Biometrika (2011), **98**, 3, *pp.* 615–631
doi: 10.1093/biomet/asr029

Printed in Great Britain

Aggregation-cokriging for highly multivariate spatial data

BY REINHARD FURRER

Institute of Mathematics, University of Zurich, 8057 Zurich, Switzerland
reinhard.furrer@math.uzh.ch

AND MARC G. GENTON

Department of Statistics, Texas A&M University, College Station, Texas 77843, U.S.A.
genton@stat.tamu.edu

SUMMARY

Best linear unbiased prediction of spatially correlated multivariate random processes, often called cokriging in geostatistics, requires the solution of a large linear system based on the covariance and cross-covariance matrix of the observations. For many problems of practical interest, it is impossible to solve the linear system with direct methods. We propose an efficient linear unbiased predictor based on a linear aggregation of the covariables. The primary variable together with this single meta-covariable is used to perform cokriging. We discuss the optimality of the approach under different covariance structures, and use it to create reanalysis type high-resolution historical temperature fields.

Some key words: Climate; Cokriging; Eigendecomposition; Intrinsic process; Linear unbiased prediction.

1. INTRODUCTION

The prediction of a geophysical quantity based on observations at nearby locations of the same quantity and on other related variables, so-called covariables, is often of interest. Typical examples are drawing pollution maps, establishing flood plans or simply predicting temperatures. Obviously, all available observations should be taken into account, and their contributions ought to be weighted by the strength of their correlation with the location of interest. In many of those applications, we have large or even massive amounts of data available, implying the need for a careful choice of methodology in order to keep the analysis computationally feasible.

For a single variable of interest, spatial best linear unbiased prediction, i.e., kriging, has been intensively studied and many approaches exist; see Sun et al. (2011) for a recent review. However, if information from several different variables is available, this information should be used for prediction and a classical approach is cokriging. Unfortunately, the problems implied by large amounts of data are then further amplified.

We assume that we have a primary variable and two or more secondary variables. We aim at predicting the primary variable at some location based on observations of the primary variable at a set of distinct locations and on observations of the secondary variables at a possibly different set of distinct locations, i.e. a particular form of cokriging.

Cokriging was extensively studied in the late 1980s and early 1990s (e.g., Davis & Greenes, 1983; Abourfirassi & Marino, 1984; Carr & McCallister, 1985). For a more theoretical discussion see, for example, Myers (1982, 1992), Ver Hoef & Cressie (1993, 1994) and references

therein. However, since the cokriging system is of order $O(N^2)$, where N is the total number of observations, practitioners often try to reduce the number of equations by taking into account only a few neighbours of the prediction location. Myers (1983) and Long & Myers (1997) provided other solutions to reduce the computational burden of cokriging, based on linear combinations or linear approximations. Another common tool is coregionalization (e.g., Journel & Huijbregts, 1978; Gelfand et al., 2004). Each component of the process is assumed to be a linear combination of orthogonal random functions. The orthogonality is exploited with a linear transformation, reducing the multivariate setting to a univariate one. However, the transformation has to be estimated, often iteratively, cancelling out the computational gains (e.g., Bourgault & Marcotte, 1991). Finally, the extremely simple alternative solution of full cokriging in a moving neighbourhood is often used in practice, see, for example, Haas (1990a,1990b). The moving neighbourhood approach is quite useful when predicting at a limited number of locations (e.g., Johns et al., 2003) but has disadvantages, such as the potential introduction of discontinuities.

For the applications of interest to us, in numerical weather prediction, the linear systems involved in calculating the best linear unbiased predictor and its mean squared prediction error are often too large to use direct methods. We present a novel approach to reduce this computational burden, by aggregating the secondary variables with carefully chosen weights. The resulting combination of secondary variables should contribute as much as possible to the prediction of the primary variable in the mean squared prediction error sense. The prediction is then performed using a cokriging approach with the primary variable and the aggregated secondary variables. This reduces the computational burden of the prediction from solving a $(n + \ell m) \times (n + \ell m)$ to solving a $(n + m) \times (n + m)$ linear system, where n and m are the numbers of observations of the primary and secondary variables, respectively, and ℓ is the number of secondary variables. We assume n and m to be such that we are able to solve a $(n + m) \times (n + m)$ linear system but not a $(n + \ell m) \times (n + \ell m)$ one. By construction we know that the resulting mean squared prediction error lies between that of kriging and of cokriging. Since the computational complexity is comparable with bikriging, i.e. cokriging with only one of the secondary variables, we aim for a mean squared prediction error between bikriging and cokriging.

2. AGGREGATION-COKRIGING

2·1. *Definitions and basic properties*

We assume that we have a primary variable, denoted by $Y_0(\cdot)$, and two or more secondary variables, denoted by $Y_1(\cdot), \ldots, Y_\ell(\cdot)$, $\ell > 1$. We aim to predict the primary variable at a new location s_p based on observations and we write $Y_0(s_p) = Y_p$, where the subscript p stands for predict. We assume n observations from distinct locations of the primary variable, denoted by Y_0, as well as m observations from distinct locations of each of the secondary variables, denoted by $Y_1, \ldots, Y_\ell$. We use a generic notation for the indices of the secondary variable(s), e.g., Y_g might be the vector Y_1, or $(Y_1^{\mathrm{T}}, \ldots, Y_\ell^{\mathrm{T}})^{\mathrm{T}}$, or a linear combination thereof. Accordingly, the elements of $\mathrm{var}(Y_g) = \Sigma_{gg}$ are determined by the covariance structure of the secondary variables $\mathrm{cov}(Y_i, Y_j) = \Sigma_{ij}$ $(i, j = 1, \ldots, \ell)$. Further, we assume that the first moment of the multivariate random process is zero and for the second moment we write

$$\mathrm{var}\begin{pmatrix} Y_p \\ Y_0 \\ Y_g \end{pmatrix} = \begin{pmatrix} \Sigma_{pp} & \Sigma_{p0} & \Sigma_{pg} \\ \Sigma_{0p} & \Sigma_{00} & \Sigma_{0g} \\ \Sigma_{gp} & \Sigma_{g0} & \Sigma_{gg} \end{pmatrix}, \tag{1}$$

where $\Sigma_{pp} = \mathrm{var}(Y_p)$, $\Sigma_{00} = \mathrm{var}(Y_0)$, $\Sigma_{p0} = \mathrm{cov}(Y_p, Y_0)$, etc. By separating between prediction Y_p and primary variable Y_0, we can incorporate measurement errors. For example, denote a

covariance function by $c(s_i, s_j)$, then $\Sigma_{pp} = c(s_p, s_p)$ and $[\Sigma_{00}]_{ij} = c(s_i, s_j) + \tau^2 I(s_i = s_j)$ with τ^2 the magnitude of the measurement error and I the indicator function.

The best linear unbiased predictor of Y_p given Y_0 and Y_g is

$$\begin{pmatrix} \Sigma_{p0} & \Sigma_{pg} \end{pmatrix} \begin{pmatrix} \Sigma_{00} & \Sigma_{0g} \\ \Sigma_{g0} & \Sigma_{gg} \end{pmatrix}^{-1} \begin{pmatrix} Y_0 \\ Y_g \end{pmatrix}, \tag{2}$$

and its mean squared prediction error is given by

$$\Sigma_{pp} - \begin{pmatrix} \Sigma_{p0} & \Sigma_{pg} \end{pmatrix} \begin{pmatrix} \Sigma_{00} & \Sigma_{0g} \\ \Sigma_{g0} & \Sigma_{gg} \end{pmatrix}^{-1} \begin{pmatrix} \Sigma_{0p} \\ \Sigma_{gp} \end{pmatrix}. \tag{3}$$

We now introduce the aggregation-cokriging method. For each location s_i $(i = 1, \ldots, m)$, where the secondary variable is observed, we need to find an ℓ-vector $a(s_i) = \{a_1(s_i), \ldots, a_\ell(s_i)\}^{\mathrm{T}}$ which defines the weights for $Y_1(s_i), \ldots, Y_\ell(s_i)$. To simplify the exposition, we define the aggregation matrix $\mathcal{A} \in \mathbb{R}^{m \times m\ell}$, such that the corresponding linear combination of the secondary variables is simply $\mathcal{A}\mathcal{Y}$, with $\mathcal{Y} = (Y_1^{\mathrm{T}}, \ldots, Y_\ell^{\mathrm{T}})^{\mathrm{T}}$. Then, in (1)–(3), Y_g is $\mathcal{A}\mathcal{Y}$, and the corresponding covariance matrices are calculated accordingly, e.g., $\Sigma_{gg} = \mathcal{A}\,\mathrm{var}(\mathcal{Y})\mathcal{A}^{\mathrm{T}}$. The matrix Σ_{gg} is positive definite and the inverse in (2) exists if $\mathcal{A}$ has full row rank.

The aggregation matrix takes the form $\mathcal{A} = \{\mathrm{diag}(L_1), \ldots, \mathrm{diag}(L_\ell)\}$, with $L_r = \{a_r(s_1), \ldots, a_r(s_m)\}^{\mathrm{T}}$. The ith row of $\mathcal{A}$ is $a(s_i)^{\mathrm{T}} \otimes e_i^{\mathrm{T}}$ where $\otimes$ is the Kronecker product and e_i is the ith canonical basis vector. In the special case where the weights do not change with the location s_i, $\mathcal{A} = (a_1 I, \ldots, a_\ell I) = a^{\mathrm{T}} \otimes I$, where $I \in \mathbb{R}^{m \times m}$ is the identity matrix. In this notation, bikriging with the first variable corresponds to $\mathcal{A} = (I, 0, \ldots, 0) = e_1^{\mathrm{T}} \otimes I$.

To relate or link the aggregation scheme to the resulting mean squared prediction error, let $E^{-1} = \mathcal{A}(\Sigma_{\mathcal{Y}\mathcal{Y}} - \Sigma_{\mathcal{Y}0}\Sigma_{00}^{-1}\Sigma_{0\mathcal{Y}})\mathcal{A}^{\mathrm{T}}$, where $\Sigma_{\mathcal{Y}\mathcal{Y}}$ is the covariance matrix of $\mathcal{Y}$ and $\Sigma_{\mathcal{Y}0} = \Sigma_{0\mathcal{Y}}^{\mathrm{T}}$ is the cross-covariance matrix between $\mathcal{Y}$ and Y_0, and define the function

$$g(\mathcal{A}) = \begin{cases} \|\mathcal{A}(\Sigma_{\mathcal{Y}p} - \Sigma_{\mathcal{Y}0}\Sigma_{00}^{-1}\Sigma_{0p})\|_E^2 & \text{if } E \text{ is positive definite,} \\ 0 & \text{otherwise,} \end{cases}$$

where $\|z\|_E^2 = z^{\mathrm{T}} E z$ and $\Sigma_{\mathcal{Y}p}$ is the cross-covariance matrix between $\mathcal{Y}$ and Y_p. The mean squared prediction error based on the aggregation matrix $\mathcal{A}$ is

$$\mathrm{MSPE}(\mathcal{A}) = \Sigma_{pp} - \Sigma_{p0}\Sigma_{00}^{-1}\Sigma_{0p} - g(\mathcal{A}), \tag{4}$$

and minimizing (4) over all admissible $\mathcal{A}$ is equivalent to maximizing $g(\mathcal{A})$. If $\mathcal{A} = a^{\mathrm{T}} \otimes I$, we abbreviate to $g(a)$. For example, $g(0) = 0$ and $g(e_r)$ are linked to the mean squared prediction error of simple kriging and bikriging with the rth variable, respectively.

Direct maximization of $g(\mathcal{A})$ over $\mathcal{A}$ is only possible in very specific cases, and numerical maximization often requires more computational effort than solving the best linear unbiased predictor with all secondary variables. We now propose a few aggregation schemes that are intuitive and suboptimal. We choose the weight vectors as the solution of

$$a(s_i) = \underset{x}{\mathrm{argmax}}\, x^{\mathrm{T}} A_i x - \lambda(x^{\mathrm{T}} B_i x - 1), \tag{5}$$

$$a(s_i) = \underset{x}{\mathrm{argmin}}\, x^{\mathrm{T}} C_i x - \lambda(x^{\mathrm{T}} \omega - 1), \tag{6}$$

where $A_i \in \mathbb{R}^{\ell\times\ell}$ is symmetric positive semidefinite, $B_i, C_i \in \mathbb{R}^{\ell\times\ell}$ are symmetric positive definite, $\omega \in \mathbb{R}^{\ell}$ has at least one nonzero element and λ is the Lagrange multiplier. Cokriging with the aggregation matrix $\mathcal{A}$ based on (5) and on (6) will be called AGG(A_i, B_i)-cokriging and AGG(C_i, ω)-cokriging, respectively. It is straightforward to show that the solution of (5) is $a(s_i) = B_i^{-1/2}\alpha_i$, where α_i is the eigenvector associated with the largest eigenvalue of $B_i^{-1/2}A_iB_i^{-1/2}$ and the solution of (6) is $a(s_i) = C_i^{-1}\omega/(\omega^{\mathrm{T}}C_i^{-1}\omega)$. If the aggregation matrix $\mathcal{A}$ is based on the vectors $a(s_i)$ $(i = 1, \ldots, m)$, we also write AGG$\{a(s_i)a(s_i)^{\mathrm{T}}, I\}$-cokriging.

Individually rescaling the weight vectors $a(s_i)$ $(i = 1, \ldots, m)$ has no effect. Hence, we will often scale the weight vectors to $a(s_i)\{a(s_i)^{\mathrm{T}}a(s_i)\}^{-1/2}$ or $a(s_i)\{a(s_i)^{\mathrm{T}}1\}^{-1}$, provided $a(s_i)^{\mathrm{T}}1 \neq 0$, to simplify theoretical concepts.

LEMMA 1. *Prediction and mean squared prediction error of aggregation-cokriging based on the weight vectors $a(s_i)$ are independent of any scaling $\gamma_i a(s_i)$, with $\gamma_i \neq 0$ $(i = 1, \ldots, m)$.*

For cleverly chosen A_i and B_i, or C_i and ω, the weight vectors obtained with this approach are not too far from the exact optimum under certain covariance models.

2·2. *Example: canonical correlation analysis*

An intuitive choice of the weight vector $a(s_i) = \{a_1(s_i), \ldots, a_\ell(s_i)\}^{\mathrm{T}}$ is such that the correlation $\mathrm{corr}\{Y_p, a_1(s_i)Y_1(s_i) + \cdots + a_\ell(s_i)Y_\ell(s_i)\}$ is maximized. The solution of this optimization is a particular case of canonical correlation analysis, see, e.g., Mardia et al. (1979).

PROPOSITION 1. *Assume that $\mathrm{cov}\{Y_p, Y_r(s_i)\} \neq 0$ for at least one r. The vector $a(s_i) = \{a_1(s_i), \ldots, a_\ell(s_i)\}^{\mathrm{T}}$ maximizing $\mathrm{corr}\{Y_p, a_1(s_i)Y_1(s_i) + \cdots + a_\ell(s_i)Y_\ell(s_i)\}$ is $a(s_i) = \mathrm{argmax}_x\, x^{\mathrm{T}}C_{ip}C_{pi}x - \lambda(x^{\mathrm{T}}C_{ii}x - 1)$ where the matrices $C_{ii} \in \mathbb{R}^{\ell\times\ell}$ and the vectors $C_{ip} \in \mathbb{R}^{\ell}$ $(i = 1, \ldots, m)$ are defined by*

$$[C_{ii}]_{rs} = \mathrm{cov}\{Y_r(s_i), Y_s(s_i)\} = [\Sigma_{rs}]_{ii} \quad (r, s = 1, \ldots, \ell), \tag{7}$$

$$[C_{ip}]_r = [C_{pi}]_r^{\mathrm{T}} = \mathrm{cov}\{Y_r(s_i), Y_p\} = [\Sigma_{rp}]_i \quad (r = 1, \ldots, \ell). \tag{8}$$

The vector $a = a(s_i)$ depends on the locations s_i where the secondary variables are observed and on the prediction location s_p. Hence, with AGG$(C_{ip}C_{pi}, C_{ii})$-cokriging, the weights $a(s_i) \in \mathbb{R}^{\ell}$ $(i = 1, \ldots, m)$ form ℓ fields, denoted by L_r $(r = 1, \ldots, \ell)$. For simplicity, we use the term weights for these ℓ fields. By construction, these weights take into account the cross-covariance structure of the secondary variables at the locations s_i $(i = 1, \ldots, m)$ and they do not take into account the covariance structure $[\Sigma_{rr}]_{ij}$ $(i \neq j)$ of the individual variables. This means that only diagonal elements of the cross-covariance matrices enter the calculation of the weight vector $a(s_i)$, e.g. (7). This fact may lead to rapidly varying weights.

Further, with compactly supported covariance functions (see, e.g. Gneiting, 2002) the cross-covariance $\mathrm{cov}\{Y(s_i), Y_p\}$ might be zero, resulting in a zero vector C_{ip}. In this case, the matrix $C_{ii}^{-1/2}C_{ip}C_{pi}C_{ii}^{-1/2}$ has only zero eigenvalues and the question arises what weight vector $a(s_i)$ we should choose. Intuitive choices are: the rth canonical basis vector, where r corresponds to the variable also used for bikriging; and $\ell^{-1/2}1$, i.e., all variables receive the same weight. Later we will see that the second case can be competitive but often inferior to the first case, for which we need to determine the variable r, implying additional computational costs.

2·3. *Example: maximum covariance analysis*

The canonical correlation based approach has the disadvantage that we need to calculate m vectors $a(s_i)$ $(i = 1, \ldots, m)$. Instead of simplistically choosing an arbitrary location s_j and using $a(s_i) = a(s_j)$ for all i, we now discuss a more intuitive procedure for which the weights do not depend on the location s_i. We propose to maximize the ordinary L_2-norm of

$$\operatorname{cov}\left(Y_p, \sum_{r=1}^{\ell} a_r Y_r\right) = \sum_{r=1}^{\ell} a_r \Sigma_{pr}, \tag{9}$$

which is

$$\operatorname{tr}\left(\sum_{r,s=1}^{\ell} a_r a_s \Sigma_{rp} \Sigma_{ps}\right) = \sum_{r,s=1}^{\ell} a_r a_s \operatorname{tr}(\Sigma_{rp} \Sigma_{ps}) = a^{\mathrm{T}} [\Sigma_{ps} \Sigma_{rp}] a,$$

under the constraint $a^{\mathrm{T}} a = 1$. Straightforward algebraic manipulations imply that a is the eigenvector associated with the largest eigenvalue of the matrix with elements $[\Sigma_{ps} \Sigma_{rp}]_{rs}$. Hence, we perform $\text{AGG}(\Sigma_{ps} \Sigma_{rp}, I)$-cokriging, where $I \in \mathbb{R}^{\ell \times \ell}$. This approach is a particular case of maximum covariance analysis, see, e.g., Jolliffe (2002).

We could also define the weight vectors $a(s_i)$ of § 2·2 via a covariance instead of a correlation approach. Or we could constrain the vector a of this section to $a^{\mathrm{T}} B a = 1$, for some symmetric positive definite matrix B. These cases would result in $\text{AGG}(C_{ip} C_{pi}, I)$- and $\text{AGG}(\Sigma_{ps} \Sigma_{rp}, B)$-cokriging.

2·4. *Generalized intrinsic model for the secondary variables*

If we consider constant weights induced by $a = (a_1, \ldots, a_\ell)^{\mathrm{T}}$, then $g(a)$ takes the form

$$g(a) = \left(\sum_{r=1}^{\ell} a_r c_r\right)^{\mathrm{T}} \left(\sum_{r,s=1}^{\ell} a_r a_s S_{rs}\right)^{-1} \left(\sum_{r=1}^{\ell} a_r c_r\right), \tag{10}$$

with $c_r = \Sigma_{rp} - \Sigma_{r0} \Sigma_{00}^{-1} \Sigma_{0p}$ and $S_{rs} = \Sigma_{rs} - \Sigma_{r0} \Sigma_{00}^{-1} \Sigma_{0s}$. Although the previous expression has well-known derivatives with respect to a_i, it is only possible in very specific cases to analytically find the maximum of $g(a)$. One example of such a simple model is based on the following specification of the covariance matrix:

$$\operatorname{var}\begin{pmatrix} Y_p \\ Y_0 \\ \mathcal{Y} \end{pmatrix} = \begin{pmatrix} \Sigma_{pp} & \Sigma_{p0} & \omega^{\mathrm{T}} \otimes \Sigma_{pc} \\ \Sigma_{0p} & \Sigma_{00} & \omega^{\mathrm{T}} \otimes \Sigma_{0c} \\ \omega \otimes \Sigma_{cp} & \omega \otimes \Sigma_{c0} & \Omega \otimes \Sigma_{cc} \end{pmatrix}, \tag{11}$$

with $\Omega \in \mathbb{R}^{\ell \times \ell}$ a symmetric positive definite matrix and $\omega \in \mathbb{R}^{\ell}$ a vector with at least one nonzero element. The matrix $\Sigma_{cc} \in \mathbb{R}^{m \times m}$ is the common correlation structure of the secondary variables. Further, up to a constant, $\Sigma_{pc} \in \mathbb{R}^{1 \times m}$ is the cross-covariance between Y_p and any secondary variable and $\Sigma_{0c} \in \mathbb{R}^{n \times m}$ is the cross-covariance between the primary variable and any secondary variable. Without explicitly stating all conditions, we assume that the matrix (11) is positive definite.

Model (11) is more general than the standard intrinsic correlation model, e.g., Wackernagel (2006, p. 154). In (11), only the secondary variables form an intrinsic correlation model.

PROPOSITION 2. *Under the model specified by* (11), AGG(Ω, ω)*-cokriging minimizes the mean squared prediction error* (3).

In practice, we often have $\mathrm{var}(\mathcal{Y}) \neq \Omega \otimes \Sigma_{cc}$ as, for example, the secondary variables may have different correlation ranges. A simple approximation of the variance in (11) is illustrated in § 4. In the case of stationarity, we could use as a first-order approximation $[\Omega_{rs}] = [\Sigma_{rs}]_{ii}$ and $\omega = [\Sigma_{r0}]_i$, for some i. A more formal approximation approach is to find the closest intrinsic approximation to the covariance structure (1), in a similar spirit to Genton (2007).

2·5. *Identical cross-covariance structure*

We work through a simple case, assuming that all the secondary variables have the same cross-covariance structure with Y_p and with the observations of the primary variable.

LEMMA 2. *Assume that* $\Sigma_{rp} = \Sigma_{1p}$ *and* $\Sigma_{r0} = \Sigma_{10}$ $(r = 2, \ldots, \ell)$. *Let* $\mathcal{A}$ *and* $\mathcal{B}$ *be aggregation matrices based on aggregation vectors* $a(s_i)$ *and* $b(s_i)$ *with* $a(s_i)^{\mathrm{T}} 1 \neq 0$ *and* $b(s_i)^{\mathrm{T}} 1 \neq 0$ $(i = 1, \ldots, m)$. *Then,* MSPE$(\mathcal{A}) <$ MSPE$(\mathcal{B})$ *if and only if the matrix*

$$\sum_{r=1}^{\ell} \sum_{s=1}^{\ell} \Sigma_{rs} \circ (H_{rs} - G_{rs}) \tag{12}$$

is positive definite, where $\circ$ *is the Schur or direct product,* $H_{rs} = \{b_r(s_1), \ldots, b_r(s_m)\}^{\mathrm{T}} \{b_s(s_1), \ldots, b_s(s_m)\}$ *and* $G_{rs} = \{a_r(s_1), \ldots, a_r(s_m)\}^{\mathrm{T}} \{a_s(s_1), \ldots, a_s(s_m)\}$.

The proof, given in the Appendix, is based on the fact that $\Sigma_{r0} \Sigma_{00}^{-1} \Sigma_{0s}$ is independent of r and s, due to the identical covariance structure. The condition based on (12) is often not practical, but $H_{rs} - G_{rs}$ being positive semidefinite for all r and s also implies positive definiteness of (12) (Horn & Johnson, 1994, Theorem 5·2·1). Further, (12) can be simplified considerably if the secondary variables all have the same cross-covariance $\Sigma_{rs} = \Sigma_{12}$ $(r \neq s = 1, \ldots, \ell)$.

In the case of constant weights, G_{rs} and H_{rs} are proportional to 11^{T}.

LEMMA 3. *Assume that* $\Sigma_{rp} = \Sigma_{1p}$ *and* $\Sigma_{r0} = \Sigma_{10}$ $(r = 2, \ldots, \ell)$. *If the aggregation matrices* $\mathcal{A}$ *and* $\mathcal{B}$ *are based on constant weights given by a and b, then* MSPE$(\mathcal{A}) <$ MSPE$(\mathcal{B})$:

(i) *if and only if* $\sum_{r=1}^{\ell} \sum_{s=1}^{\ell} \Sigma_{rs} \circ 11^{\mathrm{T}} \{b_r b_s (b^{\mathrm{T}} 1)^{-2} - a_r a_s (a^{\mathrm{T}} 1)^{-2}\}$ *is positive definite*; *or*
(ii) *if* $b_r b_s (a^{\mathrm{T}} 1)^2 \geqslant a_r a_s (b^{\mathrm{T}} 1)^2$ *for all* $r, s = 1, \ldots, \ell$, *with strict inequality for at least one pair.*

Additionally, if $\Sigma_{rs} = \Sigma_{12}$ $(r \neq s = 1, \ldots, \ell)$, *then* MSPE$(\mathcal{A}) <$ MSPE$(\mathcal{B})$ *if and only if* $\sum_{r=1}^{\ell} K_r \circ 11^{\mathrm{T}} \{b_r^2 (b^{\mathrm{T}} 1)^{-2} - a_r^2 (a^{\mathrm{T}} 1)^{-2}\}$ *with* $K_r = \Sigma_{rr} - \Sigma_{12}$, *is positive definite. Lastly, if* $\Sigma_{rr} = \Sigma_{11}$ *and* $\Sigma_{rs} = \Sigma_{12}$ $(r \neq s = 1, \ldots, \ell)$, *then* MSPE$(\mathcal{A}) <$ MSPE$(\mathcal{B})$ *if and only if the weight vectors satisfy* $b^{\mathrm{T}} b (b^{\mathrm{T}} 1)^{-2} > a^{\mathrm{T}} a (a^{\mathrm{T}} 1)^{-2}$.

COROLLARY 1. *Assume that* $\Sigma_{rp} = \Sigma_{1p}$, $\Sigma_{r0} = \Sigma_{10}$, $\Sigma_{rr} = \Sigma_{11}$ *and* $\Sigma_{rs} = \Sigma_{12}$ $(r \neq s = 1, \ldots, \ell)$. *If we use the first variable for bikriging, then*:

(i) *if* $a^{\mathrm{T}} 1 = 0$, *then* $g(a) = g(0) = 0$;
(ii) *if* $|a^{\mathrm{T}} 1| = (a^{\mathrm{T}} a)^{1/2}$, *then* $g(a) = g(e_1)$. *Further, the equality signs can be replaced with strict inequalities.*

Using (4), *the statements also relate the mean squared prediction errors.*

PROPOSITION 3. *Assume that* $\Sigma_{rp} = \Sigma_{1p}$, $\Sigma_{r0} = \Sigma_{10}$, $\Sigma_{rr} = \Sigma_{11}$ *and* $\Sigma_{rs} = \Sigma_{12}$ $(r \neq s = 1, \ldots, \ell)$. *The vector a defined in Proposition 1 and the vector a of* AGG(Ω, ω)*-cokriging are proportional to* 1 *and the mean squared prediction errors of aggregation-cokriging and cokriging are identical.*

The last statement supports using 1 as the weight vector whenever $C_{ip} = 0$ as discussed in § 2·2.

If the secondary variables do not have a common cross-covariance structure, then we cannot derive theoretical results about the performance of aggregation-cokriging. Simulations indicate, however, that aggregation-cokriging is often still very competitive.

2·6. *Computational complexity*

In order to compare the computational complexity of the methods discussed above, we assume that dot and matrix products and the calculation of inverses are performed in $O(n)$, $O(n^2)$ and $O(n^3)$ computing time, respectively, although we acknowledge the existence of iterative methods which are of lower order, e.g., Billings et al. (2002). Calculating an inverse Σ^{-1} can often be avoided by solving the associated linear system instead, which is usually done with a Cholesky decomposition of Σ, generally of order $O(n^3)$ and two triangular solves, generally of order $O(n^2)$. Assume that we know the covariance structure and its parameters. Then for prediction with zero mean, we need to solve a linear system, perform a matrix-vector and a vector-vector multiplication, see (2). For kriging, bikriging and cokriging, the associated size is n, $n + m$ and $n + \ell m$, respectively. Aggregation-cokriging is on top of calculating the weight vectors, as complex as bikriging. Calculating the weight vectors for AGG$(C_{ip}C_{pi}, C_{ii})$-cokriging and AGG$(\Sigma_{ps}\Sigma_{rp}, I)$-cokriging is of order $O(m\ell^3)$ and $O(\ell^3 + \ell m)$, negligible compared with the prediction step. If the secondary variables are second-order stationary, C_{ii} is independent of i, and we only need to calculate $C_{ii}^{-1/2}$ once.

In practice, the first two moments must often be estimated. Assuming that the mean is a linear combination of k known basis functions besides the vector-vector products, the computational complexity increases from solving one to solving $k + 1$ linear systems. As the Cholesky decomposition dominates the calculation, the order of complexity of the computation does not change.

An unknown second moment structure implies a much heavier computational burden, often requiring $O(n^2)$ or even $O(n^3)$ computing time. Hence, another advantage of AGG$(C_{ip}C_{pi}, C_{ii})$-cokriging is that it requires only the cross-covariances $\mathrm{cov}\{Y_r(s_i), Y_s(s_i)\}$ at identical locations but not the cross-covariances $\mathrm{cov}\{Y_r(s_i), Y_s(s_j)\}$ $(i \neq j)$ of the secondary variables. After the aggregation, Σ_{gg} is estimated directly from the meta-covariable $\mathcal{AY}$.

3. NUMERICAL COMPARISON

We illustrate variants of the proposed method with simple cases and contrast their mean squared prediction error with bikriging and cokriging. Throughout this section, we assume Gaussian processes with a primary variable observed at $n = 8^2$ equispaced locations and three secondary variables $(\ell = 3)$ observed at $m = 14^2$ equispaced locations in the unit square $[0, 1]^2$. We perform prediction along two paths situated near the edge and at the centre of the domain, see Fig. 1. We assume a zero first moment and that the cross-covariances are given by isotropic stationary spherical covariance functions

$$c(h; \theta_1, \theta_2) = \theta_1 \max(0, 1 - h^2\theta_2^{-2})\{1 + h(2\theta_2)^{-1}\}, \quad \theta_1, \theta_2 > 0. \tag{13}$$

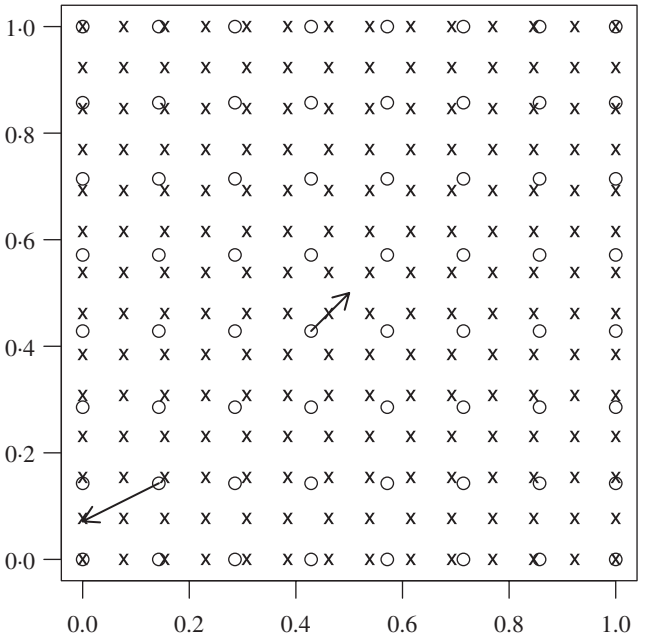

Fig. 1. Layout of the locations for the primary variables (circles) and the secondary variables (crosses). The two arrows indicate the paths along which we predicted.

Table 1. *Parameter values of the spherical covariance matrix for the four different examples*

Example	1	2	3	4
Sill θ_1	$\begin{pmatrix} 1 & 0{\cdot}3 & 0{\cdot}25 & 0{\cdot}2 \\ & 1 & 0{\cdot}2 & 0{\cdot}3 \\ & & 1 & 0{\cdot}2 \\ & & & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0{\cdot}3 & 0{\cdot}3 & 0{\cdot}1 \\ & 1 & 0{\cdot}3 & 0{\cdot}2 \\ & & 1 & 0{\cdot}3 \\ & & & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0{\cdot}45 & 0{\cdot}3 & 0{\cdot}1 \\ & 1 & 0{\cdot}3 & 0{\cdot}2 \\ & & 1 & 0{\cdot}3 \\ & & & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0{\cdot}6 & 0{\cdot}1 & 0{\cdot}1 \\ & 1 & 0{\cdot}1 & 0{\cdot}1 \\ & & 1 & 0{\cdot}1 \\ & & & 1 \end{pmatrix}$
Range θ_2	$\begin{pmatrix} 1{\cdot}2 & 0{\cdot}7 & 0{\cdot}7 & 0{\cdot}7 \\ & 0{\cdot}6 & 0{\cdot}6 & 0{\cdot}6 \\ & & 0{\cdot}6 & 0{\cdot}6 \\ & & & 0{\cdot}6 \end{pmatrix}$	$\begin{pmatrix} 1 & 0{\cdot}6 & 0{\cdot}5 & 0{\cdot}4 \\ & 0{\cdot}5 & 0{\cdot}3 & 0{\cdot}2 \\ & & 0{\cdot}2 & 0{\cdot}1 \\ & & & 0{\cdot}1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0{\cdot}6 & 0{\cdot}5 & 0{\cdot}4 \\ & 0{\cdot}5 & 0{\cdot}3 & 0{\cdot}2 \\ & & 0{\cdot}2 & 0{\cdot}1 \\ & & & 0{\cdot}1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0{\cdot}6 & 0{\cdot}1 & 0{\cdot}1 \\ & 0{\cdot}5 & 0{\cdot}1 & 0{\cdot}1 \\ & & 0{\cdot}1 & 0{\cdot}1 \\ & & & 0{\cdot}1 \end{pmatrix}$

The components match the parameters of the corresponding primary and secondary variables, e.g. the elements $[\cdot]_{12}$ contain the parameters of the cross-covariance between the primary and the first secondary variable.

We call θ_1 the sill and θ_2 the range. To simplify the notation, we assume that Y_p and Y_0 have the same second moment structure, i.e., $\tau^2 = 0$. In what follows, we discuss four examples of different cross-covariance parameter combinations all resulting in positive definite covariance matrices. The parameter values of the spherical covariances are given in Table 1. Those of Example 1 are chosen so that it represents an intrinsic model (11). Example 2 does not represent a specific covariance structure but the parameters are chosen to ensure positive definiteness. Example 3 is almost identical to Example 2 except for one different sill value, to explore the sensitivity of small changes. The last example is such that there is only considerable correlation between the primary and the first secondary variable. The secondary variables of all examples are ordered such that bikriging with the first and the third variable has the smallest and the largest mean squared prediction error, respectively.

We compare $\text{AGG}(C_{ip}C_{pi}, C_{ii})$-, $\text{AGG}(C_{ip}C_{pi}, I)$-, $\text{AGG}(\Sigma_{pr}\Sigma_{sp}, I)$-, $\text{AGG}(\Omega, \omega)$- and, $\text{AGG}(11^{\text{T}}, I)$-cokriging, i.e., aggregation-cokriging based on the vector 1. Only Example 1 is of

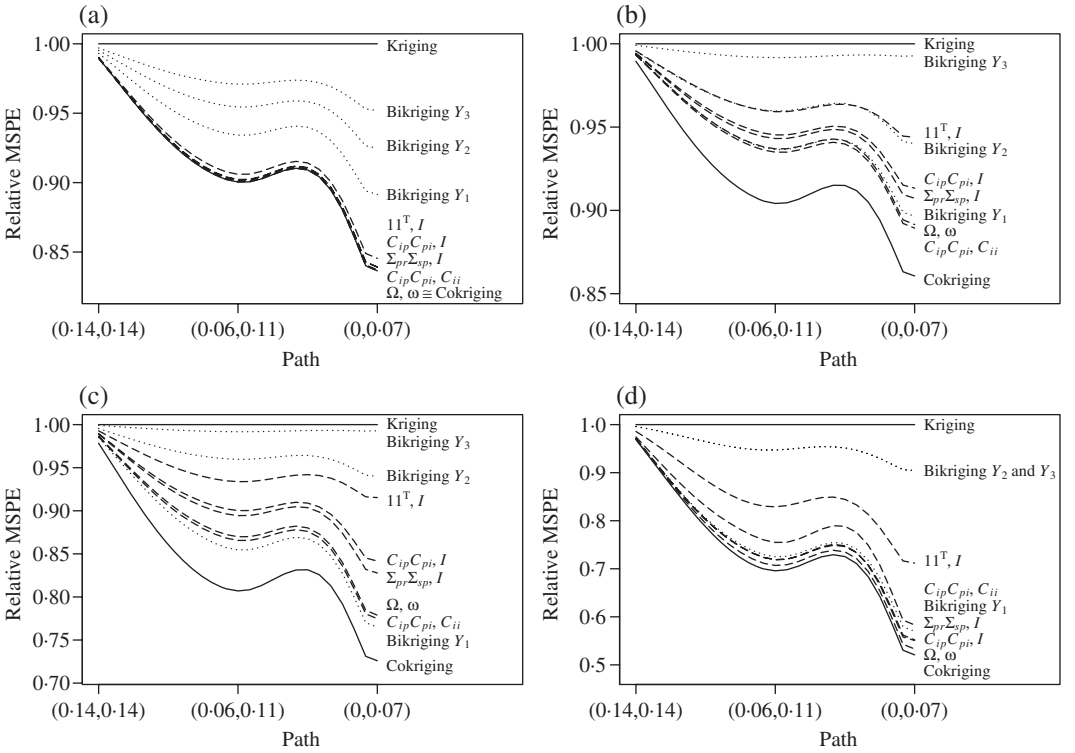

Fig. 2. Relative mean squared prediction error along the path in the lower left corner of Fig. 1 for the four examples with parameters in Table 1. Bikriging and aggregation-cokriging approaches are dotted and dashed, respectively. AGG(A_i, B_i)-cokriging is denoted by 'A_i, B_i'. The labels are ordered according to the relative mean squared prediction error at the endpoint.

the form (11) and we approximate the covariance structure of Examples 2 to 4 with an intrinsic structure $\mathrm{var}(\mathcal{Y}) \sim \Omega \otimes \Sigma_{cc}$, where Ω contains the respective sills and Σ_{cc} is unspecified.

Figure 2 gives the relative mean squared prediction error along the path in the lower left corner of Fig. 1. The mean squared prediction errors are bounded by the kriging and the cokriging curves. The local minima at (0·06,0·11) are due to the close observation of the secondary variable. Except in Example 3, several aggregation-cokriging schemes outperform bikriging. Among the aggregation-cokriging schemes, AGG($C_{ip}C_{pi}$,C_{ii})- and AGG(Ω, ω)-cokriging have in seven out of eight cases the smallest and second smallest relative mean squared prediction error.

The figure for the relative mean squared prediction error along the path at the centre of Fig. 1 is very similar and Table 2 gives the relative mean squared prediction error for its end point (0·5,0·5). Because more primary observations are available at the centre, it is much harder to compete with bikriging with the first variable, which has the lowest relative mean squared prediction error in two cases. In three cases, AGG($C_{ip}C_{pi}$,C_{ii})-cokriging has the second smallest relative mean squared prediction error and AGG(Ω, ω)-cokriging has the third smallest relative mean squared prediction error in Examples 2 to 4.

Although still competitive, AGG($C_{ip}C_{pi}$, I)-cokriging and AGG($\Sigma_{pr}\Sigma_{sp}$, I)-cokriging often have a slightly higher relative mean squared prediction error than their computationally equivalent counterparts AGG($C_{ip}C_{pi}$, C_{ii})-cokriging and AGG(Ω, ω)-cokriging, respectively. Of both approaches, AGG($\Sigma_{pr}\Sigma_{sp}$, I)-cokriging performs slightly better in general.

Introducing a nugget effect or scaling all ranges does not change the overall picture. The system is however often sensitive to a change of the sill parameters. As illustrated by Examples 2

Table 2. *Relative mean squared prediction error for the prediction at* (0·5, 0·5)

Example	Bikriging with			Aggregation-cokriging with arguments					Cokriging
	Y_1	Y_2	Y_3	$C_{ip}C_{pi}, C_{ii}$	$C_{ip}C_{pi}, I$	$\Sigma_{pr}\Sigma_{sp}, I$	Ω, ω	$11^{\mathrm{T}}, I$	
1	0·961	0·973	0·983	0·941**	0·942	0·942	0·941*	0·944	0·941
2	0·959*	0·972	0·993	0·963**	0·967	0·966	0·965	0·971	0·920
3	0·900*	0·972	0·993	0·916**	0·937	0·933	0·923	0·947	0·809
4	0·775**	0·987	0·987	0·811	0·798	0·775*	0·777	0·925	0·770

*Smallest value among bikriging and aggregation-cokriging. **Second smallest value among value among bikriging and aggregation-cokriging.

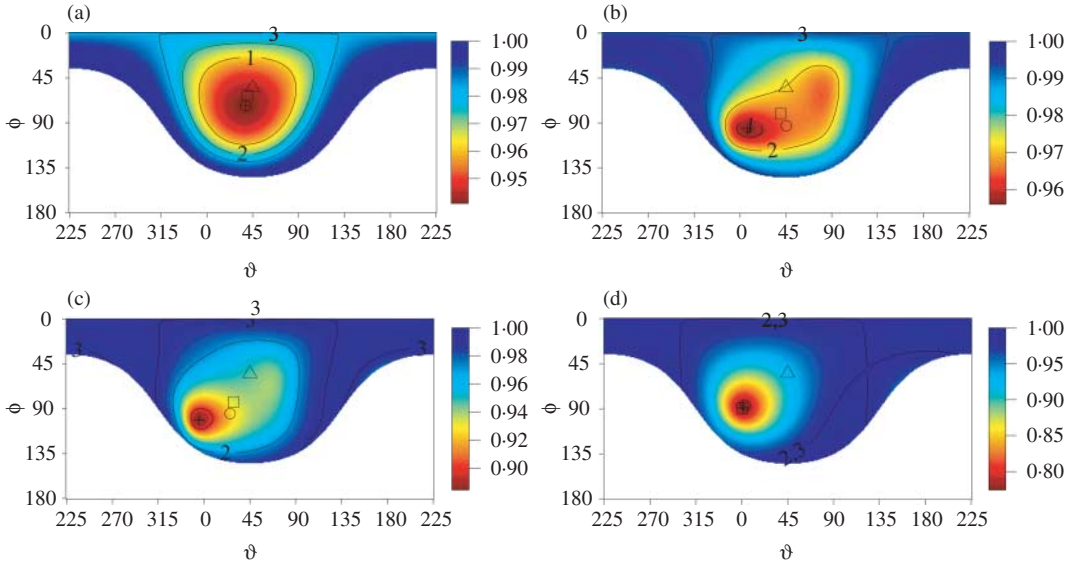

Fig. 3. Relative mean squared prediction error based on cokriging with an aggregation vector $a^{\mathrm{T}}1 \geqslant 0$ defined by the spherical angles (ϑ, ϕ) for prediction at location (0·5,0·5). The contour lines are the relative mean squared prediction error for bikriging with the three secondary variables $a = e_i$ $(i = 1, 2, 3)$, and hence passing through the spherical angles (0, 90), (90, 90) and (0, 0). The numerical minimum is denoted with +. The aggregation vectors for AGG$(\Sigma_{pr}\Sigma_{sp}, I)$-, AGG$(\Omega, \omega)$- and AGG$(11^{\mathrm{T}}, I)$-cokriging are indicated with □, ○ and △, respectively.

and 3, a change in the sill of the cross-covariance between the primary and the first secondary variable from 0·3 to 0·45 not only changes the relative mean squared prediction error dramatically, but even the corresponding weights.

The examples illustrate that there is unfortunately no general rule to choose the aggregation scheme. We advocate using AGG(Ω, ω)-cokriging, especially if the covariance structure is close to an intrinsic model (11). Otherwise, AGG$(C_{ip}C_{pi}, C_{ii})$-cokriging should be favoured or AGG$(\Sigma_{pr}\Sigma_{sp}, I)$-cokriging if computation time prohibits the former, in all cases with plug-in estimates where necessary.

For a low-dimensional setting as given here, we can numerically find the aggregation vector a by directly maximizing $g(a)$, i.e. minimizing the mean squared prediction error. As an illustration, we perform a grid search over all vectors a, $a^{\mathrm{T}}a = 1$, when predicting at (0·5,0·5). We use spherical coordinates and represent all vectors a by the angles ϑ and ϕ. Figure 3 shows relative mean squared prediction error as a function of a for $a^{\mathrm{T}}1 \geqslant 0$. The plane $a^{\mathrm{T}}1 = 0$ divides the sphere in two half spheres which are symmetric with respect to the mean squared prediction error. This explains the bowl shape of the area in Fig. 3. The contour lines give the relative mean

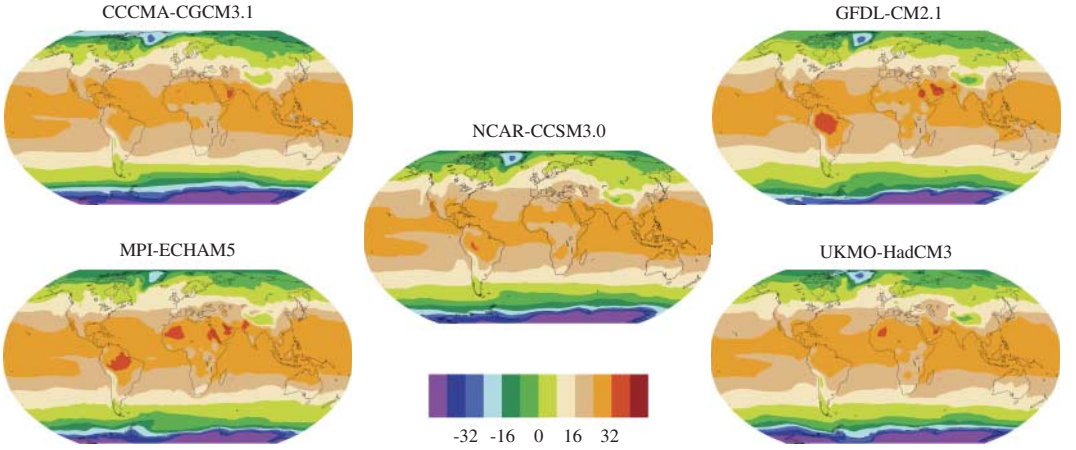

Fig. 4. September 1957 average temperature model output from five general circulation models in degrees Celsius. The data are interpolated to a 5 × 5 degree latitude-longitude resolution.

squared prediction error for the bikriging approaches with the three secondary variables, corresponding to $a = e_i$ $(i = 1, 2, 3)$ and the aggregation vectors for $\text{AGG}(\Sigma_{pr}\Sigma_{sp}, I)$-, $\text{AGG}(\Omega, \omega)$- and $\text{AGG}(11^{\text{T}}, I)$-cokriging are indicated as well. In Example 1, $\text{AGG}(\Omega, \omega)$-cokriging and the numerically found minimum mean squared prediction error coincides with the aggregation vector $(0{\cdot}754,0{\cdot}589,0{\cdot}292)^{\text{T}}$, which is (38°,73°) in spherical coordinates. For Example 4 relatively few angles lead to a smaller mean squared prediction error compared with bikriging with the first variable, in other words, the contour encloses a very small area.

4. Imputation of observed temperature fields

We illustrate the proposed methodology with an example from numerical weather prediction. A major contribution of the National Center for Environmental Prediction or the European Centre for Medium-Range Weather Forecasts is the reanalysis of temperature or precipitation fields. A reanalysis consists of blending sparse past weather observations with a numerical model and deriving best guess fields, see Kalnay et al. (1996). The construction of a reanalysis is a very time consuming computing process. Here, our goal is to illustrate our methodology to supply a similar product using much less computing time.

In order to provide a reanalysis type field, we consider temperature fields from general circulation model data. Within the context of the Fourth Assessment Report of the Intergovernmental Panel for Climate Change, several centres provided publicly available detailed climate model output for the last century (Meehl et al., 2007). For the analysis here, these fields are interpolated to a common 5 × 5 degree resolution ranging from −85 to 85 degrees of latitude ($m = 2448$). Figure 4 shows the five temperature fields considered for September 1957.

It is important to note that the climate model temperature fields represent one possible realization of a September 1957 temperature field and it is not adequate to directly compare these with observations at the same time. The differences between climate model fields and observations are representative of the natural climate variability, as well as of the differences between different climate models. An optimal reanalysis for a specific month should be based on climate model fields exhibiting similar climatological modes, e.g., similar indices for the North Atlantic Oscillation, Southern Oscillation Index. On the other hand, the commonly used reanalysis fields

Table 3. *Weighted residual sum of squares of the fit and relative, to kriging, root average squared prediction error of the prediction with each bikriging and with* AGG($\hat{\Omega}$, $\hat{\omega}$)*-cokriging. The aggregation weights of the secondary variables are given in the last row.*

	Variable				AGG($\hat{\Omega}$, $\hat{\omega}$)-cokriging
	Y_1	Y_2	Y_3	Y_4	
WRSS	1·766	1·339	1·244	1·363	18·978
Relative RASPE	0·970	0·988	1·044	0·977	0·477
Weights	0·730	0·789	0·763	0·631	

WRSS, weighted residual sum of squares; RASPE, root average squared prediction error.

differ at many grid points by more than 2°C from observations in mid latitudes (Hertzog et al., 2006), and by more than 4°C over the Antarctic sea ice among themselves (Ma et al., 2008). These differences are comparable to the differences between the climate model fields; see Fig. 4. Therefore, we avoid a lengthy introduction of real or true gridded temperature observations by using the output of NCAR-CCSM3.0 as the truth. In order to nevertheless mimic a reanalysis to at least a certain extent, we reduce this true temperature field, i.e., our primary variable, to the same locations as in the observational field for September 1957 described in Jones et al. (1999) and Brohan et al. (2006), leading to $n = 1507$. The remaining $\ell = 4$ temperature fields are used as secondary variables.

We assume a known common mean structure consisting of the point-wise mean of the temperature fields. The centred temperature variable for $i = 0, 1, \ldots, \ell = 4$ at location $s = (\delta, \vartheta)$ with δ and ϑ the latitude and longitude, is assumed to be second order stationary with spherical covariance function (13). The cross-covariances C_{rs} between different temperature variables are given by a spherical covariance function, where we relax the condition on the sill parameter to $\theta_1 \in \mathbb{R}$. In general, there is no closed form characterization of the parameters for such a multivariate process. A necessary condition on the cross-covariances though is that $f_{rs}(\omega)^2 \leqslant f_{rr}(\omega) f_{ss}(\omega)$ (e.g., Wackernagel, 2006, p. 152), where f_{rs} is the spectral density, i.e., the Fourier transform, of the covariance function C_{rs}. The spectral density of the spherical covariance function is proportional to $\{1 - \cos(\theta_2\omega)\}^2\{1 - \sin(\theta_2\omega)\}^2(\theta_2\omega)^{-4}$ and implies in practice that the range parameter has to be the same for all covariance and cross-covariance functions.

Matheron's classical variogram estimator (Matheron, 1962) is used to estimate the variograms and cross-variograms. To fit the parameters, we bin the empirical variograms according to a series of lags and then use weighted least squares, where the weights depend on the number of pairs in each bin (e.g., Cressie, 1993, § 2.4). For bikriging and AGG($\hat{\Omega}$, $\hat{\omega}$)-cokriging, a common range parameter was fitted. Table 3 gives the weighted residual sum of squares of the fit, serving as a crude measure of goodness-of-fit. The weighted residual sum of squares of aggregation-cokriging is larger because all cross-covariances among the primary and secondary variables and among the secondary variables need to be fitted as well.

Each secondary variable has $m = 2448$ observations, so straightforward implementations of bikriging are computationally challenging and cokriging is infeasible. Since the spherical covariance structure implies sparse matrices, the use of sparse matrix algebra allows the calculations to be carried out on ordinary desktop computers (Furrer et al., 2006; Furrer & Sain, 2010). With a Centrino powered GNU/Linux laptop estimation/fitting/prediction take approximately 1·57/0·00/0·28, 2·93/0·03/1·78 and 9·42/0·66/1·75 s for kriging, bikriging and aggregation-cokriging, respectively. However, depending on the parameters passed to the optimization functions, the fitting times may increase by one order of magnitude. Without the use of sparse matrix algebra, simple kriging and bikriging steps take 2·4 and 23·4 s, respectively.

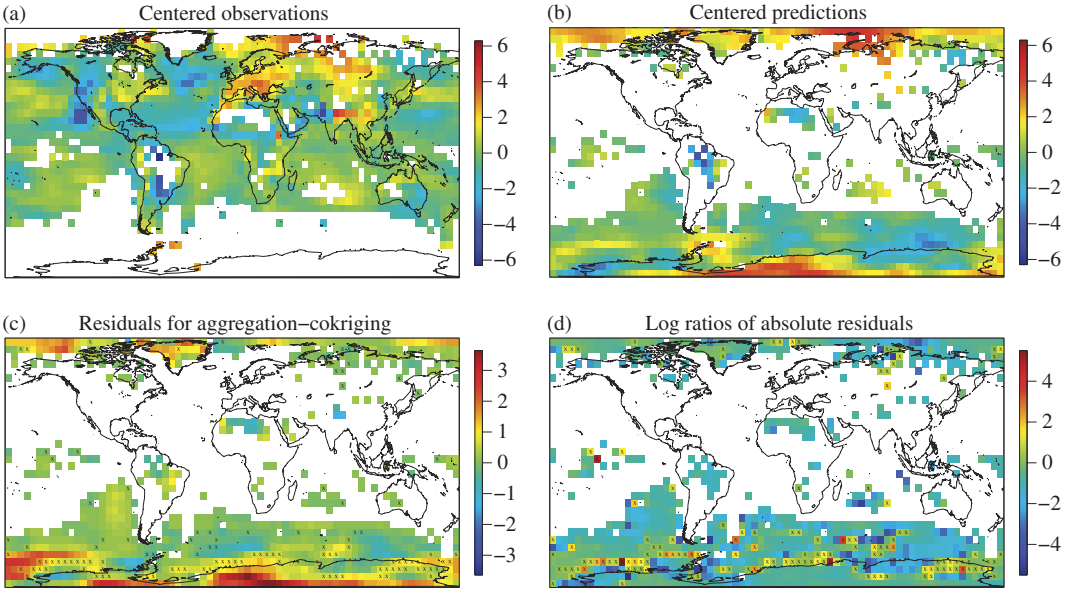

Fig. 5. Centred observed (a) and predicted (b) monthly average temperatures for September 1957 using AGG($\widehat{\Omega}$, $\widehat{\omega}$)-cokriging. (c) Prediction errors. The marks indicate the locations for which the aggregation-cokriging residuals are larger in absolute value compared with the simple kriging ones. (d) Log ratios of the absolute errors of AGG($\widehat{\Omega}$, $\widehat{\omega}$)-cokriging and of bikriging with the third variable. The marks indicate the locations for which the aggregation-cokriging residuals are larger in absolute value compared with the bikriging ones.

Figures 5(a) and 5(b) show the centred observed and predicted monthly average temperatures for September 1957 using AGG($\hat{\Omega}$, $\hat{\omega}$)-cokriging. Figures 5(c) and 5(d) show the prediction errors, i.e., the difference between the predicted monthly average temperatures for September 1957 using AGG($\hat{\Omega}$, $\hat{\omega}$)-cokriging and the actually observed temperatures. Aggregation-cokriging has a relative root average squared prediction error of 0·48. For a simple kriging approach, prediction in high and low latitudes relies only on observations that are far away. With the aggregation-cokriging approach, observations from the secondary variables are available and improve the prediction. The relative root average squared prediction error of the bikriging approaches range from 1·04 to 0·97; see also Table 3. There is no structure in the locations where bikriging with the first variable performs better than aggregation-cokriging. For this application, the aggregation-cokriging performs well because the weighted average smooths the secondary variables close to the poles, which can be interpreted as a regression towards the mean effect.

5. Extensions

The methodology was applied to gridded temperature anomaly fields from general circulation models to construct a best guess reanalysis temperature field. The approach can be applied to other multivariate settings, where we measure different variables on common locations to predict a primary variable. Examples include air pollution data like carbon monoxide and nitrogen oxides to predict tropospheric ozone production (Schmidt & Gelfand, 2003; Majumdar & Gelfand, 2007; Apanasovich & Genton, 2010) and gridded wind speed and wind direction model data to predict sea surface temperature (Berliner et al., 2000). If $n + m$ is too large to solve the associated linear system, it is possible to first apply aggregation-cokriging to

address the large ℓ problem, and then, for example, employ reduced rank representations of the spatial process (e.g. Banerjee et al., 2008; Cressie & Johannesson, 2008) for the bikriging step.

We have assumed that the secondary variables are observed at all the distinct m locations. If this is not the case, we cannot derive the aggregation matrix $\mathcal{A}$ as discussed in § 2 and it is not possible to derive optimality results. If there are only a few missing values, or the sets of locations are at least similar, we propose to proceed by first kriging the secondary variables to a common set of locations and then apply the aggregation-cokriging procedure.

When the weights do not depend on s_p, predicting at more than one location is straightforward by changing the covariance matrices in (7)–(9) accordingly.

If the primary and the secondary variables have a polynomial or regression type mean, it is possible to formulate the aggregation-cokriging solution as a minimization problem. However, except for very simple or pathological examples, we are not able to formulate closed form expressions for the optimal solutions. Nevertheless, the canonical correlation based derivation of the aggregation-cokriging weights can still be made because canonical correlation analysis is mean invariant (Mardia et al., 1979, Theorem 10.2.4).

ACKNOWLEDGEMENT

This research was sponsored by the National Science Foundation, U.S.A., and by an award made by the King Abdullah University of Science and Technology. We acknowledge the international modelling groups for providing their data for analysis. We also thank the editor, the associate editor and two referees for comments that led to a substantial improvement of the paper.

APPENDIX

Throughout the paper and the subsequent proofs, we must evaluate quantities of the form $x^{\mathrm{T}}C^{-1}w$, where the vectors and matrices involved are often partitioned. We first evaluate and simplify such quantities. In the following, assume all matrices and vectors have the required dimensions, and are positive definite where required. Then:

$$\begin{pmatrix} x^{\mathrm{T}} & y^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} C & B \\ B^{\mathrm{T}} & A \end{pmatrix}^{-1} \begin{pmatrix} x \\ y \end{pmatrix} = x^{\mathrm{T}}C^{-1}x + \|y - B^{\mathrm{T}}C^{-1}x\|_E^2$$

with $E = (A - B^{\mathrm{T}}C^{-1}B)^{-1}$ and $\|z\|_E^2 = z^{\mathrm{T}}Ez$. If D is a nonsingular matrix, then

$$\begin{pmatrix} x^{\mathrm{T}} & y^{\mathrm{T}}D^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} C & BD^{\mathrm{T}} \\ DB^{\mathrm{T}} & DAD^{\mathrm{T}} \end{pmatrix}^{-1} \begin{pmatrix} w \\ Dz \end{pmatrix} \tag{A1}$$

$$= x^{\mathrm{T}}C^{-1}w + x^{\mathrm{T}}C^{-1}BEB^{\mathrm{T}}C^{-1}w - x^{\mathrm{T}}C^{-1}BEz - y^{\mathrm{T}}E^{\mathrm{T}}BC^{-1}w + y^{\mathrm{T}}Ez. \tag{A2}$$

Proof of Lemma 1. Define the diagonal matrix containing the nonzero scalings $G = \mathrm{diag}(\gamma_1, \ldots, \gamma_m)$. Using scalings results in using $G\mathcal{A}$ instead of $\mathcal{A}$ and we have to prove that (2) and (3) are invariant under G. Both the predictor and the mean squared prediction error are of the form (A1), with D taking the role of G. Because (A2) is independent of D, the prediction and the mean squared prediction error are independent of the scalings $\gamma_1, \ldots, \gamma_m$. □

Proof of Proposition 2. With (11), we have $c_k = \omega_k(\Sigma_{cp} - \Sigma_{c0}\Sigma_{00}^{-1}\Sigma_{0p})$ and $S_{rs} = \Omega_{rs}\Sigma_{cc} - \omega_r\omega_s\Sigma_{c0}\Sigma_{00}^{-1}\Sigma_{0c}$ and (10) may be written as

$$\left(\sum_{k=1}^{\ell} a_k\omega_k\tilde{c}\right)^{\mathrm{T}} \left(\sum_{r,s=1}^{\ell} a_r a_s \Omega_{rs}\Sigma_{cc} - a_r a_s \omega_r\omega_s\tilde{S}\right)^{-1} \left(\sum_{l=1}^{\ell} a_l\omega_k\tilde{c}\right),$$

where $\tilde{c} = c_k/\omega_k$ and $\tilde{S} = \Sigma_{c0}\Sigma_{00}^{-1}\Sigma_{0c}$. Similar to Lemma 1, we scale a such that $a^\top\omega = 1$. Hence,

$$\tilde{c}^\top(a^\top\Omega a\,\Sigma_{cc} - \tilde{S})^{-1}\tilde{c} = \tilde{c}^\top U(\lambda)^{-1}\tilde{c},$$

with $\lambda = a^\top\Omega a$ and $U(\lambda) = \lambda\Sigma_{cc} - \tilde{S}$. Note that $\lambda > 0$ because $a \neq 0$ and $U(\lambda)$ is positive definite because (11) is so. For any $\tilde{c}$, maximizing $\max_\lambda \tilde{c}^\top U(\lambda)^{-1}\tilde{c}$ is equivalent to $\min_\lambda \tilde{c}^\top U(\lambda)\tilde{c}$ and to $\min_\lambda \lambda\tilde{c}^\top\Sigma_{cc}\tilde{c}$, and to $\min_a \lambda = \min_a a^\top\Omega a$, yielding the desired result. □

Proof of Lemma 2. Because $a(s_i)^\top 1 \neq 0$, we scale the weight vectors and use $\tilde{a}_i = a(s_i)\{a(s_i)^\top 1\}^{-1}$. Then with $\mathcal{A}$ based on $\tilde{a}_i$ we have

$$\mathcal{A}(\Sigma_{\mathcal{Y}0}\Sigma_{00}^{-1}\Sigma_{0p} - \Sigma_{\mathcal{Y}p}) = \sum_{r=1}^{\ell}\operatorname{diag}(L_r)c_r = c_1,$$

where $c_k = \Sigma_{k0}\Sigma_{00}^{-1}\Sigma_{0p} - \Sigma_{kp}$. Let $E_\mathcal{A} = \{\mathcal{A}(\Sigma_{\mathcal{Y}\mathcal{Y}} - \Sigma_{\mathcal{Y}0}\Sigma_{00}^{-1}\Sigma_{0\mathcal{Y}})\mathcal{A}^\top\}^{-1}$. Equivalent expressions hold for the second aggregation matrix. We need to show that $c_1^\top E_\mathcal{B} c_1 < c_1^\top E_\mathcal{A} c_1$ if and only if $0 < c_1^\top(E_\mathcal{A} - E_\mathcal{B})c_1$. Now, $E_\mathcal{A} - E_\mathcal{B}$ is positive definite if and only if $E_\mathcal{B}^{-1} - E_\mathcal{A}^{-1}$ is positive definite (Harville, 1997, Theorem 18.2.4), if and only if

$$\mathcal{B}(\Sigma_{\mathcal{Y}\mathcal{Y}} - \Sigma_{\mathcal{Y}0}\Sigma_{00}^{-1}\Sigma_{0\mathcal{Y}})\mathcal{B}^\top - \mathcal{A}(\Sigma_{\mathcal{Y}\mathcal{Y}} - \Sigma_{\mathcal{Y}0}\Sigma_{00}^{-1}\Sigma_{0\mathcal{Y}})\mathcal{A}^\top \tag{A3}$$

is positive definite. The matrix $\Sigma_{\mathcal{Y}\mathcal{Y}} - \Sigma_{\mathcal{Y}0}\Sigma_{00}^{-1}\Sigma_{0\mathcal{Y}}$ is defined by its ℓ^2 blocks, $\Sigma_{rs} - \Sigma_{r0}\Sigma_{00}^{-1}\Sigma_{0s}$, each being a positive definite $m \times m$ matrix. The second term of the blocks does not depend on s or r. Thus,

$$\begin{aligned}\mathcal{A}(\Sigma_{\mathcal{Y}\mathcal{Y}} - \Sigma_{\mathcal{Y}0}\Sigma_{00}^{-1}\Sigma_{0\mathcal{Y}})\mathcal{A}^\top &= \sum_{r=1}^{\ell}\sum_{s=1}^{\ell}\operatorname{diag}(L_r)(\Sigma_{rs} - \Sigma_{r0}\Sigma_{00}^{-1}\Sigma_{0s})\operatorname{diag}(L_s)\\ &= \sum_{r=1}^{\ell}\sum_{s=1}^{\ell}\Sigma_{rs}\circ L_r L_s^\top - \sum_{r=1}^{\ell}\operatorname{diag}(L_r)\Sigma_{10}\Sigma_{00}^{-1}\Sigma_{01}\sum_{s=1}^{\ell}\operatorname{diag}(L_s)\\ &= \sum_{r=1}^{\ell}\sum_{s=1}^{\ell}\Sigma_{rs}\circ L_r L_s^\top.\end{aligned}$$

Therefore, (A3) is positive definite if and only if (12) is positive definite. □

Proof of Lemma 3. In the case of constant weights given by $a(a^\top 1)^{-1}$ and $b(b^\top 1)^{-1}$, $G_{rs} = a_r a_s(a^\top 1)^{-2}11^\top$ and accordingly $H_{rs} = b_r b_s(b^\top 1)^{-2}11^\top$. Therefore, (12) reduces to $\sum_{r=1}^{\ell}\sum_{s=1}^{\ell}\Sigma_{rs}\circ 11^\top\{b_r b_s(b^\top 1)^{-2} - a_r a_s(a^\top 1)^{-2}\}$, proving the first statement. A sum of positive definite matrices is positive definite and the Schur product of a positive definite matrix with a nonnegative definite matrix is positive definite (Horn & Johnson, 1994, Theorem 5.2.1), proving the second statement. For the third statement, we write $\Sigma_{rr} = K_r + \Sigma_{12}$, implying $\sum_{r=1}^{\ell}\sum_{s=1}^{\ell}\Sigma_{rs} = \ell^2\Sigma_{12} + \sum_{r=1}^{\ell}K_r$. Hence, $\sum_{r=1}^{\ell}\sum_{s=1}^{\ell}\operatorname{diag}(L_r)\Sigma_{rs}\operatorname{diag}(L_r) = \ell^2\Sigma_{12} + \sum_{r=1}^{\ell}K_r \circ 11^\top a_r^2(a^\top 1)^{-2}$, because $\sum_{r=1}^{\ell}L_r = 1$. Finally, if $\Sigma_{rr} = \Sigma_{11}$, the problem reduces to $\sum_{r=1}^{\ell}b_r^2(b^\top 1)^{-2} - a_r^2(a^\top 1)^{-2}$ being positive. □

Proof of Corollary 1. The first item is trivial. In the case of bikriging, $b = e_1$ and the second item follows immediately from the last statement of Lemma 3. □

Proof of Proposition 3. The matrices C_{ii} and C_{ip} defined by (7) and (8) are of the form $b_1 I + b_2 11^\top$ and $b_3 1$. Thus, the vector $a = 1$ is an eigenvector of $C_{ii}^{-1}C_{ip}C_{pi}$ and because of the symmetry of C_{ii} also of $C_{ii}^{-1/2}C_{ip}C_{pi}C_{ii}^{-1/2}$. The covariance structure with (11) implies that ω is proportional to 1 and $\Omega = \alpha I + \beta 11^\top$, for some constants α and β. Recall that the largest eigenvalue of Ω is $\alpha + \ell\beta$ with associated eigenvector 1. Further, AGG(Ω, ω)-cokriging uses an aggregation vector $a = \Omega^{-1}1(1^\top\Omega^{-1}1)^{-1}$ and with $\Omega^{-1} = \alpha^{-1}I - \beta(\alpha^2 + \ell\alpha\beta)^{-1}11^\top$ we can conclude that a is proportional to 1.

We write the covariance between different secondary variables as Σ_{12} and the within covariance of the secondary variables as $\Sigma_{12} + \Sigma_R$. For aggregation-cokriging, we use $\mathcal{A} = 1^\top \otimes I$ to show that

$$g(1) = m(\Sigma_{1p} - \Sigma_{10}\Sigma_{00}^{-1}\Sigma_{0p})^\top(\Sigma_R + m\Sigma_{12} - \Sigma_{10}\Sigma_{00}^{-1}\Sigma_{01})^{-1}(\Sigma_{1p} - \Sigma_{10}\Sigma_{00}^{-1}\Sigma_{0p}).$$

For cokriging, we write the matrices with Kronecker products, namely

$$(\Sigma_{\mathcal{Y}p} - \Sigma_{\mathcal{Y}0}\Sigma_{00}^{-1}\Sigma_{0p})^\top(\Sigma_{\mathcal{Y}\mathcal{Y}} - \Sigma_{\mathcal{Y}0}\Sigma_{00}^{-1}\Sigma_{0\mathcal{Y}})^{-1}(\Sigma_{\mathcal{Y}p} - \Sigma_{\mathcal{Y}0}\Sigma_{00}^{-1}\Sigma_{0p}) = \{1^\top \otimes (\Sigma_{1p} - \Sigma_{10}\Sigma_{00}^{-1}\Sigma_{0p})\}^\top$$
$$\times (I \otimes \Sigma_R + 11^\top \otimes \Sigma_{12} - 11^\top \otimes \Sigma_{10}\Sigma_{00}^{-1}\Sigma_{01})^{-1}\{1^\top \otimes (\Sigma_{1p} - \Sigma_{10}\Sigma_{00}^{-1}\Sigma_{0p})\}^\top = g(1). \quad \square$$

REFERENCES

ABOURFIRASSI, M. & MARINO, M. A. (1984). Cokriging of aquifer transmissivities from field measurements of transmissivity and specific capacity. *J. Int. Assoc. Math. Geol.* **16**, 19–35.

APANASOVICH, T. V. & GENTON, M. G. (2010). Cross-covariance functions for multivariate random fields based on latent dimensions. *Biometrika* **97**, 15–30.

BANERJEE, S., GELFAND, A. E., FINLEY, A. O. & SANG, H. (2008). Gaussian predictive process models for large spatial data sets. *J. R. Statist. Soc.* B **70**, 825–48.

BERLINER, L. M., WIKLE, C. K. & CRESSIE, N. (2000). Long-lead prediction of Pacific SSTs via Bayesian dynamic modeling. *J. Climate.* **13**, 3953–68.

BILLINGS, S. D., BEATSON, R. K. & NEWSAM, G. N. (2002). Interpolation of geophysical data using continuous global surfaces. *Geophysics* **67**, 1810–22.

BOURGAULT, G. & MARCOTTE, D. (1991). Multivariable variogram and its application to the linear model of coregionalization. *Math. Geol.* **23**, 899–928.

BROHAN, P., KENNEDY J. J., HARIS, I., TETT, S. F. B. & JONES, P. D. (2006). Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. *J. Geophys. Res.* **111**, D12106.

CARR, J. R. & MCCALLISTER, P. G. (1985). An application of cokriging for estimation of tripartite earthquake response spectra. *J. Int. Assoc. Math. Geol.* **17**, 527–45.

CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*. New York: John Wiley & Sons Inc., revised reprint.

CRESSIE, N. & JOHANNESSON, G. (2008). Fixed rank kriging for very large spatial data sets. *J. R. Statist. Soc.* B **70**, 209–26.

DAVIS, B. M. & GREENES, K. A. (1983). Estimation using spatially distributed multivariate data: an example with coal quality. *J. Int. Assoc. Math. Geol.* **15**, 291–304.

FURRER, R. & SAIN, S. R. (2010). spam: A sparse matrix R package with emphasis on MCMC methods for Gaussian Markov random fields. *J. Statist. Software* **36**, 1–25.

FURRER, R., GENTON, M. G. & NYCHKA, D. (2006). Covariance tapering for interpolation of large spatial datasets. *J. Comp. Graph. Statist.* **15**, 502–23.

GELFAND, A., SCHMIDT, A., BANERJEE, S. & SIRMANS, C. (2004). Nonstationary multivariate process modelling through spatially varying coregionalization (with discussion). *Test* **13**, 1–50.

GENTON, M. G. (2007). Separable approximations of space-time covariance matrices. *Environmetrics* **18**, 681–95.

GNEITING, T. (2002). Compactly supported correlation functions. *J. Mult. Anal.* **83**, 493–508.

HAAS, T. C. (1990a). Kriging and automated variogram modeling within a moving window. *Atmosph. Envir., Part A* **24**, 1759–69.

HAAS, T. C. (1990b). Lognormal and moving window methods of estimating acid deposition. *J. Am. Statist. Assoc.* **85**, 950–63.

HARVILLE, D. A. (1997). *Matrix Algebra From a Statistician's Perspective*. New York: Springer.

HERTZOG, A., BASDEVANT, C. & VIAL, F. (2006). An assessment of ECMWF and NCEP–NCAR reanalyses in the southern hemisphere at the end of the presatellite era: results from the EOLE experiment (1971–72). *Mon. Weather Rev.* **134**, 3367–83.

HORN, R. A. & JOHNSON, C. R. (1994). *Topics in Matrix Analysis*. Cambridge: Cambridge University Press.

JOHNS, C., NYCHKA, D., KITTEL, T. & DALY, C. (2003). Infilling sparse records of spatial fields. *J. Am. Statist. Assoc.* **98**, 796–806.

JOLLIFFE, I. T. (2002). *Principal Component Analysis*. 2nd ed. New York: Springer.

JONES, P. D., NEW, M., PARKER, D. E., MARTIN, S. & RIGOR, I. G. (1999). Surface air temperature and its variations over the last 150 years. *Rev. Geophys.* **37**, 173–99.

JOURNEL, A. G. & HUIJBREGTS, C. J. (1978). *Mining Geostatistics*. London: Academic Press.

KALNAY, E., KANAMITSU, M., KISTLER, R., COLLINS, W., DEAVEN, D., GANDIN, L., IREDELL, M., SAHA, S., WHITE, G., WOOLEN, J., ZHU, Y., CHELLIAH, M., EBISUZAKI, W., HIGGINS, W., JANOWIAK, J., MO, K. C., ROPELEWSKI, C.,

Wang, J., Leetma, A., Reynolds, R., Jenne, R. & Joseph, D. (1996). The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteor. Soc.* **77**, 437–71.
Long, A. E. & Myers, D. E. (1997). A new form of the cokriging equations. *Math. Geol.* **29**, 685–703.
Ma, L., Zhang, T., Li, Q., Frauenfeld, O. W. & Qin, D. (2008). Evaluation of ERA-40, NCEP-1, and NCEP-2 reanalysis air temperatures with ground-based measurements in China. *J. Geophys. Res.* **113**, D15115.
Majumdar, A. & Gelfand, A. E. (2007). Multivariate spatial modeling using convolved covariance functions. *Math. Geol.* **39**, 225–45.
Mardia, K. V., Kent, J. T. & Bibby, J. M. (1979). *Multivariate Analysis*. London: Academic Press.
Matheron, G. (1962). Traité de géostatistique appliquée, Tome I. *In Mémoires du Bureau de Recherches Géologiques et Minières* **No. 14**, Paris: Editions Technip.
Meehl, G. A., Stocker, T. F., Collins, W. D., Friedlingstein, P., Gaye, A. T., Gregory, J. M., Kitoh, A., Knutti, R., Murphy, J. M., Noda, A., Raper, S. C. B., Watterson, I. G., Weaver, A. J. & Zhao, Z.-C. (2007). Global climate projections. In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Ed. S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor & H. L. Miller. Cambridge: Cambridge University Press.
Myers, D. E. (1982). Matrix formulation of co-kriging. *J. Int. Assoc. Math. Geol.* **14**, 249–57.
Myers, D. E. (1983). Estimation of linear combinations and co-kriging. *J. Int. Assoc. Math. Geol.* **15**, 633–7.
Myers, D. E. (1992). Kriging, co-kriging, radial basis functions and the role of positive definiteness. *Comp. Math. Appl.* **24**, 139–48.
Schmidt, A. & Gelfand, A. (2003). A Bayesian coregionalization approach for multivariate pollutant data. *J. Geophy. Res. A* **108**, D24, 8783.
Sun, Y., Li, B. & Genton, M. G. (2011). Geostatistics for large datasets. In *Advances and Challenges in Space-time Modelling of Natural Events*, Ed. J. M. Montero, E. Porcu & M. Schlather. Springer, in press. http://www.stat.tamu.edu/∼genton/2011.large.book.pdf.
Ver Hoef, J. M. & Cressie, N. (1993). Multivariable spatial prediction. *Math. Geol.* **25**, 219–39.
Ver Hoef, J. M. & Cressie, N. (1994). Errata: multivariable spatial prediction. *Math. Geol.* **26**, 273–5.
Wackernagel, H. (2006). *Multivariate Geostatistics*, 3rd ed. New York: Springer.

[*Received September* 2010. *Revised April* 2011]