University of Wollongong
Research Online

Faculty of Informatics - Papers (Archive)

Faculty of Engineering and Information Sciences

2004

Flexible spatial models for kriging and cokriging using moving averages and the fast Fourier Transform (FFT)

Jay M. Ver Hoef
Alaska Fisheries Science Center

Noel A. Cressie
University of Wollongong, ncressie@uow.edu.au

Ronald P. Barry
University of Alaska

Follow this and additional works at: https://ro.uow.edu.au/infopapers

 Part of the Physical Sciences and Mathematics Commons

Recommended Citation

Ver Hoef, Jay M.; Cressie, Noel A.; and Barry, Ronald P.: Flexible spatial models for kriging and cokriging using moving averages and the fast Fourier Transform (FFT) 2004, 265-282.
https://ro.uow.edu.au/infopapers/2366

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Flexible spatial models for kriging and cokriging using moving averages and the fast Fourier Transform (FFT)

Abstract

Models for spatial autocorrelation and cross-correlation depend on the distance and direction separating two locations, and are constrained so that for all possible sets of locations, the covariance matrices implied from the models remain nonnegative-definite. Based on spatial correlation, optimal linear predictors can be constructed that yield complete maps of spatial fields from incomplete and noisy spatial data. This methodology is called kriging if the data are of only one variable type, and it is called cokriging if it is of two or more variable types. Historically, to satisfy the nonnegative-definite condition, cokriging has used coregionalization models for cross-variograms, even though this class of models is not very flexible. Recent research has shown that moving-average functions may be used to generate a large class of valid, flexible variogram models, and that they can also be used to generate valid cross-variograms that are compatible with component variograms. There are several problems with the moving-average approach, including large numbers of parameters and difficulties with integration. This article shows how the fast Fourier Transform (FFT) solves these problems. The flexible moving-average function that we consider is composed of many small rectangles, which eliminates the integration problem. The FFT allows us to compute the cross-variogram on a set of discrete lags; we show how to interpolate the cross-variogram for any continuous lag, which allows us to fit flexible models using standard minimization routines. Simulation examples are given to demonstrate the methods.

Keywords

spatial, models, kriging, cokriging, flexible, moving, fft, averages, fast, fourier, transform

Disciplines

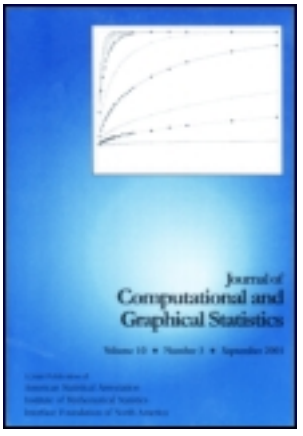
Physical Sciences and Mathematics

Publication Details

Ver Hoef, J. M., Cressie, N. A. & Barry, R. P. (2004). Flexible spatial models for kriging and cokriging using moving averages and the fast Fourier Transform (FFT). Journal of Computational and Graphical Statistics, 13 (2), 265-282.

This journal article is available at Research Online: https://ro.uow.edu.au/infopapers/2366

This article was downloaded by: [University of Wollongong]
On: 27 November 2012, At: 18:18
Publisher: Taylor & Francis
Informa Ltd Registered in England and Wales Registered Number: 1072954
Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of Computational and Graphical Statistics

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ucgs20

Flexible Spatial Models for Kriging and Cokriging Using Moving Averages and the Fast Fourier Transform (FFT)

Jay M. Ver Hoef, Noel Cressie and Ronald Paul Barry
Jay M. Ver Hoef is Biometrician, Alaska Department of Fish and Game, 1300 College Road, Fairbanks, AK 99701 . Noel Cressie is Professor, Department of Statistics, The Ohio State University, Columbus, OH, 43210-1247. Ronald Paul Barry is Professor, Department of Mathematical Sciences, University of Alaska, Fairbanks, AK 99775.
Version of record first published: 01 Jan 2012.

To cite this article: Jay M. Ver Hoef, Noel Cressie and Ronald Paul Barry (2004): Flexible Spatial Models for Kriging and Cokriging Using Moving Averages and the Fast Fourier Transform (FFT), Journal of Computational and Graphical Statistics, 13:2, 265-282

To link to this article: http://dx.doi.org/10.1198/1061860043498

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to

date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Downloaded by [University of Wollongong] at 18:18 27 November 2012

Downloaded by [University of Wollongong] at 18:18 27 November 2012

Flexible Spatial Models for Kriging and Cokriging Using Moving Averages and the Fast Fourier Transform (FFT)

Jay M. VER HOEF, Noel CRESSIE, and Ronald Paul BARRY

Models for spatial autocorrelation and cross-correlation depend on the distance and direction separating two locations, and are constrained so that for all possible sets of locations, the covariance matrices implied from the models remain nonnegative-definite. Based on spatial correlation, optimal linear predictors can be constructed that yield complete maps of spatial fields from incomplete and noisy spatial data. This methodology is called kriging if the data are of only one variable type, and it is called cokriging if it is of two or more variable types. Historically, to satisfy the nonnegative-definite condition, cokriging has used coregionalization models for cross-variograms, even though this class of models is not very flexible. Recent research has shown that moving-average functions may be used to generate a large class of valid, flexible variogram models, and that they can also be used to generate valid cross-variograms that are compatible with component variograms. There are several problems with the moving-average approach, including large numbers of parameters and difficulties with integration. This article shows how the fast Fourier Transform (FFT) solves these problems. The flexible moving-average function that we consider is composed of many small rectangles, which eliminates the integration problem. The FFT allows us to compute the cross-variogram on a set of discrete lags; we show how to interpolate the cross-variogram for any continuous lag, which allows us to fit flexible models using standard minimization routines. Simulation examples are given to demonstrate the methods.

Key Words: Geostatistics; Spatial statistics; Variogram.

1. INTRODUCTION

Spatial statistics, and in particular geostatistics, is concerned with sampling and inference in a spatial environment. The most important feature that distinguishes geostatistics

Jay M. Ver Hoef is Biometrician, Alaska Department of Fish and Game, 1300 College Road, Fairbanks, AK 99701 (E-mail: ffjmv@uaf.edu). Noel Cressie is Professor, Department of Statistics, The Ohio State University, Columbus, OH, 43210-1247. Ronald Paul Barry is Professor, Department of Mathematical Sciences, University of Alaska, Fairbanks, AK 99775.

©2004 American Statistical Association, Institute of Mathematical Statistics, and Interface Foundation of North America
Journal of Computational and Graphical Statistics, Volume 13, Number 2, Pages 265–282
DOI: 10.1198/1061860043498

Downloaded by [University of Wollongong] at 18:18 27 November 2012

from classical statistics is that geostatistics uses the spatial coordinates to model statistical dependence among data, rather than assuming that the data are independent. The spatial dependence is often termed autocorrelation, if the data are of one type (e.g., the yield of grain at various locations); if the data are of different types (e.g., the yield of grain *and* concentration of nitrogen at various, sometimes different locations), it is often termed cross-correlation. Models for spatial autocorrelation and spatial cross-correlation are constrained so that for all possible sets of locations, the covariance matrix implied from the models remains nonnegative definite. Several models have been developed for autocorrelation, the most popular being the spherical, exponential, and linear (see Cressie 1993, sec. 2.3, for a list that includes these and others). For cross-correlation, coregionalization models have been proposed (Journel and Huijbregts 1978, p. 171; Isaaks and Srivastava 1989, p. 390; Goovaerts 1997, p. 107). However, there have been problems finding very flexible classes of coregionalization models that satisfy the nonnegative-definiteness constraint referred to above. Daley (1991, chap. 5) gave cross-correlation models derived from physical principles such as geostrophy in atmospheric science. In an article that develops moving-average forms of cross-covariance, as does our article, Gaspari and Cohn (1999) gave analytical expressions for limited classes of autocovariance and cross-covariance functions.

In geostatistics, it is common to express spatial dependence through variograms and cross-variograms rather than through autocorrelation and cross-correlation. Variograms and cross-variograms have the potential to yield slightly more general classes of spatial dependence. However, to realize the potential of this extra generality has proved problematic. For example, the coregionalization models require that the cross-variogram and each of the component variograms must share the same basic set of variogram models, and so it is quite restrictive (Papritz, Kunsch, and Webster 1993; Helterbrand and Cressie 1994; Goovaerts 1997, p. 123; Ver Hoef and Barry 1998; and Yao and Journel 1998). It will be seen that by considering autocorrelation and cross-correlation and building models based on the fast Fourier transform (FFT), considerable progress can be made. Yao and Journel (1998) also considered using the FFT, but their approach is different than ours. First, they modeled nonparametrically by computing the empirical covariance and then smoothing it in the spectral domain to attain the positive-definiteness condition. Thus, they did not specify any covariance or variogram a priori. Second, they did not actually produce covariance models for continuous lags. Because of the discrete nature of the FFT, the back transformation from the spectral domain to the spatial domain produces a covariance table for discrete lag values only; Yao and Journel (1998) suggested using the nearest available lag value from the covariance table.

Moving averages are an intuitive way to construct valid statistical models for spatial data, and this has been recognized sporadically in the past, with textbook treatments given by Yaglom (1987a,b), Matern (1986), and Thiebaux and Pedder (1987). However, there are some practical problems associated with these moving-average representations. First, they depend on integrals that may be difficult or impossible to solve analytically. [Gaspari and Cohn (1999) gave analytical expressions for limited classes of autocovariance and cross-covariance functions.] When integral solutions are possible, the method is flexible enough

to model nonstationary data (Higdon 1998; Hidgon, Swall, and Kern 1999) and count and proportion data (Wolpert and Ickstadt 1998). Analytical solutions are desirable because we often want to fit the models to data, which usually involves some numerical minimization. If the integrals of the variogram/cross-variogram models also require numerical solutions, the computational burden may be too great. To get around this problem, Barry and Ver Hoef (1996) considered moving-average functions that are composed of many small rectangles. This allows for easy integration but also creates the problem of many parameters to estimate, because the height over each rectangle can take on a unique value. The ideas are readily extended to cokriging (Ver Hoef and Barry 1998). A similar problem of minimizing an integral function of many parameters was studied by Mockus (1998), although his models are different than ours. The objectives of this article are to show how the fast Fourier Transform (FFT) can be used to help solve the problems of integration of many parameters, and how that allows more flexible models for kriging and cokriging. Several simulation examples demonstrate the efficacy of the methods.

Downloaded by [University of Wollongong] at 18:18 27 November 2012

2. MODELS FOR AUTOCORRELATION AND CROSS-CORRELATION

Consider a vector-valued spatial process $\{\mathbf{Z}(\mathbf{s}) : \mathbf{s} \in D\}$, where $D \subseteq \mathcal{R}^d$ is the d-dimensional spatial domain of interest and $\mathbf{Z}(\mathbf{s}) \equiv [Z_1(\mathbf{s}), Z_2(\mathbf{s}), \ldots, Z_L(\mathbf{s})]'$. Define second-order stationarity as follows: Assume that, for the jth spatial process $Z_j(\bullet)$, the mean $E[Z_j(\mathbf{s})]=\mu_j$, for all $\mathbf{s} \in \mathcal{D}$, and the covariance,

$$C_{jj}(\mathbf{h}) \equiv \text{cov}[Z_j(\mathbf{s}), Z_j(\mathbf{s}+\mathbf{h})], \tag{2.1}$$

exists. The quantity $C_{jj}(\mathbf{h})$ is called the autocovariance. Data on $Z_j(\bullet)$ are collected at J_j locations; and assume that the data are a realization of the random vector $\mathbf{Z}_j \equiv [Z_j(\mathbf{s}_{1,j}), Z_j(\mathbf{s}_{2,j}), \ldots, Z_j(\mathbf{s}_{J_j,j})]'$.

The cross-covariance is defined as

$$C_{jk}(\mathbf{s}, \mathbf{u}) \equiv \text{cov}[Z_j(\mathbf{s}), Z_k(\mathbf{u})]. \tag{2.2}$$

Another quantity that measures spatial dependence across variables is the cross-variogram,

$$2\gamma_{jk}(\mathbf{s}, \mathbf{u}) \equiv \text{var}[Z_j(\mathbf{s}) - Z_k(\mathbf{u})]. \tag{2.3}$$

Notice that (2.3) is different from another quantity, which is often used in geostatistics to express cross-spatial dependence; namely,

$$2\tau_{jk}(\mathbf{s}, \mathbf{u}) \equiv \text{cov}[Z_j(\mathbf{s}) - Z_j(\mathbf{u}), Z_k(\mathbf{s}) - Z_k(\mathbf{u})], \tag{2.4}$$

and this has also been called a cross-variogram. There is now a considerable body of literature (Cressie 1993, p. 140; Myers 1991; Ver Hoef and Cressie 1993; Papritz et al. 1993; Cressie and Wikle 1998; Ver Hoef and Barry 1998; Royle and Berliner 1999) that demonstrates the limitations of (2.4) as a measure of cross-spatial dependence and we shall

not consider it further in this article. Also, we will promote the use of estimation using restricted maximum likelihood, so we consider only $C_{jk}(\mathbf{s}, \mathbf{u})$ from now on. By working with autocorrelation and cross-correlation as building blocks, we shall build models in (2.2) that depend only on $\mathbf{h} = \mathbf{u} - \mathbf{s}$, and in anticipation of this we write the cross-covariance as $2C_{jk}(\mathbf{h})$; note that $C_{jk}(\mathbf{s}, \mathbf{u})$ has domain $\mathcal{R}^d \times \mathcal{R}^d$, $C_{jk}(\mathbf{h})$ has domain $\mathcal{R}^d$, and $C_{jk}(||\mathbf{h}||)$ has domain $[0, \infty)$, where $||\mathbf{h}||$ is Euclidean distance.

2.1 MOVING-AVERAGE REPRESENTATIONS

Barry and Ver Hoef (1996) showed that a large class of variograms (with a sill) can be developed by integration of a moving-average function over a white-noise random process. Start with the following spatial processes: $W_k(\mathbf{x})$ is a zero mean white noise random process with $\mathbf{x} \in \mathcal{R}^d$; $k = 0, 1, 2, \ldots, L$; that is, $E[W_k(\mathbf{x})] = 0$, $\text{var}[\int_A W_k(\mathbf{x})d\mathbf{x}] = \mid A \mid$, $\text{cov}[\int_A W_k(\mathbf{x})d\mathbf{x}, \int_B W_k(\mathbf{x})d\mathbf{x}] = 0$ when $A \cap B = \emptyset$ and $W_k(\mathbf{x})$ is independent of $W_m(\mathbf{x})$ when $k \neq m$. Now, define

$$Y_k(\mathbf{x} \mid \rho_k, \boldsymbol{\Delta}_k) = \sqrt{1 - \rho_k^2} W_k(\mathbf{x}) + \rho_k W_0(\mathbf{x} - \boldsymbol{\Delta}_k).$$

Let $U_k(\bullet)$ be another white noise process with $E[U_k(\mathbf{x})] = 0$ and $\text{var}[U_k(\mathbf{x})] = 1$. Also, let $U_k(\mathbf{x})$ be independent of $U_k(\mathbf{t})$ for all $\mathbf{x} \neq \mathbf{t}$, and let $U_k(\mathbf{x})$ be independent of $U_m(\mathbf{t})$ for all $k \neq m$ and all $\mathbf{x}$ and $\mathbf{t}$.

Then for some $\mathbf{s} \in \mathcal{D} \subset \mathcal{R}^d$ let

$$Z_k(\mathbf{s}|\theta_k, \nu_k, \mu_k, \rho_k, \Delta_k) \equiv \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g_k(\mathbf{x} - \mathbf{s}|\boldsymbol{\theta}_k) Y_k(\mathbf{x} \mid \rho_k, \boldsymbol{\Delta}_k) d\mathbf{x} + \nu_k U_k(\mathbf{s}) + \mu_k, \tag{2.5}$$

where $g_k(\mathbf{x}|\boldsymbol{\theta}_k)$ is Riemann-integrable. The moving-average construction allows a valid autocovariance to be expressed as

$$C_{jj}(\mathbf{h}|\boldsymbol{\theta}_j, \nu_j) = \begin{cases} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (g_j(\mathbf{u}|\boldsymbol{\theta}_j))^2 d\mathbf{u} + \nu_j^2, & \text{for} \quad \mathbf{h} = \mathbf{0}, \\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g_j(\mathbf{u}|\boldsymbol{\theta}_j) g_j(\mathbf{u} - \mathbf{h}|\boldsymbol{\theta}_j) d\mathbf{u}, & \text{for} \quad \mathbf{h} \neq \mathbf{0}, \end{cases} \tag{2.6}$$

where we assume that the integrals exist; here $g_j(\mathbf{u}|\theta_j)$ is called the moving-average function and it is defined on $\mathcal{R}^d$. Note that for this class of models, the autocovariance (2.6) has a nugget effect (a discontinuity at the origin that adds an additional variance component ν_j^2 when $\mathbf{h} = \mathbf{0}$). The remaining part, $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g_j(\mathbf{u}|\theta_j) g_j(\mathbf{u} - \mathbf{h}|\boldsymbol{\theta}_j) d\mathbf{u}$, controls the behavior of the autocovariance as $\mathbf{h}$ changes, and it can be thought of as the well-known "autocorrelation" function from Fourier analysis. Autocovariance models have sills, although it is possible to construct moving averages for models for variograms without sills (Lindstrom 1993).

It is equivalent to develop the cross-covariance for the quantities $R_j(\mathbf{s}) \equiv Z_j(\mathbf{s}) - \mu_j$, where recall that μ_j is the mean of the jth process. Ver Hoef and Barry (1998) showed this construction for cross-variograms; the cross-covariance (2.3) for the mean-centered

Downloaded by [University of Wollongong] at 18:18 27 November 2012

processes $R_j(\bullet)$ is defined to be:

$$C_{jk}(\mathbf{h}|\boldsymbol{\theta}, \rho, \Delta) = \rho_j \rho_k \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g_j(\mathbf{u}|\boldsymbol{\theta}_j) g_k(\mathbf{u} - \mathbf{h} + \Delta_k - \boldsymbol{\Delta}_j|\theta_k) d\mathbf{u}, \qquad (2.7)$$

such that $\boldsymbol{\theta} \equiv (\theta_j', \boldsymbol{\theta}_k')'$, $\rho \equiv (\rho_j, \rho_k)'$, and $\boldsymbol{\Delta} \equiv (\Delta_j', \boldsymbol{\Delta}_k')'$. Notice that for cross-covariances, there are parameters $\boldsymbol{\rho}$ and $\boldsymbol{\Delta}$ that express the strength and shift-asymmetry of cross-spatial dependence; that is, $\boldsymbol{\rho}$ is the cross-correlation between white noise processes (Barry and Ver Hoef 1996). In practice, we can let $\boldsymbol{\Delta}_1 = \mathbf{0}$ and all subsequent $\boldsymbol{\Delta}_j; j \neq 1$ are relative to $\boldsymbol{\Delta}_1$. Also note that for $L = 2$ variables, ρ_j and ρ_k will not be identifiable, but their product is identifiable, and for $L > 2$, ρ_j and ρ_k will be identifiable. Observe that the number of cross-covariances grows as $(L-1)L/2$, while the number of parameters $\{\boldsymbol{\Delta}_j\}$ grows as $L-1$ and the number of parameters $\{\rho_j\}$ grows as L. It would also be possible to allow cross-covariance between nugget-effect components; however, we often think of this as measurement error and so we do not pursue it further here. It would require additional cross-correlation parameters similar to ρ_j and ρ_k.

The moving-average constructions (2.6) and (2.7) are attractive because they are flexible; we can choose any pair of moving-average functions $g_j(\mathbf{u}|\boldsymbol{\theta}_j)$ and $g_k(\mathbf{u}|\boldsymbol{\theta}_k)$ that are square integrable. The resulting autocovariances and cross-covariances will yield valid kriging and cokriging equations. However, as we discussed in the introduction, analytical evaluations of the integrals in (2.6) and (2.7) can be difficult. Numerical evaluations of the integrals are not practical if the functions will then be fitted to empirical data using numerical minimization, as is often the case. We solve this problem by using many small rectangles in (2.7), which have easy integrals.

2.2 FLEXIBLE SPATIAL COVARIANCES FROM SMALL RECTANGLES

The FFT allows us to compute the autocovariance on a set of discrete shifts, but we desire an autocovariance for any continuous $\mathbf{h}$. Barry and Ver Hoef (1996) showed that a very flexible autocovariance can be used if the moving-average function $g(\mathbf{u}; \boldsymbol{\theta})$ is composed of many small rectangles. In one dimension, let

$$g_j(x|a_1, \ldots, a_M, c, M) = \sum_{m=1}^{M} a_m \mathcal{I}\left(\frac{2(m-1)c}{M} < x + c \leq \frac{2mc}{M}\right), \qquad (2.8)$$

where $\mathcal{I}(\bullet)$ is the indicator function. The function (2.8) has support $(-c, c]$, and consists of M steps of equal width $2c/M$ and heights $a_1, \ldots, a_M$ (Figure 1). Let $h_p^* = 2pc/M$, where p is any integer. That is, $2c/M$ is the width of the rectangles in (2.8), and the sides of the rectangles line up at $\{h_p^*\}$; see Figure 1. Then it is easy to see from (2.6) and (2.8) that

$$C_{jj}(h_p^*) = \big(2c/M\big) \sum_{m=1}^{M} a_m a_{m+p},$$

Downloaded by [University of Wollongong] at 18:18 27 November 2012

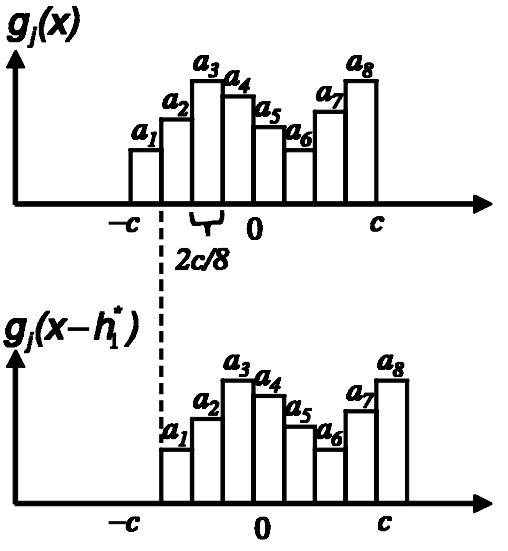


Figure 1. Moving-average construction in one-dimension, where the moving-average function is composed of small rectangles. The figure shows a lag where the sides of the rectangles line up at h_1^*.

where $a_m = 0$ if $m < 1$ or $m > M$. For parameter estimation, it is convenient to rescale by using $\widetilde{a}_m = a_m / \sqrt{\sum_{m=1}^{M} a_m^2}$. Then the autocovariance becomes

$$C_{jj}(h_p^*) = \sigma_j^2 \sum_{m=1}^{M} \widetilde{a}_m \widetilde{a}_{m+p},$$

where $\sigma_j^2 \equiv \frac{2c}{M} \sum_{m=1}^{M} a_m^2$ is the usual sill parameter found in many semivariogram models. Now, let $h_L = \left\lfloor \frac{Mh}{2c} \right\rfloor \frac{2c}{M}$ and $h_U = \left\lceil \frac{Mh}{2c} \right\rceil \frac{2c}{M}$ for any given h, where $\lfloor x \rfloor$ is the nearest integer less than x and $\lceil x \rceil$ is the nearest integer greater than x. If $f = (h - h_L)M/2c$ is the fraction of the distance that h is from h_L to h_U, then Barry and Ver Hoef (1996) showed that for the moving-average functions given by (2.8),

$$C_{jj}(h) = (1 - f)C_{jj}(h_L) + fC_{jj}(h_U),$$

for any h. Notice that this is a linear interpolation of the upper and lower autocovariance values. This construction does not include a nugget effect ν_j^2, which can easily be added.

The extension of these ideas to $d = 2$ dimensions is as follows. Let the moving-average function be

$$g_j(x, y; a_{11}, \ldots, a_{MN}, c, d, M, N) = \sum_{m=1}^{M} \sum_{n=1}^{N} a_{mn} \mathcal{I}\left(\frac{2(m-1)c}{M} < x + c \le \frac{2mc}{M} \quad \text{and} \quad \frac{2(n-1)d}{N} < y + d \le \frac{2nd}{N} \right). \tag{2.9}$$

Let $\mathbf{h}_{pq}^* = (2pc/M, 2qd/N)'$, where p and q are positive integers, negative integers, or zero.

Downloaded by [University of Wollongong] at 18:18 27 November 2012

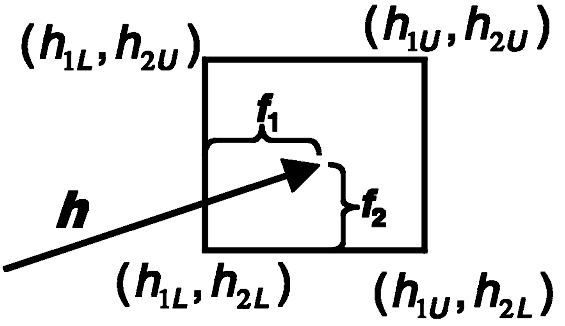


Figure 2. Moving-average construction in two dimensions, where the spatial lag $\mathbf{h}$ *is some fraction of the grid used for the FFT.*

Downloaded by [University of Wollongong] at 18:18 27 November 2012

Then from (2.6), with $\nu_j^2 = 0$,

$$C_{jj}(\mathbf{h}_{pq}^*) = [(4cd)/(MN)] \sum_{m=1}^{M} \sum_{n=1}^{N} a_{mn} a_{m+p,n+q},$$

where $a_{mn} = 0$ if $m < 1$, $m > M$, $n < 1$, or $n > N$. Again, we rescale so that

$$C_{jj}(\mathbf{h}_{pq}^*) = \sigma_j^2 \sum_{m=1}^{M} \sum_{n=1}^{N} \widetilde{a}_{mn} \widetilde{a}_{m+p,n+q}, \tag{2.10}$$

where $\sigma_j^2 \equiv \frac{4cd}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} a_{mn}^2$. Here, let $h_{1L} = \lfloor \frac{Mh_1}{2c} \rfloor \frac{2c}{M}$, $h_{1U} = \lceil \frac{Mh_1}{2c} \rceil \frac{2c}{M}$, $h_{2L} = \lfloor \frac{Nh_2}{2d} \rfloor \frac{2d}{N}$, $h_{2U} = \lceil \frac{Nh_2}{2d} \rceil \frac{2d}{N}$, for any $\mathbf{h} = (h_1, h_2)'$, and let $f_1 = (h_1 - h_{1L})M/2c$ and $f_2 = (h_2 - h_{2L})N/2d$. Then

$$\begin{aligned} C_{jj}(\mathbf{h}) = (1-f_1)(1-f_2) C_{jj}(h_{1L}, h_{2L}) + (1-f_1) f_2 C_{jj}(h_{1L}, h_{2U}) \\ + f_1 (1-f_2) C_{jj}(h_{1U}, h_{2L}) + f_1 f_2 C_{jj}(h_{1U}, h_{2U}), \end{aligned} \tag{2.11}$$

for any $\mathbf{h}$ (see Figure 2); this is easily verified using (2.6) and (2.9). Notice that this is a linear interpolation of the four autocovariance values. This construction does not include a nugget effect ν_j^2, which can be easily added.

Likewise, if we consider a second variable, $Z_k(\mathbf{s})$ with

$$\begin{aligned} &g_k(x, y; b_{11}, \ldots, b_{MN}, c, d, M, N) \\ &= \sum_{m=1}^{M} \sum_{n=1}^{N} b_{mn} \mathcal{I}\left(\frac{2(m-1)c}{M} < x + c \le \frac{2mc}{M} \quad \text{and} \quad \frac{2(n-1)d}{N} < y + d \le \frac{2nd}{N} \right), \end{aligned}$$

then from (2.6), with $\nu_k^2 = 0$,

$$C_{kk}(\mathbf{h}_{pq}^*) = \sigma_k^2 \sum_{m=1}^{M} \sum_{n=1}^{N} \widetilde{b}_{mn} \widetilde{b}_{m+p,n+q}, \tag{2.12}$$

where $\sigma_k^2 \equiv \frac{4cd}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} b_{mn}^2$. Futhermore, from (2.7) with $\boldsymbol{\Delta}_j = \boldsymbol{\Delta}_k = \mathbf{0}$,

$$C_{jk}(\mathbf{h}_{pq}^*) = \rho \sigma_j \sigma_k \sum_{m=1}^{M} \sum_{n=1}^{N} \widetilde{a}_{mn} \widetilde{b}_{m+p,n+q}. \tag{2.13}$$

Then it can be shown likewise that,

$$C_{jk}(\mathbf{h}) = (1-f_1)(1-f_2)C_{jk}(h_{1L},h_{2L}) + (1-f_1)f_2C_{jk}(h_{1L},h_{2U}) + f_1(1-f_2)C_{jk}(h_{1U},h_{2L}) + f_1f_2C_{jk}(h_{1U},h_{2U}), \quad (2.14)$$

for any $\mathbf{h}$ (see Figure 2), and once again this is a linear interpolation of the four cross-covariance values. Nonzero shift parameters in (2.13) are easily incorporated; the cross-covariance becomes $C_{jk}(\mathbf{h}+\Delta_j-\boldsymbol{\Delta}_k)$. A separate model is also possible for cross-covariance between nugget-effect components but we do not pursue it here.

Downloaded by [University of Wollongong] at 18:18 27 November 2012

2.3 THE FAST FOURIER TRANSFORM (FFT)

Now (2.6) and (2.7) can be viewed as convolutions of functions, which are well known to transform to products in the spectral domain. Hence, it makes sense to take the Fourier transform of the moving-average functions, and the FFT allows us to obtain (2.10), (2.12), and (2.13) rapidly for all $\mathbf{h}^*_{pq}$, as we shall now describe. Let $x_p^+ = \frac{2pc-c(M-1)}{M}$, for $p = 0,\ 1,\ldots,M-1$, and $y_q^+ = \frac{2qd-d(N-1)}{N}$, for $q = 0,\ 1,\ldots,N-1$, which are the coordinates for the centers of the rectangular grid described in (2.9). The two-dimensional FFT is

$$G_j\left(\frac{m}{c},\frac{n}{d}\right) = \sum_{q=0}^{N-1}\left[\sum_{p=0}^{M-1} g_j\left(x_p^+,y_q^+\right)e^{-i2\pi mp/M}\right]e^{-i2\pi nq/N},$$

for $m = 0,1,\ldots,M-1$ and $n = 0,1,\ldots,N-1$, where $i=\sqrt{-1}$. Then the autocorrelation and cross-correlation functions are denoted as $r_{jk}(p,q)$, which is

$$\frac{1}{MN}\sum_{n=0}^{N-1}\left[\sum_{m=0}^{M-1} G_j\left(\frac{m}{c},\frac{n}{d}\right)G_k^*\left(\frac{m}{c},\frac{n}{d}\right)e^{-i2\pi mp/M}\right]e^{-i2\pi nq/N},$$

for $p = 0,\ 1,\ldots,M-1$ and $q = 0,\ 1,\ldots,N-1$, where $G_k^*(x,y)$ is the complex conjugate of $G_k(x,y)$. In practice, the values in $r_{jk}(p,q)$ need to be rearranged (see Brigham 1988, p. 244) and rescaled so that $r_{jk}(0,0) = \sum_{m=1}^{M}\sum_{n=1}^{N}\widetilde{a}_{mn}\widetilde{b}_{m,n}$. After rearranging and rescaling, we obtain $r_{jk}(h_p^*,h_q^*) = \sum_{m=1}^{M}\sum_{n=1}^{N}\widetilde{a}_{mn}\widetilde{b}_{m+p,n+q}$, for $h_p^* = 2pc/M$, with p an integer such that $-M/2 \le p \le M/2$, and q an integer such that $-N/2 \le q \le N/2$. Using (2.10), we obtain the autocovariance (without any nugget effect) for discrete lags as

$$C_{jj}(\mathbf{h}^*_{pq}) = \sigma_j^2 r_{jj}\left(h_p^*,h_q^*\right), \quad (2.15)$$

and $C_{jj}(\mathbf{h})$ for any other $\mathbf{h}$ is given by (2.11). The cross-covariance (2.13) (without shifts) for discrete lags is

$$C_{jk}(\mathbf{h}^*_{pq}) = \rho\sigma_j\sigma_k r_{jk}\left(h_p^*,h_q^*\right), \quad (2.16)$$

and $C_{jk}(\mathbf{h})$ for any $\mathbf{h}$ is given by (2.14). Notice that when $j = k$ (so that $\rho = 1$ in (2.16)), $C_{jj}(\mathbf{h}^*_{pq}) = C_{jk}(\mathbf{h}^*_{pq})$, as it should.

The importance of using the FFT is that it is very fast. Computing $C_{jk}(\mathbf{h}^*_{pq})$ for all possible $\mathbf{h}^*_{pq}$, for p an integer such that $-M/2 \le p \le M/2$ and q an integer such that $-N/2 \le q \le N/2$, by brute force is on the order of $(MN)^2$ operations, but the FFT only requires on the order of $MN \log_2(MN)$ operations.

2.4 REDUCING THE NUMBER OF PARAMETERS

So far, we have allowed all $\{a_{mn}\}$ to be unconstrained, so we have MN free parameters. We can reduce the number of parameters by allowing some functional relationships among the $\{a_{mn}\}$. For example, let (x^+_m, y^+_n) be the coordinates at the center of the (m,n)th rectangle from the moving-average function. We can construct a variety of moving-average functions in the following way. Let some function be defined between -1 and 1; for example, let the moving-average function of (2.6) be

$$g(x,y) = \sqrt{1-x^2-y^2}\,\mathcal{I}\left(\sqrt{x^2+y^2} < 1\right); \quad -1 \le x \le 1, \quad -1 \le y \le 1,$$

which is half of a unit sphere defined on the square $[-1,1]^2$. We simply let the heights of the boxes be $a_{mn} = g(x^+_m, y^+_n)$. We use this example to illustrate the type of cross-covariances that can be generated. In Figure 3, the moving-average function for variable 1 is on the left, the moving-average function for variable 2 is in the center (they are scaled using $\widetilde{a}_m = a_m/\sqrt{\sum_{m=1}^M a_m^2}$ and $\widetilde{b}_m = b_m/\sqrt{\sum_{m=1}^M b_m^2}$), and the resulting cross-covariance surface is on the right. Figure 3(a) shows moving-average functions that are the same. We can perform scale transformations on the x- and y-coordinates (shrink or expand them) to make the half sphere larger, or shrink x more than y (Figure 3(b)) and give the coordinates a rotation (Figure 3(c)), which yields an anisotropic function of moving-average heights,

$$g_t(x,y|\theta) = \sqrt{1-x_t^2-y_t^2}\,\mathcal{I}\left(\sqrt{x_t^2+y_t^2} < 1\right); \quad -1 \le x \le 1, \quad -1 \le y \le 1,$$

where $x_t = (x\cos(\theta_1) - y\sin(\theta_1))/\theta_2$, and $y_t = (x\sin(\theta_1) + y\cos(\theta_1))/\theta_3$. The parameter $-\pi/2 \le \theta_1 \le \pi/2$ is the rotation parameter, $\theta_2 > 0$ is a range parameter that scales the x-coordinates, and $\theta_3 > 0$ is a range parameter that scales the y-coordinates. In Figure 3(c), we let $\theta_1 = 45$ degrees, $\theta_2 = 0.5$ and $\theta_3 = 1$. We can further change the shape of $g_t(x,y|\theta)$ by including a power parameter $\theta_4 > 0$:

$$g_t(x,y|\theta) = (1-x_t^2-y_t^2)^{\theta_4}\,\mathcal{I}\left(\sqrt{x_t^2+y_t^2} < 1\right); \quad -1 \le x \le 1, \quad -1 \le y \le 1. \quad (2.17)$$

In Figure 3(d), we let $\theta_4 = 10$ for the moving-average function for variable 2, rotate with $\theta_1 = -\pi/4$, and set $\rho = 1$. We use (2.16) because we have a cross-covariance. An autocovariance (2.15) is just a special case of (2.16) where both moving averages are the same. In Figure 3(e), we allow a shift in the coordinates for the moving-average function for variable 2; $x_t = ((x-\Delta_1)\cos(\theta_1) - (y-\Delta_2)\sin(\theta_1))/\theta_2$, and $y_t = ((x-\Delta_1)\sin(\theta_1) + (y-\Delta_2)\cos(\theta_1))/\theta_3$. When there is a shift, our notational convention is to absorb the extra (shift) parameters into the vector θ. In Figure 3(e), we let $\Delta_1 = \Delta_2 = 0.5$. Notice that

Downloaded by [University of Wollongong] at 18:18 27 November 2012

Downloaded by [University of Wollongong] at 18:18 27 November 2012

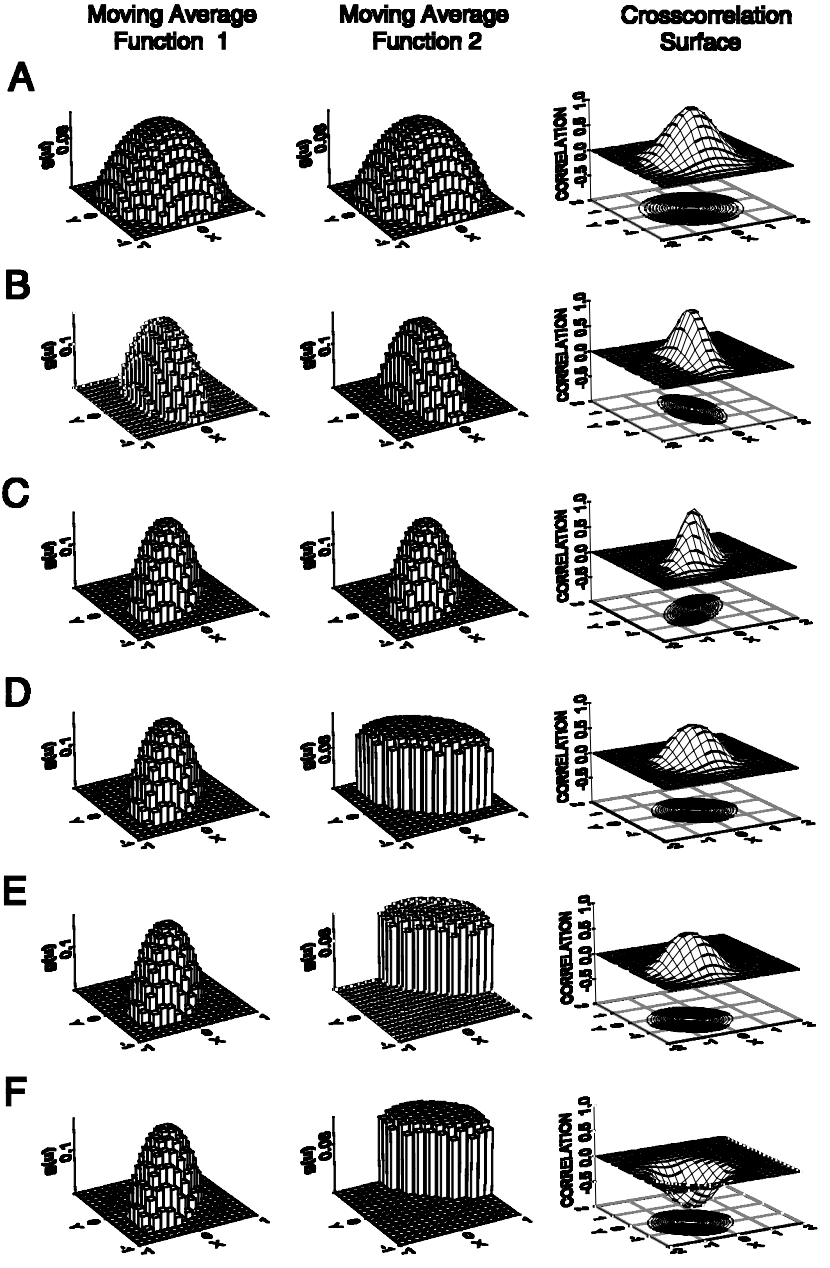


Figure 3. Moving-average construction in two dimensions. The left panel is the moving-average function for variable 1, the middle panel is the moving-average function for variable 2, and the right panel is the corresponding autocovariance/cross-covariance. (a) The height of the small boxes follow a half-sphere function. (b) The moving-average function is transformed. (c) The moving-average function is rotated. (d) The moving-average function for variable 2 is rotated differently than for variable 1, and the shape is changed. (e) The moving-average function for variable 2 is shifted. (f) The cross-correlation parameter is made negative rather than positive.

the resulting shift causes an asymmetry in the cross-covariance. Finally, we allow negative cross-correlation by taking $\rho = -0.9$ (Figure 3(f)).

In spatial modeling, it is a difficult problem to pick autocorrelation functions. For example, should one choose a spherical, exponential, or K-Bessel variogram model? Based on the formulation given in this article, the equivalent problem is to choose a moving-average function. Many of the same principles apply here as for variograms; one can choose moving average functions based on: (1) the fit of the model to empirical data, (2) likelihood or AIC values, (3) cross-validation methods, or (4) a class rich enough to handle most situations. The K-Bessel (Matern) variogram satisfies this last criterion because it has an extra parameter that controls the behavior of the variogram near the origin (Stein 1999, p. 49). The model described by $g(x, y|\theta)$ (2.17) also yields flexible covariances near the origin due to the parameter θ_4. This is especially important for cokriging because it is often difficult to model the spatial autocorrelation and cross-correlation if the moving-average functions are too simple. Because the moving average function implies an autocovariance and a variogram, these can be checked against empirical autocovariances and variograms; Cressie and Ver Hoef (2001) used (2.17) for agricultural data and compare the fitted model to empirical variograms. Cressie and Pavlicova (2002) showed how to choose a moving-average model based on desired properties of a given autocovariance near the origin.

Downloaded by [University of Wollongong] at 18:18 27 November 2012

3. FITTING THE AUTOCOVARIANCE AND CROSS-COVARIANCE TO THE DATA

We used restricted maximum likelihood (REML) to fit moving-average models, and consequently autocovariance and cross-covariances, to data. Let us collect all parameters into one vector, $\mathbf{\Phi} \equiv (\boldsymbol{\theta}'\boldsymbol{\nu}'\boldsymbol{\rho}'\boldsymbol{\Delta}')'$. Using (2.6) and (2.7), we obtain the covariance matrix $\mathbf{\Sigma}_{\mathbf{\Phi}}$, which depends on parameters $\mathbf{\Phi}$, for a set of spatial data. The REML equation to be minimized is

$$L(\mathbf{\Phi}) = (n-p)\log(2\pi) + \log|\mathbf{\Sigma}_{\mathbf{\Phi}}| + (\mathbf{z} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{\mathbf{\Phi}})'\mathbf{\Sigma}_{\mathbf{\Phi}}^{-1}(\mathbf{z} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{\mathbf{\Phi}}) + \log|\mathbf{X}'\mathbf{\Sigma}_{\mathbf{\Phi}}^{-1}\mathbf{X}|, \quad (3.1)$$

where $\widehat{\boldsymbol{\beta}}_{\mathbf{\Phi}} = (\mathbf{X}'\mathbf{\Sigma}_{\mathbf{\Phi}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}_{\mathbf{\Phi}}^{-1}\mathbf{z}$. If there are many observations, then inverting the matrix $\mathbf{\Sigma}_{\mathbf{\Phi}}$ is not practical. In the example that follows, we used a method similar to the suggestion of Stein (1999, p. 172). Suppose we have 2,000 values for $Z_2(\bullet)$ and 100 values for $Z_1(\bullet)$. Then we randomly divided the 2,000 values for $Z_2(\bullet)$ into 10 subsets. Each subset was combined with the 100 values for $Z_1(\bullet)$, and we then summed the restricted likelihoods for each. This is equivalent to minimizing (3.1) for

$$\mathbf{\Sigma}_{\mathbf{\Phi}} = \begin{pmatrix} \mathbf{\Sigma}_{p,p} & \mathbf{\Sigma}_{p,1} & \mathbf{\Sigma}_{p,2} & \cdots & \mathbf{\Sigma}_{p,10} \\ \mathbf{\Sigma}_{p,1} & \mathbf{\Sigma}_{1,1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{\Sigma}_{p,2} & \mathbf{0} & \mathbf{\Sigma}_{2,2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{p,10} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Sigma}_{10,10} \end{pmatrix},$$

Table 1. Parameters Used to Simulate Data Using a Half-Sphere Moving-Average Function. The simulated data were used to estimate the true parameters, for both the univariate case for variable 1 only, and the bivariate case including both variables.

Parameter	*True value*	*Univariate estimate*	*Bivariate estimate*
Variable 1			
θ_1 (rotate)	0.222π	0.231π	0.237π
θ_2 (scale *x*)	0.400	0.268	0.301
θ_3 (scale *y*)	0.800	0.533	0.544
θ_4 (power)	2.000	1.567	1.050
σ (partial sill)	10.00	8.197	9.611
ν (nugget)	0.100	0.031	0.005
Variable 1			
θ_1 (rotate)	0.278π		0.277π
θ_2 (scale *x*)	0.500		0.511
θ_3 (scale *y*)	1.000		1.054
θ_4 (power)	0.500		1.186
σ (partial sill)	5.000		4.158
ν (nugget)	0.100		0.116
Cross-Variable			
Δ_x (shift *x*)	0.100		0.111
Δ_y (shift *y*)	0.100		0.093
ρ (correlation)	0.900		0.898

Downloaded by [University of Wollongong] at 18:18 27 November 2012

where the subscript p indicates the $Z_1(\bullet)$ data were used in computing the relevant covariance matrices, and the subscripts $1, 2, \ldots, 10$ indicate that the 10 randomly selected subsets of the $Z_2(\bullet)$ data were used in computing the relevant covariance matrices. For kriging, Stein (1999, p. 172) recommended spatially compact blocks rather than random ones, but it is important to estimate the sill well in order to estimate ρ (which controls the cross-covariance) well, so we wanted some longer spatial lags. This is an area that needs further research, although we used this only for estimation of covariance parameters—for prediction and prediction standard errors, we used the full kriging and cokriging equations based on neighborhood constraints; see Section 4. Rapid computing using the FFT allows us to compute models directly, which is essential when minimizing (3.1). Several examples are given in the next section.

4. A SIMULATION EXAMPLE

To demonstrate that we can recover true spatial dependencies, and use them to carry out spatial prediction, we simulated bivariate spatial data using moving averages. Simulations used moving averages over a systematic grid of 100×100 nodes in the range $-3 \le x \le 3$ and $-3 \le y \le 3$. At each node, we generated a pair of $N(0, 1)$ variables with correlation ρ, with each pair independent of any other pair; denote them by $W_1(\mathbf{s}_k)$ and $W_2(\mathbf{s}_k)$, for $k = 1, 2, \ldots, 10{,}000$. The spatially dependent variables were then simulated by

$$Z_i(\mathbf{s}) = \frac{\sigma_i}{\sqrt{\sum_{k=1}^{10,000} g_t(\mathbf{s} - \mathbf{s}_k|\boldsymbol{\theta}_i)^2}} \sum_{k=1}^{10,000} g_t(\mathbf{s}_k - \mathbf{s}|\boldsymbol{\theta}_i) W_i(\mathbf{s}_k - \boldsymbol{\Delta}_i) + \nu_i U_i(\mathbf{s}) + \mu_i,$$

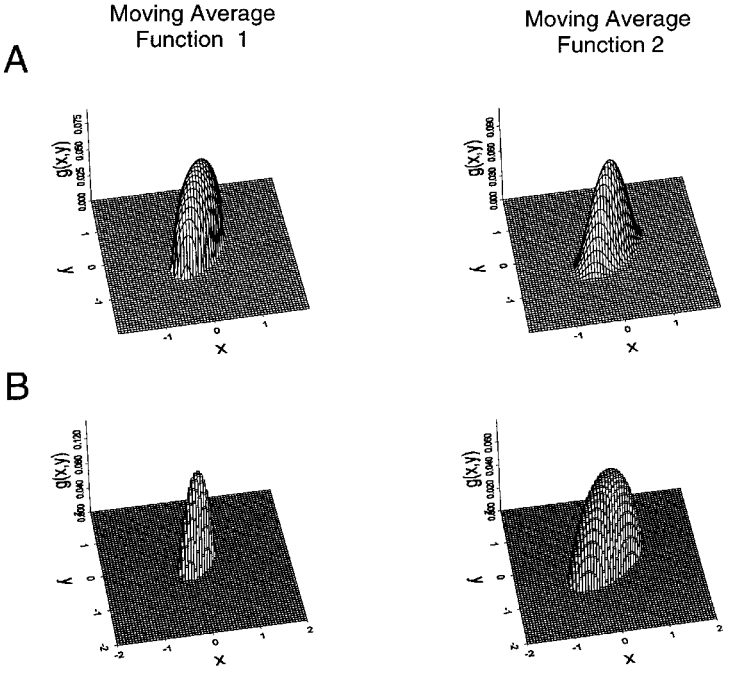


Figure 4. True and estimated moving-average models for simulated data. (a) The true moving-average models for variables 1 and 2. (b) The fitted moving average models for variables 1 and 2.

where $U_i(\mathbf{s})$ is an independent $N(0,1)$ variable at each $\mathbf{s}$, $g_t(\mathbf{s}|\boldsymbol{\theta}_i)$ is given by (2.17), and $\mu_i \equiv E[Z_i(\mathbf{s})]$, for $i = 1, 2$. Note that we used independent standard Gaussian random variables as our "building blocks," but other distributions could be used to create non-Gaussian data. All covariance parameters used to simulate the data are given in the first column of Table 1; regarding the shifts, we let $\boldsymbol{\Delta}_1 = \mathbf{0}$, and $\boldsymbol{\Delta}_2 = (\Delta_x, \Delta_y)'$, which are shown in Table 1. The true moving-average models, with $g_t(x, y)$ in (2.17) for variables 1 and 2, are shown in Figure 4(a). We let $\mu_1 = 100$ and $\mu_2 = 0$. Bivariate spatial data were obtained at 2,000 spatial locations, which were randomly chosen on $[-2, 2] \times [-2, 2]$, according to the uniform probability mass function. By simulating at locations within this window, we avoided edge effects when constructing the moving average. More details were given by Ver Hoef and Barry (1998); see also Oliver (1995).

For cokriging, one variable is usually expensive or difficult to obtain, while the other is easier and less expensive. Therefore, we randomly selected 100 of the 2,000 locations of variable 1 to be used as the dataset. Hereafter, think of variable 1 as the more expensive but more sparsely sampled variable that we wish to predict. These 100 data values on variable 1 were used for kriging and then, they and the additional 2,000 values on variable 2, were used for cokriging variable 1. In practice, we used those values of variable 2 that were within a distance of 0.2 from the prediction location. This cokriging neighborhood is important for computational reasons, keeping matrix inversions possible for the covariance matrices.

The remaining 1,900 values for variable 1 were used as a validation dataset, with each

Downloaded by [University of Wollongong] at 18:18 27 November 2012

value predicted by both kriging and cokriging. In this way, the predicted value could be compared to the true (simulated) value. The kriging and cokriging equations were given by Ver Hoef and Cressie (1993) and Ver Hoef and Barry (1998). Other literature relevant to multivariate spatial prediction are Gotway and Hartford (1996), Wackernagel (1998), and Royle (2000).

We first estimated the covariance parameters defined in (2.17), by minimizing the REML Equations (3.1). We initially minimized (3.1) based on only the 100 observations for variable 1, which we label univariate estimation in Table 1. These parameter estimates are used for kriging. Then we minimized the REML equations for both variables and all parameters simultaneously, which we label bivariate estimation in Table 1. These parameter estimates are used for cokriging. The fitted moving averages are shown in Figure 4(b), using the discrete approximation of $g_t(x, y)$ given by (2.17), where we let $M = N = 64$, and $c = d = 2$. Table 1 and Figure 4 show that we can obtain good estimates of moving-average functions using REML.

To check model estimation and to compare methods, we used some summary statistics from validation. Let $\widehat{Z}_1(\mathbf{s}_i)$ be the (co)kriged value of the variable 1 at location $\mathbf{s}_i$ for the ith datum in the 1,900 values of variable 1 that were set aside for validation; for now, we suppress the difference in notation between cokriging and kriging. We checked for bias in both kriging and cokriging by considering $(1/n_1)\sum_{i=1}^{n_1}[\widehat{Z}_1(\mathbf{s}_i) - Z_1(\mathbf{s}_i)]$, where $n_1 = 1{,}900$. We used the root-mean-squared-prediction error,

$$\text{RMSPE} \equiv \sqrt{\sum_{i=1}^{n_1}(\widehat{Z}_1(\mathbf{s}_i) - Z_1(\mathbf{s}_i))^2/n_1},$$

to assess the predictive ability of both kriging and cokriging. Let $\widehat{\text{var}}[\widehat{Z}_1(\mathbf{s}_i)]$ be the estimated prediction variance at location $\mathbf{s}_i$. Then let $\text{RMEV} \equiv (\sum_{i=1}^{n_1}\widehat{\text{var}}[\widehat{Z}_1(\mathbf{s}_i)]/n_1)^{1/2}$. If the estimated prediction variances are correct, then RMEV should be close to RMSPE. We also wanted to assess whether the estimated prediction standard errors were valid. If we denote $\widehat{\text{se}}(\widehat{Z}_1(s_i)) \equiv (\widehat{\text{var}}[\widehat{Z}_1(\mathbf{s}_i)])^{1/2}$, then the prediction interval coverage,

$$80\%\text{PI} \equiv \sum_{i=1}^{n_1}\mathcal{I}[|\widehat{Z}_1(s_i) - Z_1(\mathbf{s}_i)| < 1.28 * \widehat{\text{se}}(\widehat{Z}_1(s_i))]$$

should be about 0.80, where recall that $\mathcal{I}[\bullet]$ denotes the indicator function. For kriging, we used the fitted model based on variable-1 data, and for cokriging we used the fitted model based on the data from both variables. The results are given in Table 2. From Table 2, there is negligible bias, there is a 6.3% improvement when using cokriging over kriging based on RMSPE, and both methods appear to have valid prediction variances and prediction intervals.

At the suggestion of a referee, we did one more simulation to highlight the comparison of kriging to cokriging. This time, we kept 200 $\{Z_1(\mathbf{s}_i)\}$ values and predicted the remaining 1,800. Also, we thinned the $\{Z_1(\mathbf{s}_i)\}$ values going from left to right (Figure 5, top panel). Then, we computed the RMSPE for all predictions within the bands $x = -2$ to -1.5,

Downloaded by [University of Wollongong] at 18:18 27 November 2012

Table 2. Validation Statistics for Kriging and Cokriging

Validation statistics	*Kriging*	*Cokriging*
Bias	0.258	0.212
RMSPE	1.821	1.707
RMEV	1.725	1.717
80%PI	0.797	0.807

-1.5 to -1.0, etc. These are plotted in the bottom panel of Figure 5. Notice that when the primary variable is dense, there is little advantage for cokriging. However, as the primary variable gets thinner, cokriging becomes more advantageous. A plot of cokriging predictions versus true values, for each band, is given in Figure 6. Notice that cokriging is always a better

Downloaded by [University of Wollongong] at 18:18 27 November 2012

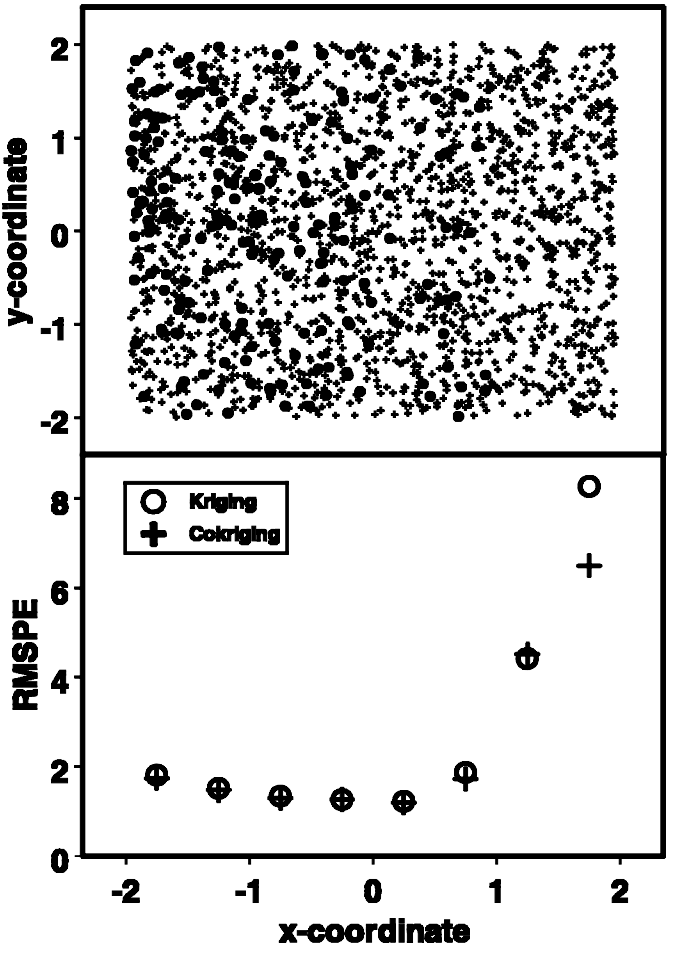


Figure 5. Simulated data that are progressively thinned as x increases. In the top panel, the solid circles are the 200 locations for variable 1 and the crosses are the 2,000 locations for variable 2. The bottom panel shows the validation RMSPE for kriging and cokriging for bands (of 0.5 width) along the x-coordinate.

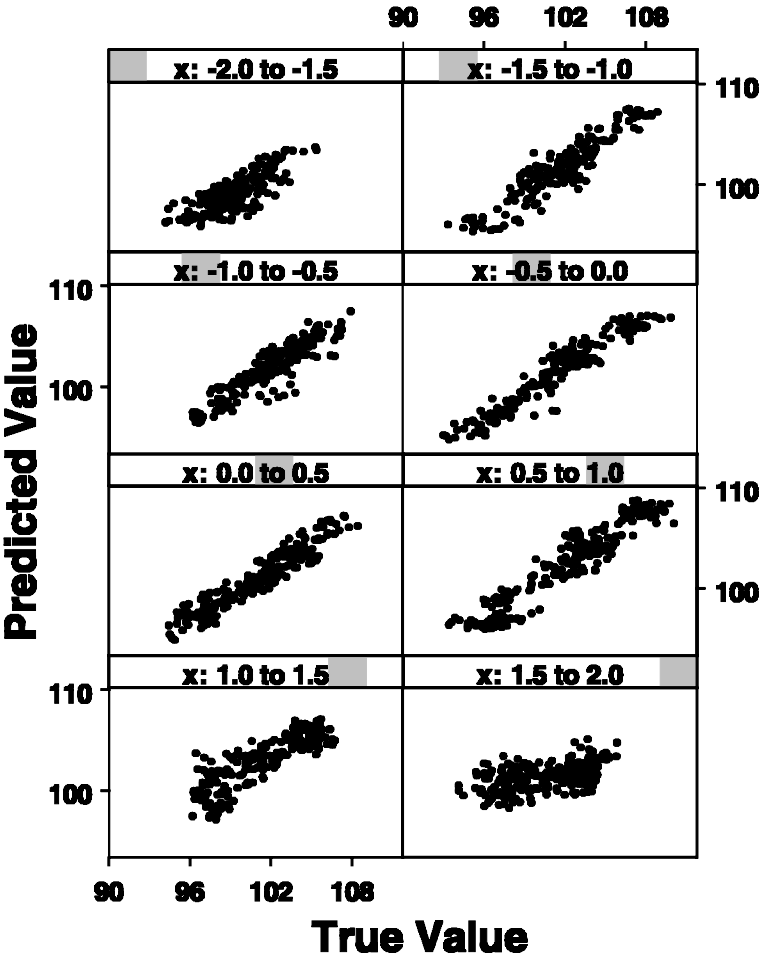


Figure 6. Validation scatterplots showing the true simulated value and the predicted value using cokriging for each band (of 0.5 width) along the x-coordinate.

predictor than kriging, but its performance does show deterioration as the primary variable becomes thinner.

All programs were developed using the two-dimensional FFT function and generic minimization functions in the software package S-Plus.

5. DISCUSSION

The moving-average construction, using the FFT, is a flexible way to model autocovariance and cross-covariance. This article shows how to use moving-average functions composed of small boxes to create valid autocovariances and cross-covariances for multivariate spatial data. The number of parameters in the moving-average functions may be constrained by requiring that the height of the boxes follow some functional form. When the boxes are shifted so that their edges coincide, then the FFT allows rapid evaluation of

Downloaded by [University of Wollongong] at 18:18 27 November 2012

the integrals to obtain the autocovariance and cross-covariance functions. For spatial lag values where the edges do not line up, a simple linear interpolator (2.14) still allows rapid evaluation of the integrals. Because the FFT allows such fast evaluation of the autocovariances and cross-covariances, numerical minimization algorithms are still highly feasible for fitting them. The advantage of the moving-average construction is that it allows the use of a large variety of autocovariances, and the resulting cross-covariances are valid by construction. By choosing appropriate moving-average functions, scientists can develop their own autocovariances and cross-covariances to meet their specific needs for environmental and other problems [e.g., Cressie and Ver Hoef (2001) applied these methods to precision agriculture data, where global positioning systems, GPS, are used to precisely locate harvest yields within fields].

The examples of Section 4 illustrate several other computational issues. First, it was shown that the data can be partitioned and restricted likelihoods summed when there are large amounts of data, and good parameter estimates were still obtained. Second, REML has an attractive automatic feature. Recall the differences in means, $\mu_1 = 100$ and $\mu_2 = 0$, for the simulated data; when fitting cross-variograms using weighted least squares, these mean differences cause problems that require standardization of the data (see Cressie 1993, p. 141; Cressie and Wikle 1998; Ver Hoef and Barry 1998). There are also issues regarding the relative weightings of weighted least squares for fitting variograms and cross-variograms (Ver Hoef and Barry 1998), and there is always the issue of binning the cross-variogram lags. The use of REML eliminated these issues. Finally, the use of cokriging neighborhoods allowed the computation of the needed inverses for cokriging.

In summary, the use of the FFT, along with REML and cokriging neighborhoods, make significant computational advances in modeling, estimation, and cokriging. We have shown that cokriging can offer significant improvements over kriging for spatial prediction, and the methods we have developed should allow more flexible models and wider use of kriging and cokriging.

Downloaded by [University of Wollongong] at 18:18 27 November 2012

ACKNOWLEDGMENTS

The authors would like to thank the editors and referees for their very helpful comments. Ver Hoef's research was supported by Federal Aid in Wildlife Restoration to the Alaska Department of Fish and Game. Cressie's research was supported by a U.S. Environmental Protection Agency STAR grant R827-257-01-0 and an Office of Naval Research grant N00014-02-1-0052.

[Received May 2001. Revised March 2003.]

REFERENCES

Barry, R. P., and Ver Hoef, J. M. (1996), "Blackbox Kriging: Spatial Prediction Without Specifying Variogram Models," *Journal of Agricultural, Biological, and Environmental Statistics*, 3, 1–25.

Brigham, E. O. (1988), *The Fast Fourier Transform and Its Applications*, Englewood Cliffs, NJ: Prentice Hall.

Cressie, N. (1993), *Statistics for Spatial Data* (rev. ed.), New York: Wiley.

Cressie, N., and Pavlicova, M. (2002), "Calibrated Spatial Moving Average Simulations," *Statistical Modelling*, 2, 1–13.

Cressie, N., and Ver Hoef, J. M. (2001), "Multivariate Geostatistics for Precision Agriculture" (with discussion), *Bulletin of the International Statistical Institute*, Invited Papers Volume 59, Book 1, 407–410.

Cressie, N., and Wikle, C. K. (1998), "The Variance-Based Cross-Variogram: You Can Add Apples and Oranges," *Mathematical Geology* 30, 789–799.

Daley, R. (1991), *Atmospheric Data Analysis*, New York: Cambridge University Press.

Gaspari, G., and Cohn, S. E. (1999), "Construction of Correlation Functions in Two and Three Dimensions," *Quarterly Journal of the Royal Meteorological Society*, 125, 723–757.

Goovaerts, P. (1997), *Geostatistics for Natural Resources Evaluation*, New York: Oxford University Press.

Gotway, C. A., and Hartford, A. H. (1996), "Geostatistical Methods for Incorporating Auxiliary Information in the Prediction of Spatial Variables," *Journal of Agricultural, Biological, and Environmental Statistics*, 1, 17–39.

Helterbrand, J. D., and Cressie, N. (1994), "Universal Cokriging Under Intrinsic Coregionalization," *Mathematical Geology*, 26, 205–226.

Higdon, D. (1998), "A Process-Convolution Approach to Modelling Temperatures in the North Atlantic Ocean," *Environmental and Ecological Statistics*, 5, 173–190.

Higdon, D., Swall, J., and Kern, J. (1999), "Non-stationary Spatial Modeling," in *Bayesian Statistics* 6, New York: Oxford University Press, pp. 761–768.

Isaaks, E. H., and Srivastava, R. M. (1989), *An Introduction to Applied Geostatistics*, New York: Oxford University Press.

Journel, A. G., and Huijbregts, C. J. (1978), *Mining Geostatistics*, London: Academic Press.

Lindstrom, T. (1993), "Fractional Brownian Fields as Integrals of White Noise," *The Bulletin of the London Mathematical Society*, 25, 83–88.

Matern, B. (1986), *Spatial Variation* (2nd ed.), New York: Springer-Verlag.

Mockus, A. (1998), "Estimating Dependencies From Spatial Averages," *Journal of Computational and Graphical Statistics* 7, 501–513.

Myers, D. E. (1991), "Pseudo Crossvariograms, Positive-Definiteness, and Cokriging," *Mathematical Geology* 23, 805–816.

Oliver, D. S. (1995), "Moving Averages for Gaussian Simulation in Two and Three Dimensions," *Mathematical Geology*, 27, 939–960.

Papritz, A., Kunsch, H. R., and Webster, R. (1993), "On the Pseudo Cross-Variogram," *Mathematical Geology*, 25, 1015–1026.

Royle, J. A. (2000), "Multivariate Spatial Models," in *Studies in the Atmospheric Sciences*, eds. M. L. Berliner, D. Nychka, and T. Hoar, New York: Springer, pp. 23–44.

Royle, J. A., and Berliner, M. (1999), "A Hierarchical Approach to Multivariate Spatial Modeling and Prediction," *Journal of Agricultural, Biological, and Environmental Statistics*, 4, 29–56.

Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, New York: Springer-Verlag.

Thiebaux, H. J., and Pedder, M. A. (1987), *Spatial Objective Analysis: with Applications in Atmospheric Science*, San Diego: Academic Press.

Ver Hoef, J. M., and Barry, R. P. (1998), "Constructing and Fitting Models for Cokriging and Multivariable Spatial Prediction," *Journal of Statistical Planning and Inference*, 69, 275–294.

Ver Hoef, J. M., and Cressie, N. (1993), "Multivariable Spatial Prediction," *Mathematical Geology*, 25: 219–240; Correction (1994), *Mathematical Geology*, 26, 273–275.

Wackernagel, H. (1998), *Multivariate Geostatistics* (2nd ed.), Berlin: Springer.

Wolpert, R., and Ickstadt, K. (1998), "Poisson/Gamma Random Field Models for Spatial Statistics," *Biometrika*, 85, 251–267.

Yaglom, A. M. (1987a), *Correlation Theory of Stationary and Related Random Functions* (vol. I), New York: Springer-Verlag.

——— (1987b), *Correlation Theory of Stationary and Related Random Functions* (vol. II.), New York: Springer-Verlag.

Yao, T., and Journel, A. G. (1998), "Automatic Modeling of (Cross) Covariance Tables Using Fast Fourier Transform," *Mathematical Geology*, 30, 589–615.

Downloaded by [University of Wollongong] at 18:18 27 November 2012