earlier arguments, while the expressions for the entries of B follow from the dominated convergence theorem applied to the sequence of functions $\{h_{i,m}h_{j,m}\}_{m=1}^{\infty}$ where

$$h_{i,m}(\xi,\eta) = \int_{D_m} C_i(\xi,\lambda)C_Y(\lambda,\eta)d\lambda$$

and similarly for $h_{j,m}$. The functions $\{h_{i,m}h_{j,m}\}_{m=1}^{\infty}$ are dominated by a summable function $\vartheta_i\vartheta_j$ as is shown in (A.0.33).

This concludes the proof of Theorem 1.5.6.

Proof of Theorem 1.5.7:

To prove the result, it would be sufficient if

$$\lim_{r_m\to\infty} r_m^d|A_{1,m} - A_m| = 0 \tag{A.0.34}$$

and for (1.5.93)

$$\lim_{r_m\to\infty} r_m^{3d}|B_{1,m} - B_m| = 0 \tag{A.0.35}$$

while for (1.5.95)

$$\lim_{r_m\to\infty} r_m^{2d+z+t}|B_{1,m} - B_m| = 0. \tag{A.0.36}$$

Clearly, if (A.0.36) holds and $z + t > d$ then (A.0.35) must hold. Let $\phi_{i,m}$ denote the function ϕ_i of (A.0.12) for the domain D_m. (One should not confuse

the current notation $\phi_{i,m}$ with the $\phi_{i,n}$ in the sections on convergence in a fixed compact domain, where n is the number of observations for the fixed domain D. In the current notation $\phi_{i,m}$ depends on the domain, and in the notation of the previous sections is the limit of the $\phi_{i,n}$ for the domain D_m as $n \to \infty$.) Let $\phi_{Y,m}$ be defined in a similar fashion with ϕ_Y replacing ϕ_i. The following holds

Lemma A.0.3. *If there exists a positive number α such that the condition*

$$|\mathbf{C}_i(\rho)| \leq \alpha\rho^{-z} \tag{A.0.37}$$

holds with $z > d/2$ then (A.0.34) holds and (A.0.35) holds. If, furthermore

$$\int_0^\infty |\mathbf{C}_i(\rho))|\nu(\rho)\rho^{d-1}d\rho < \infty \tag{A.0.38}$$

then (A.0.34) holds and (A.0.36) holds.

Before proving the lemma, two other lemmas will be stated and proved:

Lemma A.0.4. *If the function Q defined in (A.0.9) is bounded on $\mathbf{R}^{2d}$ and the condition in (A.0.37) holds with $z > 0$, the bound*

$$|\phi_{i,m}(x_{l_1}, x_{l_2}) - C_i(x_{l_1}, x_{l_2})| \leq kr_m^{-z} \tag{A.0.39}$$

holds uniformly on D_m^2 for $1 \leq i \leq q$ for some constant k. If the condition in (A.0.38) also holds, the bound (A.0.39) holds with $-z$ replaced by $-d$. The same is true if $\phi_{i,m}$ and C_i are replaced by $\phi_{Y,m}$ and C_Y, respectively.

Proof of Lemma A.0.4.

It is an easy observation that

$$\int_{D_m} |C_i(\xi, \lambda)|d\lambda \leq \int_0^{\text{diam}(D_m)} |\mathbf{C}_i(\rho)|\rho^{d-1}d\rho. \tag{A.0.40}$$

Assuming $|Q| \leq k_1 < \infty$ from (1.5.31) one has

$$\begin{aligned}|\phi_{i,m}(x_{l_1}, x_{l_2}) - C_i(x_{l_1}, x_{l_2})| = \Big| &\int_{D_m} Q(x_{l_1}, \xi) C_i(\xi, x_{l_2}) f_m(\xi) d\xi \\ &+ \int_{D_m} Q(x_{l_2}, \xi) C_i(x_{l_1}, \xi) f_m(\xi) d\xi \\ &- \int_{D_m} \int_{D_m} Q(x_{l_1}, \xi) Q(x_{l_2}, \eta) C_i(\xi, \eta) f_m(\xi) f_m(\eta) d\xi d\eta \Big| \\ \leq k_1 \Big(&\int_{D_m} |C_i(\xi, x_{l_2})| f_m(\xi) d\xi + \int_{D_m} |C_i(x_{l_1}, \xi)| f_m(\xi) d\xi \Big) \\ &+ k_1^2 \int_{D_m} \int_{D_m} |C_i(\xi, \eta)| f_m(\xi) f_m(\eta) d\xi d\eta \\ &\leq (2k_1 \alpha_G r_m^{-d} + k_1^2 \alpha_G r_m^{-d}) \int_0^{\operatorname{diam}(D_m)} |\mathbf{C}_i(\rho)| \rho^{d-1} d\rho.\end{aligned}$$

Thus if (A.0.38) holds, the lemma holds. (In particular, this is the case if $\phi_{i,m}$ and C_i are replaced by $\phi_{Y,m}$ and C_Y, respectively, by the hypothesis of the main Theorem 1.5.7.)

Otherwise if (A.0.37) holds, one has

$$\begin{aligned}\int_1^{\operatorname{diam}(D_m)} |\mathbf{C}(\rho)| \rho^{d-1} d\rho &\leq \alpha \int_1^{r_m \operatorname{diam}(D)} \rho^{-z} \rho^{d-1} d\rho \\ &\leq \frac{\alpha}{d-z} \Big((r_m \operatorname{diam}(D))^{d-z} - 1 \Big).\end{aligned}$$

Hence

$$\int_0^{\operatorname{diam}(D_m)} |\mathbf{C}(\rho)| \rho^{d-1} d\rho \leq k_z r_m^{d-z}$$

for some constant k_z, and

$$|\phi_{i,m}(x_{l_1}, x_{l_2}) - C_i(x_{l_1}, x_{l_2})| \leq k r_m^{-z}$$

for a certain constant k and thus Lemma A.0.4 follows.

For the hypotheses of (A.0.39) to hold, it suffices to show that the function Q of (A.0.12) is bounded on $\mathbf{R}^{2d}$ uniformly in m. The following lemma guarantees the validity of the hypothesis if the regressors r_k making up the matrix X are homogeneous functions. The function $r_k(\xi)$ is homogeneous of degree $\gamma_k \geq 0$ if $r_k(r\xi) = r^{\gamma_k} r_k(\xi)$ for all real r, and all $\xi \in D$. If $r_k(\xi)$ is a monomial involving various powers of the components of ξ, it is homogeneous and its order equals the sum of those powers.

Lemma A.0.5. *If the regressors r_k are homogeneous functions, then there exists a finite bound k_Q such that*

$$\forall_m \forall_{(\xi,\eta)\in D_m^2} Q_m(\xi,\eta) \leq k_Q$$

where the function Q_m is defined as in (A.0.9) for the domain D_m.

Proof:

For the domain D_m one has

$$Q_m(x_{l_1}, x_{l_2}) = \sum_{k_1=1}^{p} \sum_{k_2=1}^{p} r_{k_1}(x_{l_1}) r_{k_2}(x_{l_2}) S_m(k_1, k_2)$$

where

$$S_m = R_m^{-1}$$

and

$$R_m(k_1,k_2) = \int_{D_m} r_{k_1}(\xi) r_{k_2}(\xi) f_m(\xi) d\xi = \int_D r_{k_1}(r_m\xi) r_{k_2}(r_m\xi) f(\xi) d\xi$$
$$= r_m^{\gamma_{k_1}+\gamma_{k_2}} \int_D r_{k_1}(\xi) r_{k_2}(\xi) f(\xi) d\xi.$$

Thus

$$S_m = \begin{bmatrix} r_m^{-\gamma_1} & & 0 \\ & \ddots & \\ 0 & & r_m^{-\gamma_p} \end{bmatrix} S \begin{bmatrix} r_m^{-\gamma_1} & & 0 \\ & \ddots & \\ 0 & & r_m^{-\gamma_p} \end{bmatrix}. \tag{A.0.41}$$

Hence

$$\begin{aligned} \sup_{\xi,\eta\in D_m} |Q_m(\xi,\eta)| &= \sup_{\xi,\eta\in D} |Q_m(r_m\xi, r_m\eta)| \\ &= \sup_{\xi,\eta\in D} |r_m^{\gamma_{k_1}} r_{k_1}(\xi) r_m^{\gamma_{k_2}} r_{k_2}(\eta) r_m^{-(\gamma_{k_1}+\gamma_{k_2})} S(k_1,k_2)| \\ &= \sup_{\xi,\eta\in D} |r_{k_1}(\xi) r_{k_2}(\eta) S(k_1,k_2)| \end{aligned}$$

which proves the lemma.

Proof of Lemma A.0.3:

If the bound (A.0.39) holds, one has

$$\begin{aligned} &|\phi_{i,m}(\xi_1,\xi_2)\phi_{j,m}(\xi_2,\xi_1) - C_i(\xi_1,\xi_2)C_j(\xi_2,\xi_1)| \\ &\qquad \le k r_m^{-z}(|C_i(\xi_1,\xi_2)| + |C_j(\xi_1,\xi_2)|) + k^2 r_m^{-2z} \end{aligned} \tag{A.0.42}$$

uniformly on D_m^2 and hence

$$\begin{aligned}
&\lim_{r_m\to\infty} r_m^d |A_{1,m} - A_m| \\
&\quad \le \lim_{r_m\to\infty} r_m^d \int_{D_m}\int_{D_m} \Big|\phi_{i,m}(\xi_1,\xi_2)\phi_{j,m}(\xi_2,\xi_1) - C_i(\xi_1,\xi_2)C_j(\xi_2,\xi_1)\Big| \\
&\qquad\qquad v(\xi_1,\xi_2) f_m(\xi_1) f_m(\xi_2) d\xi_1 d\xi_2 \\
&\qquad \le \lim_{r_m\to\infty} r_m^d \Bigg(k\alpha_G r_m^{-d} r_m^{-z} \Bigg(\int_0^{\mathrm{diam}(D_m)} |\mathbf{C}_i(\rho)|\nu(\rho)\rho^{d-1} d\rho \\
&\quad + \int_0^{\mathrm{diam}(D_m)} |\mathbf{C}_j(\rho)|\nu(\rho)\rho^{d-1} d\rho \Bigg) + k^2 r_m^{-2z} \int_0^{\mathrm{diam}(D_m)} \nu(\rho)\rho^{d-1} d\rho \Bigg) \\
&\qquad \le \lim_{r_m\to\infty} r_m^d (2kk_z\alpha_G r_m^{-d} r_m^{d-z-t} r_m^{-z} + k^2 r_m^{-2z-t}) \le \lim_{r_m\to\infty} k_2 r_m^{d-2z-t} = 0
\end{aligned}$$

since $z > d/2$, which proves (A.0.34). This clearly also holds if C_i satisfies (1.5.72) with $p = 1$. From (A.0.42) it also follows that

$$\begin{aligned}
&\lim_{r_m\to\infty} r_m^{2z+t} |h_{1,i,m}(\xi_1,\xi_2) - h_{i,m}(\xi_1,\xi_2)| \\
&\le \lim_{r_m\to\infty} r_m^{2z+t} \int_{D_m} |\phi_{i,m}(\xi_1,\lambda)\phi_{Y,m}(\lambda,\xi_2) - C_i(\xi_1,\lambda)C_Y(\lambda,\xi_2)| v(\xi_1,\lambda) f_m(\lambda) d\lambda \\
&\qquad \le \lim_{r_m\to\infty} r_m^{2z+t} \Bigg(k r_m^{-z} r_m^{-d} \int_{D_m} \Big(|C_i(\xi_1,\lambda) + C_Y(\xi_2,\lambda)| \Big) v(\xi_1,\lambda) d\lambda \\
&\qquad\qquad + k^2 r_m^{-2z} \int_{D_m} v(\xi_1,\lambda) f_m(\lambda) d\lambda \Bigg) \\
&\qquad \le \lim_{r_m\to\infty} r_m^{2z+t} (2kk_z r_m^{-z} r_m^{-d} r_m^{d-z-t} + k^2 k_v r_m^{-2z} r_m^{-d} r_m^{d-t}) \le k_3 < \infty \quad \text{(A.0.43)}
\end{aligned}$$

for some k_3 and this bound is uniform in (ξ_1,ξ_2). If $v(\xi,\eta) \equiv 1$ then $t = 0$. Clearly, if C_i satisfies (1.5.72) with $p = 1$, the result above holds for $t = 0$ with $z = d$. It follows that

$$\lim_{r_m\to\infty} r_m^{d+2z+t}|h_{1,i,m}(\xi_1,\xi_2)h_{1,j,m}(\xi_2,\xi_1) - h_{i,m}(\xi_1,\xi_2)h_{j,m}(\xi_2,\xi_1)|$$

$$= \lim_{r_m\to\infty} r_m^{d+2z+t}\Bigg|\int_{D_m} \phi_{i,m}(\xi_1,\lambda)\phi_{Y,m}(\lambda,\xi_2)v(\xi_1,\lambda)f_m(\lambda)d\lambda$$

$$\int_{D_m} \phi_{j,m}(\xi_1,\lambda)\phi_{Y,m}(\lambda,\xi_2)v(\xi_1,\lambda)f_m(\lambda)d\lambda$$

$$- \int_{D_m} C_i(\xi_1,\lambda)C_Y(\lambda,\xi_2)v(\xi_1,\lambda)f_m(\lambda)d\lambda$$

$$\int_{D_m} C_j(\xi_1,\lambda)C_Y(\lambda,\xi_2)v(\xi_1,\lambda)f_m(\lambda)d\lambda\Bigg|$$

$$\leq \lim_{r_m\to\infty} r_m^{d+2z+t}\Bigg(\int_{D_m} |\phi_{i,m}(\xi_1,\lambda)\phi_{Y,m}(\lambda,\xi_2)$$

$$- C_i(\xi_1,\lambda)C_Y(\lambda,\xi_2)|v(\xi_1,\lambda)f_m(\lambda)d\lambda$$

$$\int_{D_m} |C_j(\xi_1,\lambda)C_Y(\lambda,\xi_2)|v(\xi_1,\lambda)f_m(\lambda)d\lambda$$

$$+ \int_{D_m} |\phi_{j,m}(\xi_1,\lambda)\phi_{Y,m}(\lambda,\xi_2) - C_j(\xi_1,\lambda)C_Y(\lambda,\xi_2)|v(\xi_1,\lambda)f_m(\lambda)d\lambda$$

$$\int_{D_m} |C_i(\xi_1,\lambda)C_Y(\lambda,\xi_2)|v(\xi_1,\lambda)f_m(\lambda)d\lambda$$

$$+ \int_{D_m} |\phi_{i,m}(\xi_1,\lambda)\phi_{Y,m}(\lambda,\xi_2) - C_i(\xi_1,\lambda)C_Y(\lambda,\xi_2)|v(\xi_1,\lambda)f_m(\lambda)d\lambda$$

$$\int_{D_m} |\phi_{j,m}(\xi_1,\lambda)\phi_{Y,m}(\lambda,\xi_2) - C_j(\xi_1,\lambda)C_Y(\lambda,\xi_2)|v(\xi_1,\lambda)f_m(\lambda)d\lambda\Bigg)$$

$$\leq \lim_{r_m\to\infty} (r_m^d k_3 r_m^{-d}(\vartheta_i(\xi_1-\xi_2) + \vartheta_j(\xi_1-\xi_2)) + r_m^{d-2z-t}k_3^2)$$

$$\leq k_4(\vartheta_i(\xi_1-\xi_2) + \vartheta_j(\xi_1-\xi_2))$$

since $z > d/2$ and this bound also holds uniformly in (ξ_1, ξ_2) with

$$\begin{aligned}\vartheta_i(\xi-\eta) &= \int_{\mathbf{R}^d} |C_i(\xi,\lambda)v(\xi,\lambda)C_Y(\lambda,\eta)|d\lambda \\ &= \int_{\mathbf{R}^d} |C_i(\xi-\eta-\lambda,0)v(\xi-\eta-\lambda,0)C_Y(\lambda,0)|d\lambda \\ &= \int_{\mathbf{R}^d} |\psi_i(\xi-\eta-\lambda)\psi_Y(\lambda)|d\lambda = (|\psi_i| * |\psi_Y|)(\xi-\eta).\end{aligned}$$

It is easy to see that if (1.5.72) holds for C_i with $p = 1$ or $z + t > d$ then the function ψ_i is in L_1. Since by assumption $\psi_Y \in L_1$ as well, by an elementary property of the convolution, so is ϑ_i. In any case, by an earlier argument ϑ_i is at least guaranteed to be in L_2, and by Lemma A.0.2 it follows that

$$\lim_{r_m\to\infty} r_m^{-d/2} \int_{D_m}\int_{D_m} \vartheta(\xi-\eta)d\xi d\eta < \infty.$$

Finally

$$\begin{aligned}&\lim_{r_m\to\infty} r_m^{2d+z+t}|B_{1,m}(i,j) - B_m(i,j)| \\ &\qquad \le \lim_{r_m\to\infty} r_m^{2d+z+t} \\ &\int_{D_m}\int_{D_m} \left|h_{1,i,m}(\xi_1,\xi_2)h_{1,j,m}(\xi_2,\xi_1) - h_{i,m}(\xi_1,\xi_2)h_{j,m}(\xi_2,\xi_1)\right| f_m(\xi_1)f_m(\xi_2)d\xi_1 d\xi_2 \\ &\quad \le \lim_{r_m\to\infty} r_m^{d-z}k_4 \int_{D_m}\int_{D_m} (\vartheta_i(\xi_1-\xi_2) + \vartheta_j(\xi_1-\xi_2))f_m(\xi_1)f_m(\xi_2)d\xi_1 d\xi_2 \\ &\qquad \le \lim_{r_m\to\infty} r_m^{d-z}k_4\alpha_G r_m^{-d} \int_0^{\mathrm{diam}(D_m)} (\kappa_i(\rho) + \kappa_j(\rho))\rho^{d-1}d\rho \\ &\qquad\qquad \le \lim_{r_m\to\infty} r_m^{d-z}k_4k_5\alpha_1 r_m^{-d}r_m^{d/2} = 0\end{aligned}$$

where Lemma A.0.2 was applied to the functions κ_i and κ_j defined as in (A.0.31) and $z > d/2$. This proves (A.0.35) and (A.0.36). Thus Lemma A.0.3 is proved. Hence Theorem 1.5.7 follows.

Proof of Theorem 1.5.8.

If the nugget effect is included in the model, most of the proof of Theorem 1.5.7 carries over directly. It is clear that

$$\lim_{m\to\infty} a_m(i) = a(i) = \mathbf{C}_i(0).$$

The only non-trivial parts are (1.5.106), (1.5.108), (1.5.107) and (1.5.109). Firstly, one observes

$$\lim_{r_m\to\infty} r_m^{2d}\left|\int_{D_m}\int_{D_m} h_{j,m}(\xi,\eta)C_Y(\eta,\xi)f_m(\xi)f_m(\eta)d\xi d\eta\right| \leq \alpha_G r_m^{-2d}\int_0^{\mathrm{diam}(D_m)} |C_Y(\rho)|\kappa_j(\rho)\rho^{d-1}d\rho < \infty \quad \text{(A.0.44)}$$

and

$$\lim_{r_m\to\infty} r_m^{d}\left|\int_{D_m}\int_{D_m} C_Y(\xi,\eta)C_Y(\eta,\xi)f_m(\xi)f_m(\eta)d\xi d\eta\right| \leq \alpha_G r_m^{-d} d\int_0^{\mathrm{diam}(D_m)} (C_Y(\rho))^2\rho^{d-1}d\rho < \infty. \quad \text{(A.0.45)}$$

Next one observes that

$$h_{1,j,m}\phi_{Y,m} - h_{j,m}C_Y = (h_{1,j,m} - h_{j,m})(\phi_{Y,m} - C_Y) + (h_{1,j,m} - h_{j,m})C_Y + h_{j,m}(\phi_{Y,m} - C_Y).$$

Since the assumptions of Theorem 1.5.8 are the same as those of Theorem 1.5.7 one has

$$|h_{1,j,m}(\xi_1,\xi_2) - h_{j,m}(\xi_1,\xi_2)| \leq k_1 r_m^{-2z-t}$$

for some k_1 by (A.0.43). Moreover

$$|\phi_{Y,m}(\xi,\eta) - C_Y(\xi,\eta)| \leq k_2 r_m^{-d}$$

for some k_2 by Lemma A.0.4. Hence

$$\lim_{r_m\to\infty} r_m^{2d}\int_{D_m}\int_{D_m}|h_{1,j,m}(\xi,\eta)\phi_{Y,m}(\xi,\eta)-h_{j,m}(\xi,\eta)C_Y(\xi,\eta)|f_m(\xi)f_m(\eta)d\xi d\eta$$
$$\leq k_1k_2r_m^{2d-2z-t-d}+\alpha_G k_1 r_m^{2d-2z-t-d}\int_0^{\mathrm{diam}(D_m)}|\mathbf{C}_Y(\rho)|\rho^{d-1}d\rho$$
$$+\alpha_G k_2 r_m^{2d-d-2d}\int_0^{\mathrm{diam}(D_m)}|\kappa_i(\rho)|\rho^{d-1}d\rho \quad \text{(A.0.46)}$$

where κ_i is defined as in (A.0.31). Since $z > d/2$ and by Lemma A.0.2 applied to κ_i, the last limit is zero. Therefore

$$\lim_{r_m\to\infty} r_m^{2d}|b_m(j)|$$
$$=\lim_{r_m\to\infty} r_m^{2d}\left|\int_{D_m}\int_{D_m}h_{1,j,m}(\xi,\eta)\phi_{Y,m}(\eta,\xi)f_m(\xi)f_m(\eta)d\xi d\eta\right|<\infty.$$

Similarly,

$$\lim_{r_m\to\infty} r_m^{d}\int_{D_m}\int_{D_m}|\phi_{Y,m}(\xi,\eta)\phi_{Y,m}(\eta,\xi)-C_Y(\xi,\eta)C_Y(\eta,\xi)|f_m(\xi)f_m(\eta)d\xi d\eta$$
$$\leq\lim_{r_m\to\infty} r_m^{d}\int_{D_m}\int_{D_m}(2kr_m^{-d}|C_Y(\xi,\eta)|+k^2r_m^{-2d})f_m(\xi)f_m(\eta)d\xi d\eta=0$$

where Lemma A.0.4 is used for $\phi_{Y,m}$ and C_Y with $z=d$. Hence

$$\lim_{r_m\to\infty} r_m^{d}|b_{0,m}|=\lim_{r_m\to\infty} r_m^{d}\left|\int_{D_m}\int_{D_m}\phi_{Y,m}(\xi,\eta)\phi_{Y,m}(\eta,\xi)f_m(\xi)f_m(\eta)d\xi d\eta\right|=0.$$

The above inequalities imply that the following limits exist

$$b(j)=\lim_{m\to\infty} r_m^{2d}b_m(j)$$

$$b_0=\lim_{m\to\infty} r_m^{d}b_{0,m}.$$

Together with Theorem 1.5.8 this shows (1.5.106) and (1.5.108). The limit (1.5.107) is shown as follows

$$\begin{aligned}
&\lim_{n\to\infty} r_m^d \begin{bmatrix} 1 & a'_m \\ 0 & A_{1,m}^{-1} \end{bmatrix} \begin{bmatrix} b_{0,m} & b'_m \\ b_m & B_{1,m} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ a_m & A_{1,m}^{-1} \end{bmatrix} \\
&\quad = \lim_{n\to\infty} r_m^d \begin{bmatrix} 1 & a'_m \\ 0 & A_{1,m}^{-1} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & r_m^{-d} I_q \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & r_m^{d} I_q \end{bmatrix} \begin{bmatrix} b_{0,m} & b'_m \\ b_m & B_{1,m} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & r_m^{d} I_q \end{bmatrix} \\
&\qquad\qquad \begin{bmatrix} 1 & 0 \\ 0 & r_m^{-d} I_q \end{bmatrix} \begin{bmatrix} 1 & 0 \\ a_m & A_{1,m}^{-1} \end{bmatrix} \\
&\quad = \begin{bmatrix} 1 & a' \\ 0 & A^{-1} \end{bmatrix} \left(\lim_{n\to\infty} r_m^d \begin{bmatrix} 1 & 0 \\ 0 & r_m^{d} I_q \end{bmatrix} \begin{bmatrix} b_{0,m} & b'_m \\ b_m & B_{1,m} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & r_m^{d} I_q \end{bmatrix} \right) \begin{bmatrix} 1 & 0 \\ a & A^{-1} \end{bmatrix} \\
&\qquad\qquad = \begin{bmatrix} 1 & a' \\ 0 & A^{-1} \end{bmatrix} \begin{bmatrix} b_0 & b' \\ b & B \end{bmatrix} \begin{bmatrix} 1 & 0 \\ a & A^{-1} \end{bmatrix} \qquad \text{(A.0.47)}
\end{aligned}$$

where Theorem 1.5.8 was applied. The limit (1.5.109) is shown in a similar way, by applying Theorem 1.5.8 again.

Proof of Theorem 1.5.9:

The limit of A_R in (1.5.9) is straightforward

$$\begin{aligned}
A_R(i,j) &= \alpha_G \int_0^\infty \mathbf{C}_i(\rho)\mathbf{C}_j(\rho)\nu_{R,t}(\rho)\rho^{d-1}d\rho \\
&= \alpha_G \int_0^R \mathbf{C}_i(\rho)\mathbf{C}_j(\rho)\rho^{d-1}d\rho + \alpha_G \int_R^\infty \mathbf{C}_i(\rho)\mathbf{C}_j(\rho)\nu_{R,t}(\rho)\rho^{d-1}d\rho
\end{aligned}$$

it is clear that as $R \to \infty$ the last term in the sum above converges to zero, while the first one converges to $A(i,j)$. The limit of M_R is proved in the same way.

The notation of the proof of Theorem 1.5.6 will be used. To prove the limit for B_R one first observes

$$\kappa_{R,i}(\| \xi - \eta \|) = \zeta_{R,i}(\xi - \eta) = h_{R,i}(\xi, \eta)$$
$$= \int_{\mathbf{R}^d} \psi_{R,i}(\xi - \eta - \lambda)\psi_Y(\lambda)d\lambda = \int_{\mathbf{R}^d} \psi_{R,i}(\lambda)\psi_Y(\xi - \eta - \lambda)d\lambda$$
$$= \int_{\|\lambda\| \leq R} \psi_i(\lambda)\psi_Y(\xi - \eta - \lambda)d\lambda + \int_{\|\lambda\| > R} \psi_{R,i}(\lambda)\psi_Y(\xi - \eta - \lambda)d\lambda \quad \text{(A.0.48)}$$

again, clearly the last term of the last sum approaches zero, while the first term approaches $h_i(\xi, \eta)$. Next, one observes

$$\kappa_i(\| \xi - \eta \|) = \vartheta_i(\xi, \eta) = \int_{\mathbf{R}^d} |C_i(\xi, \lambda)C_Y(\lambda, \eta)|d\lambda \geq \vartheta_{R,i}(\xi, \eta)$$
$$= \int_{\mathbf{R}^d} |\psi_{R,i}(\xi - \eta - \lambda)\psi_Y(\lambda)|d\lambda \geq |h_{R,i}(\xi, \eta)|.$$

Finally, it is observed that

$$\int_0^\infty |\kappa_i(\rho)\kappa_j(\rho)|\rho^{d-1}d\rho < \infty$$

as is shown in the argument leading to (A.0.32). Hence Lebesgue's dominated convergence theorem applies and

$$\lim_{R\to\infty} \int_0^\infty H_{R,i}(\rho)H_{R,j}(\rho)\rho^{-d}d\rho = \int_0^\infty H_i(\rho)H_j(\rho)\rho^{-d}d\rho$$

as required. The proofs for θ_R and E_R are trivial. Thus the proof of Theorem 1.5.9 is complete.

Before proving Theorem 1.5.10, the following lemma will be established:

Lemma A.0.6. *Let $A_{m,n(m)}(\mathbf{X})$ be an $n(m) \times n(m)$ matrix whose (i, j) element is*

$$n(m)^{-2} \operatorname{tr}(K_{i,m,n(m)}(\mathbf{X})K_{j,m,n(m)}(\mathbf{X})) \quad \text{(A.0.49)}$$

and $M_{m,n(m)}(\mathbf{X})$ *be an* $n(m) \times 1$ *vector whose* i*-th element is*

$$n(m)^{-2} \operatorname{tr}(K_{i,m,n(m)}(\mathbf{X}) K_{Y,m,n(m)}(\mathbf{X})) \tag{A.0.50}$$

(where $K_{Y,m,n(m)}(\mathbf{X})$ *is defined analogously to* $K_{i,m,n(m)}(\mathbf{X})$*). Finally, let* $B_{m,n(m)}(\mathbf{X})$ *be an* $n(m) \times n(m)$ *matrix whose* (i,j) *element is*

$$n(m)^{-4} \operatorname{tr}(K_{i,m,n(m)}(\mathbf{X}) K_{Y,m,n(m)}(\mathbf{X}) K_{j,m,n(m)}(\mathbf{X}) K_{Y,m,n(m)}(\mathbf{X}))$$

If $\lim_{m\to\infty} r_m^d n(m)^{-1} = 0$ *then*

$$r_m^d A_{m,n(m)}(\mathbf{X}) - A \to_{p(\mathbf{X})} \mathbf{0} \tag{A.0.51}$$

$$r_m^d M_{m,n(m)}(\mathbf{X}) - M \to_{p(\mathbf{X})} \mathbf{0} \tag{A.0.52}$$

where A *and* M *are given by Theorem 1.5.6 and hence*

$$A_{m,n(m)}(\mathbf{X})^{-1} M_{m,n(m)}(\mathbf{X}) - \theta \to_{p(\mathbf{X})} \mathbf{0}. \tag{A.0.53}$$

Moreover, if $\lim_{m\to\infty} r_m^{3d} n(m)^{-1} = 0$ *then*

$$r_m^{3d} B_{m,n(m)}(\mathbf{X}) - B \to_{p(\mathbf{X})} \mathbf{0} \tag{A.0.54}$$

and hence

$$r_m^{3d/2} \sqrt{B_{m,n(m)}(\mathbf{X})(i,i)} - \sqrt{B(i,i)} \to_{p(\mathbf{X})} \mathbf{0}. \tag{A.0.55}$$

Proof of Lemma A.0.6:

As is shown in the proof of Lemma 1.5.6,

$$A_{m,n(m)}(\mathbf{X})(i,j) = n(m)^{-2} \binom{n(m)}{2} U_{n(m)} + V_{n(m)}$$

where U_n is a U-statistic whose kernel is $h(x_1, x_2) = C_i(x_1, x_2) C_j(x_2, x_1)$ and $V_{n(m)} = O(n(m)^{-1})$. Thus

$$\mathbf{E}[U_{n(m)}] = r_m^{-2d} \int_{D_m} \int_{D_m} C_i(\xi, \eta) C_j(\eta, \xi) d\xi d\eta$$

and from the proof of Theorem 1.5.6 it follows that

$$\lim_{m\to\infty} r_m^d \mathbf{E}[U_{n(m)}] = A(i,j).$$

Moreover, by Lemma A of Section 5.2.1 of Serfling (1980), one has

$$\begin{aligned}\mathrm{var}(U_{n(m)}) \leq 2n(m)^{-1}\Bigg(r_m^{-2d}\int_{D_m}\int_{D_m}(C_i(\xi,\eta)C_j(\eta,\xi))^2 d\xi d\eta \\ - r_m^{-4d}\left(\int_{D_m}\int_{D_m} C_i(\xi,\eta)C_j(\eta,\xi)d\xi d\eta\right)^2\Bigg) \leq c_1 n(m)^{-1} r_m^{-d}\end{aligned}$$

for some constant c_1 (the last inequality follows by an argument similar to the one in the proof of Theorem 1.5.6). Hence (A.0.51) of Lemma A.0.6 follows. In a similar fashion one proves (A.0.52) of Lemma A.0.6. The proof of A.0.53 of Lemma A.0.6 proceeds in the same fashion as that of proof of Lemma 1.5.6.

The proof of (A.0.54) is similar to that of (A.0.51). Firstly, it is sufficient to consider only the diagonal elements of $B_{m,n(m)}(\mathbf{X})$

$$B_{m,n(m)}(\mathbf{X})(i,i) = n(m)^{-4}\binom{n(m)}{4}U_{n(m)} + V_{n(m)}$$

where U_n is a U-statistic whose kernel is

$$h(x_1,x_2,x_3,x_4) = C_i(x_1,x_2)C_Y(x_2,x_3)C_i(x_3,x_4)C_Y(x_4,x_1)$$

and $V_{n(m)} = O(n(m)^{-1})$. Thus

$$\begin{aligned}&\mathbf{E}[U_{n(m)}] \\ &= r_m^{-4d}\int_{D_m}\int_{D_m}\left(\int_{D_m} C_i(\xi,\lambda)C_Y(\lambda,\eta)d\lambda\right)\left(\int_{D_m} C_j(\xi,\lambda)C_Y(\lambda,\eta)d\lambda\right)d\xi d\eta\end{aligned}$$

and from the proof of Theorem 1.5.6 it follows that

$$\lim_{m\to\infty} r_m^{3d}\mathbf{E}[U_{n(m)}] = B(i,i).$$

Moreover, by Lemma A of Section 5.2.1 of Serfling (1980), one has

$$\begin{aligned}\mathrm{var}(U_{n(m)}) \leq 4n(m)^{-1}\Bigg(r_m^{-4d}\int_{D_m}\int_{D_m}\Bigg(\int_{D_m}(C_i(\xi,\lambda)C_Y(\lambda,\eta))^2 d\lambda\Bigg)\\ \Bigg(\int_{D_m}(C_j(\xi,\lambda)C_Y(\lambda,\eta))^2 d\lambda\Bigg)d\xi d\eta\\ - r_m^{-8d}\Bigg(\int_{D_m}\int_{D_m}\Bigg(\int_{D_m}C_i(\xi,\lambda)C_Y(\lambda,\eta)d\lambda\Bigg)\\ \Bigg(\int_{D_m}C_j(\xi,\lambda)C_Y(\lambda,\eta)d\lambda\Bigg)d\xi d\eta\Bigg)^2\Bigg) \leq c_2 n(m)^{-1}r_m^{-3d}\end{aligned}$$

for some constant c_2 (the last inequality follows by an argument similar to the one in the proof of Theorem 1.5.6). Hence (A.0.54) follows. Given that $B(i,i) > 0$ the proof of (A.0.55) is straightforward.

Proof of Theorem 1.5.10:

Let the assumptions hold. It follows from (A.0.53) that

$$\mathbf{E}_\omega[\hat{\theta}_{m,n(m)}(\omega,\mathbf{X})] - \theta \rightarrow_{p(\omega,\mathbf{X})} \mathbf{0}$$

and therefore it is sufficient to show

$$\hat{\theta}_{m,n(m)}(\omega,\mathbf{X}) - \mathbf{E}_\omega[\hat{\theta}_{m,n(m)}(\omega,\mathbf{X})] \rightarrow_{p(\omega,\mathbf{X})} \mathbf{0}.$$

By definition

$$\hat{\theta}_{m,n(m)}(\omega,\mathbf{X}) = A_{m,n(m)}(\mathbf{X})^{-1}Z_{m,n(m)}(\omega,\mathbf{X})$$

where $A_{m,n(m)}(\mathbf{X})$ is as in (A.0.49) and $Z_{m,n(m)}(\omega,\mathbf{X})$ is an $n(m)\times 1$ vector whose i-th element is

$$n(m)^{-2}\,\mathrm{tr}(K_{i,m,n(m)}(\mathbf{X})Y_{m,n(m)}(\omega,\mathbf{X})Y_{m,n(m)}(\omega,\mathbf{X})') \tag{A.0.56}$$

and

$$\mathbf{E}[\hat{\theta}_{m,n(m)}(\omega,\mathbf{X})] = A_{m,n(m)}(\mathbf{X})^{-1}M_{m,n(m)}(\mathbf{X}).$$

It follows from (A.0.51) that

$$(r_m^d A_{m,n(m)}(\mathbf{X}))^{-1} - A^{-1} \to_{p(\omega,\mathbf{X})} 0.$$

Thus by Slutzky's theorem it would be sufficient to show that

$$r_m^d Z_{m,n(m)}(\omega, \mathbf{X}) - r_m^d M_{m,n(m)}(\mathbf{X}) \to_{p(\omega,\mathbf{X})} 0.$$

By (1.4.25) one has the bound

$$\mathrm{var}_\omega(r_m^d Z_{m,n(m)}(\omega, \mathbf{X})(i)|\mathbf{X} = X) \leq c r_m^{2d} B_{m,n(m)}(X)(i,i).$$

Let $\delta > 0$ and $\epsilon > 0$ and $\delta_1 > 0$. The set $S_{\mathbf{X}}(m)$ will be defined so that

$$X \in S_{\mathbf{X}}(m) \iff |\, r_m^{3d/2} \sqrt{B_{m,n(m)}(X)(i,i)} - \sqrt{B(i,i)}\,| < \delta_1$$

from (A.0.55) it follows that there exists M_1 such that

$$m \geq M_1 \Rightarrow \mathbf{P}_{\mathbf{X}}(X \in S_{\mathbf{X},m}) \geq 1 - \frac{\epsilon}{2}. \tag{A.0.57}$$

From Chebyshev's inequality, one has for all X

$$\mathbf{P}_\omega\left(|r_m^d Z_{m,n(m)}(\omega, X) - r_m^d M_{m,n(m)}(X)| > \sqrt{2c/\epsilon}\, r_m^d \sqrt{B_{m,n(m)}(X)(i,i)}\right) < \frac{\epsilon}{2}. \tag{A.0.58}$$

Let the set $S_\omega^1(m, X)$ be defined by

$$\omega \in S_\omega^1(m, X) \iff |r_m^d Z_{m,n(m)}(\omega, X) - r_m^d M_{m,n(m)}(X)| \leq \sqrt{2c/\epsilon}\, r_m^d \sqrt{B_{m,n(m)}(X)(i,i)}.$$

Hence

$$\mathbf{P}_\omega(\omega \in S_\omega^1(m, X)) \geq 1 - \frac{\epsilon}{2}. \tag{A.0.59}$$

If $X \in S_{\mathbf{X}}(m)$ one has

$$\sqrt{B_{m,n(m)}(X)(i,i)} \leq r_m^{-3d/2}\left(\delta_1 + \sqrt{B(i,i)}\right).$$

Let the set $S^2_\omega(m,X)$ be defined by

$$\omega \in S^2_\omega(m,X) \iff |r_m^d Z_{m,n(m)}(\omega,X) - r_m^d M_{m,n(m)}(X)| \leq \sqrt{2c/\epsilon}\, r_m^{-d/2}\left(\delta_1 + \sqrt{B(i,i)}\right).$$

Hence if $X \in S_{\mathbf{X}}(m)$ one has $S^1_\omega(m,X) \subseteq S^2_\omega(m,X)$. Let M_2 be such that

$$m \geq M_2 \Rightarrow \sqrt{2c/\epsilon}\, r_m^{-d/2}\left(\delta_1 + \sqrt{B(i,i)}\right) \leq \delta.$$

Now for $m \geq \max\{M_1, M_2\}$ one has

$$\begin{aligned}
&\mathbf{P}_{(\omega,\mathbf{X})}\left(|r_m^d Z_{m,n(m)}(\omega,X) - r_m^d M_{m,n(m)}(X)| > \delta\right) \\
&\quad = \mathbf{P}_\omega\left(|r_m^d Z_{m,n(m)}(\omega,X) - r_m^d M_{m,n(m)}(X)| > \delta | X \in S_{\mathbf{X}}(m)\right) \\
&\qquad \mathbf{P}_{\mathbf{X}}\left(X \in S_{\mathbf{X}}(m)\right) \\
&\quad + \mathbf{P}_\omega\left(|r_m^d Z_{m,n(m)}(\omega,X) - r_m^d M_{m,n(m)}(X)| > \delta | X \in S_{\mathbf{X}}(m)^c\right) \\
&\qquad \mathbf{P}_{\mathbf{X}}\left(X \in S_{\mathbf{X}}(m)^c\right) \\
&\quad \leq \mathbf{P}_\omega\left(\omega \in S^2_\omega(m,X)^c | X \in S_{\mathbf{X}}(m)\right) + \frac{\epsilon}{2} \\
&\quad \leq \mathbf{P}_\omega\left(\omega \in S^1_\omega(m,X)^c | X \in S_{\mathbf{X}}(m)\right) + \frac{\epsilon}{2} \leq \epsilon.
\end{aligned}$$

This concludes the proof of Teorem 1.5.10.

BIBLIOGRAPHIE

ADLER R.J. (1980). *The Geometry of Random Fields.* John Wiley & Sons.

BARTLETT, M.S. (1964). Spectral analysis of two-dimensional point processes. *Biometrika* **44** 299-311.

BROWN, K.G. (1978). Estimation of variance components using residuals. *Journal of the American Statistical Association* **73** 141 - 146.

CHILES, J.P., DELFINER, P. (1999). *Geostatistics: modeling spatial uncertainty.* John Wiley & Sons.

CRESSIE, N.A.C (1993). *Statistics for Spatial Data.* Revised edition. John Wiley & Sons.

CRESSIE, N.A.C. and GRONDONA, M. O. (1992). A comparison of variogram estimation with covariogram estimation. In *The Art of Statistical Science*, edited by K.V. Mardia, Wiley, Chichester, 191-208.

DAVIS, J.C. (1973). *Statistics and Data Analysis in Geology.* John Wiley & Sons.

DEUTSCH, C. V. and JOURNEL, A. G. (1992). *GSLIB: Geostatistical Software Library and User's Guide.* Oxford University Press.

DIGGLE, P.J. (1983). *Statistical Analysis of Spatial Point Patterns.* Academic Press.

HALL, P. and PATIL, P. (1994). Properties of nonparametric estimators of autocovariance for stationary random field. *Probability Theory and Related Fields* **99** 399-424.

ISAAKS E. H. and SRIVASTAVA, R. M. (1989) *Applied Geostatistics.* Oxford University Press.

KITANIDIS, P. K. (1985). Minimum variance unbiased quadratic estimation of covariances of regionalized variables. *Journal of the International Association for Mathematical Geology* **17** 195-208.

LAHIRI, S. N. (1996). On inconsistency of estimators under infill asymptotics for spatial data. *Sankhya* Ser. A, **58** 403-417.

LAHIRI, S. N., KAISER, M. S. , CRESSIE, N. and HSU, N. (1999). Prediction of spatial cumulative distribution functions using subsampling. *Journal of the American Statistical Association* **94** 86-97.

MATHERON, G. (1965). *Les variables régionalisées et leur estimation.* Masson et Cie.

PARTHASARATY, K. (1967). *Probability Measures on Metric Spaces.* Academic Press.

POWOJOWSKI, M. (1999a). Additive Covariogram Models and Estimation through Projections. In preparation.

POWOJOWSKI, M. (1999b). Estimation of the Covariogram of an Isotropic Process through Spectral Component Additive Models. In preparation.

POWOJOWSKI, M. (1999c). Model Selection in Covariogram Estimation. In preparation.

RAO, C.R. (1971). Estimation of variance and covariance components - MINQUE theory. *Journal of Multivariate Analysis* **1** 257 - 275.

RAO, C.R. and KLEFFE, J. (1988). *Estimation of Variance Components and Applications.* North-Holland.

RIPLEY, B.D. (1988). *Statistical Inference for Spatial Processes.* Cambridge University Press.

SCHOENBERG, I. J. (1938). Metric spaces and completely monotone functions. *The Annals of Mathematics* **39** 811-841.

SERFLING, R. J (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons.

SHAPIRO, A. and BOTHA, J.D. (1991). Variogram fitting with a general class of conditionally nonnegative definite functions. *Computational Statistics and Data Analysis* **11** 87-96.

SNEDDON, I.N (1972). *The Use of Integral Transforms.* McGraw-Hill.

STEIN, M. L. (1987). Minimum norm quadratic estimation of spatial variograms. *Journal of the American Statistical Association* **82** 765-772.

STEIN, M. L. (1989). Asymptotic distribution of minimum norm quadratic estimators of the covariance function of a Gaussian random field. *The Annals of Statistics* **17** 980-1000.

STEIN, M. L. and HANDCOCK, M. S. (1989). Some asymptotic properties of kriging when the covariance function is misspecified. *Journal of the International Association for Mathematical Geology* **21** 839-861.

VERDOOREN, L.R. (1988). Least squares estimators and non-negative estimators of variance components. *Communications in Statistics A - Theory Methods* **17** 1287 - 1309.

WACKERNAGEL, H. (1995). *Multivariate geostatistics.* Springer-Verlag.