Area-to-point Kriging with inequality-type data

E.-H. Yoo and P. C. Kyriakidis

Department of Geography, University of California Santa Barbara,
Ellison Hall 5831, Santa Barbara, CA 93106-4060, U.S.A.
Telephone: +1 (805) 259-7541
FAX: +1 (805) 893-3146
E-mail:yoo@geog.ucsb.edu

Abstract

In practical applications of area-to-point spatial interpolation, inequality constraints, such as non-negativity, or more general constraints on the maximum and/or minimum allowable value of the resulting predictions, should be taken into account. The geostatistical framework proposed in this paper deals with area-to-point interpolation problems under such constraints, while: (i) explicitly accounting for support differences between sample data and unknown values, (ii) guaranteeing coherent predictions, and (iii) providing a measure of reliability for the resulting predictions. The analogy between the dual form of area-to-point interpolation and a spline allows to solve constrained area-to-point interpolation problems via a constrained quadratic minimization algorithm, after accounting for the following three issues: (i) equality and inequality constraints could be applied to different supports, and such support differences should be considered explicitly in the problem formulation, (ii) if inequality constraints are enforced on the entire set of points discretizing the areal data, it is impossible to obtain a solution of the quadratic programming problem, and (iii) the uniqueness and existence of the solution has to be diagnosed. In this work, stable and efficient computation of point predictions is achieved through the following two steps: (i) initial prediction at all locations via unconstrained area-to-point interpolation, and (ii) constrained area-to-point interpolation with inequality information only at those points whose initial predicted values violate the inequality constraints. Last, the application of the proposed method to area-to-point spatial interpolation with inequality constraints in one and two dimensions is demonstrated using realistically simulated data.

1 Introduction

In many problems of modeling spatial data, measurements are rarely enough to reconstruct the true surface from which they were sampled. For example, population density surfaces, which are conventionally derived as aggregates over irregularly shaped geographic regions, may be highly biased due to the insufficient information content of the data. Under these circumstances, ancillary information, which may take the form of inequality constraints, such as non-negativity for population density surfaces, can complement the lack of measurements.

Surface reconstruction is a common problem across several disciplines so various methods have been proposed for incorporating inequality-type data into the construction of surface models with measurement data. A simple traditional approach is to reset the violating initial predictions to the nearest bound of the physical interval, which often produces artifact discontinuities in the resulting surface.

In the geostatistical literature, more elaborate methods are found, such as (i) constraints on the Kriging weights, (ii) the soft-Kriging approach, and (iii) constrained predictions. Various approaches have been proposed to constrain Kriging weights, attributing violating Kriging predictions to negative Kriging weights. For example, Barnes and Johnson (1984) provided an algorithm to produce non-negative Kriging weights, using an iterative solution based on the Kuhn-Tucker theorem, although the convergence of this algorithm is not guaranteed, and Limic and Mikelic (1984) showed how to add positivity constraints on Kriging weights to obtain positive predictions using Quadratic Programming (QP) techniques. Recently, Szidarovszky, Baafi and Kim (1987) and Deutsch (1996) also provided algorithms that lead to non-negative Kriging weights. However, it is important to recognize that positive weights is a sufficient, but non-necessary condition for positive Kriging predictions (Journel, 1986; Walvoort and de Gruijter, 2001), and may produce abrupt changes at locations where the constraints are imposed. An alternative approach is to impose constraints on the Kriging predictions rather than on the Kriging weights through the use of indicator constraint intervals, as suggested by Journel (1986). The “soft-Kriging approach”, whose formalism allows building a predictor which honors prior constraint intervals, including a constraint interval at all points in the study area, addresses the confidence intervals for the resulting predictions, but requires internal consistency of the covariance models used for the indicator variable, which is hard to obtain in practice. Another approach is to constrain the Kriging predictions via non-linear optimization techniques (Barnes and You, 1992; Dubrule and Kostov, 1986; Kostov and Dubrule, 1986); this approach will be extended in this paper to address constrained area-to-point prediction. Outside geostatitics, pioneering work has been done by Tobler (1979), who proposed a pychnophylactic interpolation method to account for constraints, including mass-preservation and and non-negativity. Furthermore, his work motivated follow-up solutions to “volume-matching problems” (or “area-matching problems”), which are connected to the variational spline theories in statistics (Dyn and Wong, 1987; Wong, 1980).

In this paper, we adopt a geostatistical approach to constrain prediction in the realm of non-linear optimization problems. Two main methods will be applied to the task of area-to-point prediction, while taking into account the support differences between areal data and point predictions explicitly, and providing the uncertainty of such predictions. The first method is based on the primal form of Kriging so that the minimization of Kriging variance will be extended to incorporate the inequality constraints, while the second method is based on the analogy of dual Kriging formalism with splines. The second approach produces very similar results with the solution proposed by Tobler (1979) and Dyn and Wong (1987) in a particular case. In the following sections, we outline the two methods of constrained area-to-point prediction, and present comparative results in a case study.

2 Theory

Consider the problem of predicting the values of a continuous attribute z at a set of P prediction locations $\{\mathbf{u}_p, p = 1, \ldots, P\}$ within a study area A, using areal data defined on a set of K supports $\{v_k, k = 1, \ldots, K\}$. Here, the notation $\mathbf{u}_p$ denotes the coordinate vector of the p-th location, and v_k denotes the k-th support with centroid $\mathbf{u}_k$ and otherwise arbitrary shape. The number of prediction locations within the k-th support v_k is denoted as P_k, with $\sum_{k=1}^{K} P_k = P$. Note that a special case of an areal support whose measure is infinitesimally small is a point support: $P_k = 1$, if the k-th support v_k is reduced to a point $\mathbf{u}_k$. In the following discussion, the unknown value at the location $\mathbf{u}$, denoted as $z(\mathbf{u})$, is viewed as a realization of a random variable (RV), denoted as $Z(\mathbf{u})$, and the set of all points RVs in the study domain $\{Z(\mathbf{u}),\ \mathbf{u} \in \mathbf{A}\}$ constitutes a random function (RF).

In this paper, we focus on area-to-point Kriging, which is an optimal linear predictor in the sense that it minimizes the prediction error variance under the constraint of unbiasedness of the predictor (Chilès and Delfiner, 1999; Cressie, 1993; Isaaks and Srivastava, 1989). The following discussion is based on the intrinsically stationary RF, where the process is characterized by the mean $m_Z(\mathbf{u})$ and the covariance function $C_Z(\mathbf{u}, \mathbf{u}')$ or the variogram $\gamma_Z(\mathbf{u}, \mathbf{u}')$. Under intrinsic stationarity, the expected value of the difference (increment) between any two RVs $Z(\mathbf{u})$ and $Z(\mathbf{u}+\mathbf{h})$ separated by a distance vector $\mathbf{h} = |\mathbf{u}' - \mathbf{u}|,\ \forall\ \mathbf{u}, \mathbf{u}' \in A$ is zero: $E[Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u})] = 0$, which entails that the expected value of any RV $Z(\mathbf{u})$ is constant but unknown: $E[Z(\mathbf{u}+\mathbf{h})] = m_Z,\ \forall \mathbf{u} \in A$. The variance of the difference of any two RVs $Z(\mathbf{u})$ and $Z(\mathbf{u}+\mathbf{h})$ depends only on the separation distance vector $\mathbf{h}$: $Var[Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u})] = 2\gamma_Z(\mathbf{h})$.

Suppose that the areal datum corresponding to the k-th support v_k, denoted $d(v_k)$, is linked to point support values via a sampling function g_k:

$$d(v_k) = \int_{\mathbf{u} \in v_k} g_k(\mathbf{u}) z(\mathbf{u}) d\mathbf{u} \simeq \sum_{p=1}^{P_k} g_k(\mathbf{u}_p) z(\mathbf{u}_p)$$

where P_k represents the number of points used to discretize the areal support v_k, and $g_k(\mathbf{u}_p)$ denotes the value of the sampling function with respect to the k-th support v_k at the location $\mathbf{u}_p$. Some typical examples of the sampling function (Kyriakidis and Yoo, 2005) include an indicator function $g_k(\mathbf{u}_p) = 1$, if $\mathbf{u}_p \in v_k$, 0 if not, where the areal datum $d(v_k)$ is simply the sum of all points within support v_k. Another sampling function that has a wide application is an indicator function normalized by the total number P_k of points within any support v_k: $g_k(\mathbf{u}_p) = 1/P_k$, if $\mathbf{u}_p \in v_k$, 0 if not. In this case, the areal datum $d(v_k)$ is simply the mean of all P_k point values within support v_k.

Since the k-th areal support datum $d(v_k)$ is viewed as a realization of a new RV $D(v_k)$, the moments of the K areal RVs $\{D(v_k),\ k = 1, \ldots, K\}$ are functionally related to those of the point RF $\{Z(\mathbf{u}), \mathbf{u} \in A\}$. The mean of the k-th areal RV $D(v_k)$, denoted as $m_D(v_k)$, is a linear function of the mean of point support values:

$$\begin{aligned} E\{D(v_k)\} &= E\left\{\sum_{p=1}^{P_k} g_k(\mathbf{u}_p) z(\mathbf{u}_p)\right\} = \sum_{p=1}^{P_k} g_k(\mathbf{u}_p) E\{Z(\mathbf{u}_p)\} \\ &= \sum_{p=1}^{P_k} g_k(\mathbf{u}_p) m_Z = m_D(v_k) \end{aligned}$$

The covariance between any two areal RVs $D(v_k)$ and $D(v_{k'})$ is a double weighted linear combination of point covariance values $C_Z(|\mathbf{u}_p - \mathbf{u}_{p'}|)$ between all possible vectors are formed by any two prediction locations $\mathbf{u}_p \in v_k,\ \mathbf{u}_{p'} \in v_{k'}$ (Arbia, 1989):

$$Cov\{D(v_k), D(v_{k'})\} = \sum_{p=1}^{P_k} \sum_{p'=1}^{P_{k'}} g_k(\mathbf{u}_p) g_k(\mathbf{u}_{p'}) C_Z(|\mathbf{u}_p - \mathbf{u}_{p'}|) = C_D(v_k, v_{k'})$$

The cross-covariance between any areal RV $D(v_k)$ and any point RV $Z(\mathbf{u}_{p'})$, denoted as $C_{DZ}(v_k, \mathbf{u}_{p'})$, is derived as:

$$\begin{aligned} Cov\{D(v_k), Z(\mathbf{u}_{p'})\} &= Cov\left\{\left[\sum_{p=1}^{P_k} g_k(\mathbf{u}_p) Z(\mathbf{u}_p)\right], Z(\mathbf{u}_{p'})\right\} = \sum_{p=1}^{P_k} g_k(\mathbf{u}_p) Cov\{Z(\mathbf{u}_p), Z(\mathbf{u}_{p'})\} \\ &= \sum_{p=1}^{P_k} g_k(\mathbf{u}_p) C_Z(|\mathbf{u}_p - \mathbf{u}_{p'}|) = C_{DZ}(v_k, \mathbf{u}_{p'}) \end{aligned}$$

Note that the cross-covariance C_{DZ} is nonstationary in the general case of unequal supports, $P_k \neq P_{k'}$ or different sampling functions $g_k \neq g_{k'}$. In other words, for any pair of supports $v_k, v_{k'}$ whose centroids $\mathbf{u}_k, \mathbf{u}_{k'}$ are equidistant from a prediction location $\mathbf{u}_{p'}, C_{DZ}(v_k, \mathbf{u}_{p'}) \neq C_{DZ}(v_{k'}, \mathbf{u}_{p'})$, and the cross-covariance C_{DZ} is stationary only for fixed supports and a constant sampling function. For a complete exposition to geostatistical area-to-point prediction, see Kyriakidis (2004).

2.1 Kriging as a nonlinear optimization problem

Kriging is a family of generalized linear regression algorithms so that all variants of Kriging predictors can be expressed as a basic linear regression estimator: the sum of the mean and the weighted linear combination of residuals. The residuals are weighted by Kriging weights which are determined so that they minimize the prediction error variance subject to unbiasedness constraints (Isaaks and Srivastava, 1989; Journel and Huijbregts, 1978). Interestingly enough, however, the method of Kriging has been studied not only in regression and geostatistics, but also in the field of optimization by Luenberger (1969) whose complete theory of a vectorial approach to optimization applies to the Kriging system of equations as a special case (Olea, 1999).
In the following section, the primal form of area-to-point ordinary Kriging (OK) prediction is reviewed from the perspective of an optimization problem, and one of its variants, a bounded ordinary Kriging (BOK) subject to inequality constraints, is discussed.

2.1.1 Area-to-point Kriging subject to equality constraints only

Let Z be an intrinsic RF, where the mean of Z is assumed to be constant but unknown over the study domain A. The point prediction at location $\mathbf{u}_p$, using the $(K \times 1)$ vector $\mathbf{d} = [d(v_1), \ldots, d(v_K)]'$ of areal data, is given as:

$$\hat{z}(\mathbf{u}_p) = \sum_{k=1}^{K} \lambda_k(\mathbf{u}_p) d(v_k) = \sum_{k=1}^{K} \lambda_k(\mathbf{u}_p) d(v_k) + m_Z \left\{1 - \sum_{k=1}^{K} \sum_{p=1}^{P_k} g_k(\mathbf{u}_p) \lambda_k(\mathbf{u}_p)\right\}$$

$$= \quad [\boldsymbol{\lambda}_p' \quad m_Z][\begin{array}{c}\mathbf{d}\\ 0\end{array}] = \boldsymbol{\lambda}_p'\mathbf{d}, \quad \forall\, \mathbf{u}_p \in A \tag{1}$$

subject to:

$$\sum_{k=1}^{K}\sum_{p=1}^{P_k} g_k(\mathbf{u}_p)\lambda_k(\mathbf{u}_p) = \boldsymbol{\lambda}_p'\mathbf{G}\mathbf{1} = 1, \quad \forall\, \mathbf{u}_p \in A$$

where $\lambda_k(\mathbf{u}_p)$ is the OK weight assigned to the k-th areal datum for the prediction at location $\mathbf{u}_p$, and m_Z is an unknown but constant mean filtered from the linear predictor by forcing the Kriging weights to sum to 1. Since the mean of the k-th areal RV $m_D(v_k)$ is a linear function of the mean of point support values via a sampling function $g_k(\mathbf{u}_p)$, the weighted mean of the areal RVs is expressed as a double weighted combination of the mean of point RVs, and is written as: $\sum_{k=1}^{K}\lambda_k(\mathbf{u}_p)m_D(v_k) = \sum_{k=1}^{K}\sum_{p=1}^{P_k} g_k(\mathbf{u}_p)\lambda_k(\mathbf{u}_p)m_Z$. In matrix notation, $\boldsymbol{\lambda}_p$ denotes a $(K\times 1)$ vector of OK weights for the K areal data, $\mathbf{G}$ is a $(K\times P)$ matrix of sampling functions, and $\mathbf{1}$ is a $(P\times 1)$ vector of ones.

The Kriging weights are determined by minimizing the prediction error variance subject to the unbiasedness constraint on the weights. Therefore, the problem of area-to-point OK prediction can be restated as the problem of "finding K weights $\boldsymbol{\lambda}_p$ minimizing $Var\{\hat{Z}(\mathbf{u}_p) - Z(\mathbf{u}_p)\}$ subject to linear constraints" of the following form:

$$\text{Min} \quad \sigma^2(\mathbf{u}_p) = C_Z(\mathbf{0}) + \boldsymbol{\lambda}_p'\mathbf{C}_D\boldsymbol{\lambda}_p - 2\boldsymbol{\lambda}_p'\mathbf{c}_{DZ}^p$$

subject to

$$\boldsymbol{\lambda}_p'\mathbf{G}\mathbf{1} = 1$$

where $C_Z(\mathbf{0})$ denotes the "a priori" variance of the point support RF $\{Z(\mathbf{u}),\ \mathbf{u}\in A\}$, and $\mathbf{C}_D$ and $\mathbf{c}_{DZ}^p$ are a $(K\times K)$ matrix of covariances among the K areal supports and a $(K\times 1)$ vector of cross-covariances between the K areal supports and the prediction location $\mathbf{u}_p$, respectively. This problem is classically solved by converting the constrained minimization problem into an unconstrained one using the method of Lagrange multipliers. Let ξ be a Lagrange multiplier, then the Lagrange function for area-to-point OK is written as:

$$L(\boldsymbol{\lambda}_p,\xi) = \boldsymbol{\lambda}_p'\mathbf{C}_D\boldsymbol{\lambda}_p - 2\boldsymbol{\lambda}_p'\mathbf{c}_{DZ}^p + 2\xi(\boldsymbol{\lambda}_p'\mathbf{G}\mathbf{1} - 1) \tag{2}$$

The Lagrange function in Equation (2) is a quadratic expression in the unknown OK weights, and the constraint on the weights is linear. In such a case, the necessary and sufficient condition to have a unique global minimum is that $\boldsymbol{\lambda}_p'\mathbf{C}_D\boldsymbol{\lambda}_p \geq 0$, which is already satisfied by the positive definite condition on $\mathbf{C}_D$ (Olea, 1999). The unconstrained minimum variance problem in Equation (2) is obtained by equating the partial derivatives of the Lagrange function to zero, as follows:

$$\frac{1}{2}\frac{\partial L(\boldsymbol{\lambda}_p,\xi)}{\partial\boldsymbol{\lambda}_p} = \mathbf{C}_D\boldsymbol{\lambda}_p - \mathbf{c}_{DZ}^p + \xi\mathbf{G}\mathbf{1} = 0$$

$$\frac{1}{2}\frac{\partial L(\boldsymbol{\lambda}_p,\xi)}{\partial\xi} = \boldsymbol{\lambda}_p'\mathbf{G}\mathbf{1} - 1 = \mathbf{1}'\mathbf{G}'\boldsymbol{\lambda}_p - 1 = 0$$

This yields a system of $(K+1)$ linear equations with $(K+1)$ unknowns. The solution of this system provides the optimal weights $\boldsymbol{\lambda}_p$ for area-to-point prediction. In matrix notation, the OK system of equations becomes:

$$\begin{bmatrix} \mathbf{C}_D & \mathbf{G1} \\ \mathbf{1}'\mathbf{G}' & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda}_p \\ \xi \end{bmatrix} = \begin{bmatrix} \mathbf{c}_{DZ}^p \\ 1 \end{bmatrix} \tag{3}$$

2.1.2 Area-to-point Kriging subject to inequality constraints

In this section, a simple extension of area-to-point OK prediction is considered to place the predicted values between explicit lower and upper bounds, which is called “bounded ordinary Kriging” (BOK). The BOK predictor incorporates the bounds into geostatistical interpolation by a small algorithmic change and additional computing effort (Barnes and You, 1992).
Consider again the task of predicting the attribute value at the p-th location using the K areal data, but subject to the additional constraints at Q point locations $\{\mathbf{u}_q,\ q = 1,\ldots,Q(=Q_l+Q_u)\}$, where the inequality constraints consist of Q_l lower bounds and Q_u upper bounds. The new constraints do not affect the form of the predictor so that OK prediction at the p-th location is expressed as the weighted linear combination of the areal data, as written in Equation (1). However, the Kriging weights are altered by restricting the solution to satisfy bound constraints. The quadratic minimizer for the BOK predictor is written as:

$$\text{Min} \quad \sigma^2(\mathbf{u}_p) = C_Z(\mathbf{0}) + \boldsymbol{\lambda}_p'\mathbf{C}_D\boldsymbol{\lambda}_p - 2\boldsymbol{\lambda}_p'\mathbf{c}_{DZ}^p \tag{4}$$

subject to

$$\begin{aligned} \boldsymbol{\lambda}_p'\mathbf{G1} &= 1, & p &= 1,\ldots,P \\ \hat{z}(\mathbf{u}_p) &\geq z_l, & p &= (P+1),\ldots,(P+Q_l) \\ \hat{z}(\mathbf{u}_p) &\leq z_u, & p &= (P+q_l+1),\ldots,(P+Q) \end{aligned}$$

where z_l and z_u denote the lower and upper bound values, respectively. Due to the additional inequality constraints, the quadratic form of the minimization problem can not be solved using the classical Lagrange multiplier technique any more. As an alternative, the Kuhn-Tucker theorem, that is the central theorem for all constrained nonlinear optimization (Künzi, Tzschach and Zehnder, 1971), can provide the necessary and sufficient conditions for a solution to the BOK prediction *.
The application of the Kuhn-Tucker conditions to BOK yields the following Lagrange function:

$$L(\boldsymbol{\lambda}_p,\boldsymbol{\xi}) = \boldsymbol{\lambda}_p'\mathbf{C}_D\boldsymbol{\lambda}_p - 2\boldsymbol{\lambda}_p'\mathbf{c}_{DZ}^p + 2\xi_0(\boldsymbol{\lambda}_p'\mathbf{G1}-1) - 2\xi_1(\boldsymbol{\lambda}_p'\mathbf{d}-z_l) + 2\xi_2(\boldsymbol{\lambda}_p'\mathbf{d}-z_u) \tag{5}$$

where $\boldsymbol{\xi} = [\xi_0 \quad \xi_1 \quad \xi_2]'$ denotes the vector of Lagrange multipliers associated with unbiasedness and lower and upper bound constraints, respectively. Solving Equation (5) for the Kriging weights $\boldsymbol{\lambda}_p$ and the Lagrange multipliers ξ_0,ξ_1,ξ_2, we can derive the following system of equations (Barnes and You, 1992):

$$\frac{\partial L(\boldsymbol{\lambda}_p,\boldsymbol{\xi})}{\partial \boldsymbol{\lambda}_p} = \mathbf{C}_D\boldsymbol{\lambda}_p - \mathbf{c}_{DZ}^p + \xi_0\mathbf{G1} - \xi_1\mathbf{d} + \xi_2\mathbf{d} = 0 \tag{6}$$

*When the objective function and constraints are a family of convex functions, the Kuhn-Tucker equations provide both necessary and sufficient conditions for a global solution to the constrained minimization problem

$$
\begin{aligned}
\frac{\partial L(\boldsymbol{\lambda}_p, \partial \boldsymbol{\xi})}{\partial \xi_0} &= \boldsymbol{\lambda}_p' \mathbf{G} \mathbf{1} - 1 = 0 && (7) \\
\frac{\partial L(\boldsymbol{\lambda}_p, \boldsymbol{\xi})}{\partial \xi_1} &= \boldsymbol{\lambda}_p' \mathbf{d} - z_l \geq 0 && (8) \\
\xi_1 \frac{\partial L(\boldsymbol{\lambda}_p, \boldsymbol{\xi})}{\partial \xi_1} &= \xi_1 (\boldsymbol{\lambda}_p' \mathbf{d} - z_l) = 0 && (9) \\
\frac{\partial L(\boldsymbol{\lambda}_p, \boldsymbol{\xi})}{\partial \xi_2} &= \boldsymbol{\lambda}_p' \mathbf{d} - z_u \leq 0 && (10) \\
\xi_2 \frac{\partial L(\boldsymbol{\lambda}_p, \boldsymbol{\xi})}{\partial \xi_2} &= \xi_2 (\boldsymbol{\lambda}_p' \mathbf{d} - z_u) = 0 && (11) \\
\xi_1 &\geq 0 \\
\xi_2 &\geq 0
\end{aligned}
$$

For a better understanding of the BOK system, consider a case where the BOK prediction at the p-th location is greater than the lower bound ($\boldsymbol{\lambda}_p' \mathbf{d} > z_l$) and lower than the upper bound ($\boldsymbol{\lambda}_p' \mathbf{d} < z_u$). Per Equations (9) and (11), the two Lagrange multipliers for the upper and lower bounds, ξ_1 and ξ_2, must be zero. These conditions lead the objective function Equation (5) to be equivalent to OK prediction without inequality constraints. On the other hand, if the Lagrange multipliers associated with inequality constraints are not equal to zeros ($\xi_1 > 0$ or $\xi_2 > 0$), per Equations (9) and (11), the BOK predictions must equal to the bound values ($\boldsymbol{\lambda}_p' \mathbf{d} - z_l = 0$ or $\boldsymbol{\lambda}_p' \mathbf{d} - z_u = 0$). In words, whenever a predicted value violates a bound value, the optimal bounded prediction in BOK is fixed to that bound value.

The same argument holds for the BOK prediction error variance. When a BOK prediction falls between the lower and upper bounds, the corresponding BOK prediction error variance is the same as the variance derived without inequality constraints. However, when the predicted value violates the inequality constraints, i.e., when a predicted value is smaller than the lower bound so that the predicted value at the p-th location is bounded to the lower bound value, the Kuhn-Tucker conditions are summarized in the following matrix form:

$$
\begin{bmatrix} \mathbf{C}_D & \mathbf{G1} & \mathbf{d} \\ \mathbf{1}'\mathbf{G}' & 0 & 0 \\ \mathbf{d}' & 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda}_p \\ \xi_0 \\ -\xi_1 \end{bmatrix} = \begin{bmatrix} \mathbf{c}_{DZ}^p \\ 1 \\ z_l \end{bmatrix} \tag{12}
$$

Solving the BOK system in Equation (12), the following BOK prediction error variance, denoted as $\hat{\sigma}^2_{BOK}(\mathbf{u}_p)$, is derived:

$$
\begin{aligned}
\hat{\sigma}^2_{BOK}(\mathbf{u}_p) &= C_Z(\mathbf{0}) - [[\mathbf{c}_{DZ}^p]' \; 1 \; z_l] \begin{bmatrix} \mathbf{C}_D & \mathbf{G1} & \mathbf{d} \\ \mathbf{1}'\mathbf{G}' & 0 & 0 \\ \mathbf{d}' & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c}_{DZ}^p \\ 1 \\ z_l \end{bmatrix} && (13) \\
&= \hat{\sigma}^2_{OK}(\mathbf{u}_p) + [\hat{z}(\mathbf{u}_p) - z_l]^2 / \eta && (14)
\end{aligned}
$$

where $\hat{\sigma}^2_{OK}(\mathbf{u}_p)$ denotes the prediction error variance without inequality constraints. The derivation of BOK prediction error variance in Equation (14) from Equation (13) requires a few algebraic manipulations; see Appendix I for detailed derivation. The correction term in BOK variance is the difference between the unconstrained prediction and the bound value weighted by η, which involves the dual Kriging weights as:

$$\eta = [\mathbf{d}' \; 0] \begin{bmatrix} \mathbf{C}_D & \mathbf{G1} \\ \mathbf{1}'\mathbf{G}' & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{d} \\ 0 \end{bmatrix} \tag{15}$$

In words, the BOK variance is increased proportional to the discrepancy between the OK prediction and the bound value whenever a prediction violates inequality constraints, because the bound value is considered a known datum in the BOK prediction at the inequality constraint point. Therefore, as the discrepancy between the OK prediction and the bound value increases, the prediction error at location $\mathbf{u}_p$ is increased as well, and vice versa.
In summary, the BOK system allows incorporating inequality constraints into area-to-point prediction in an intuitive way with minimum computing effort, while accounting for support differences explicitly and reporting the uncertainty of the BOK predictions. In the application of BOK to area-to-point prediction, however, it is important to recognize that this solution is based on a "point by point" formalism of Kriging, which implies that the bound at one location does not affect the prediction at another location. However, it is highly expected that inequalities impact not only the points at which they are applied to, but also their neighborhood (Dubrule and Kostov, 1986). In consequence, the resulting area-to-point BOK prediction after considering inequality constraints does not guarantee the areal data reproduction property (or "coherence" of prediction) anymore. Furthermore, it is highly expected that the predicted surface may show "clamping effects[†]" at the bound locations, which may cause the surface to exhibit discontinuities often found due to a high nugget effect. In addition, the BOK system can not handle more general inequality constraints, such as interval inequality constraints, i.e. $z_l < \hat{z}(\mathbf{u}) < z_u$.

2.2 Kriging as an interpolation problem

In a classic interpolation problem, the main objective is to find an unknown function to represent reality without reference to the original data. Usually, this unknown function is approximated by a parametric function whose form is postulated in advance, either explicitly or implicitly. The parameters of the function are chosen so as to optimize some criterion of the goodness-of-fit at data locations, which can be statistical or deterministic (Chilès and Delfiner, 1999). Recall that the main objective of Kriging is also to reconstruct the unknown surface on the basis of values observed at a limited number of points or areas. In this context, Kriging can be viewed as an interpolator, although there exists a fundamental difference between Kriging and standard interpolators, such as splines (Chilès and Delfiner, 1999; Cressie, 1990; Wahba, 1990*a*). Interpolators focus on modeling the interpolated surface in a purely deterministic way, whereas Kriging is concerned more about modeling the spatial phenomenon itself based on the statistical characteristics demonstrated by the data via a data-based covariance (Chilès and Delfiner, 1999; Cressie, 1990).

[†]the interpolation function has exactly equal values at the set of bounds

In the following section, we examine the problem of defining an interpolation function based on the Kriging system so that inequality-type data can be more easily incorporated into the Kriging system based on the analogy with constrained splines.

2.2.1 Area-to-point Kriging interpolator

An alternative to the primary formalism of Kriging is given by Matheron (1981) whereby the Kriging predictor is expressed as a linear combination of covariance values weighted by the dual Kriging coefficients, instead of a weighted combination of data values. This dual formulation has drawn attention due to some interesting properties, such as its formal equivalence with splines and an efficient and fast algorithm to compute Kriging predictions with a unique neighborhood (Chilès and Delfiner, 1999; Galli, Murillo and Thomann, 1984; Goovaerts, 1997; Royer and Vieira, 1984). The dual formalism of ordinary Kriging (DOK) is is easily derived by rewriting the primary form of OK in Equation (1) with the normal equations in Equation (3), and using the fact that the inverse of covariance among areal data is symmetric, i.e., $[\mathbf{C}_D^{-1}]' = \mathbf{C}_D^{-1}$:

$$\begin{aligned} \hat{z}(\mathbf{u}_p) &= [\mathbf{d}' \quad 0][\begin{matrix} \boldsymbol{\lambda}_p \\ m_Z^p \end{matrix}] = [\mathbf{d}' \quad 0]\begin{bmatrix} \mathbf{C}_D & \mathbf{G1} \\ \mathbf{1}'\mathbf{G}' & 0 \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{c}_{DZ}^p \\ 1 \end{bmatrix} \\ &= [\boldsymbol{\omega}' \quad a_0][\begin{matrix} \mathbf{c}_{DZ}^p \\ 1 \end{matrix}] = \boldsymbol{\omega}'\mathbf{c}_{DZ}^p + a_0, \quad \forall\, \mathbf{u}_p \in A \\ &= \sum_{k=1}^{K} \omega(v_k) C_{DZ}(v_k, \mathbf{u}_p) + a_0 \end{aligned} \tag{16}$$

where a $(K \times 1)$ vector of DOK weights is denoted as $\boldsymbol{\omega} = [\omega(v_k), k = 1 \ldots, K]'$. Note that the k-th DOK weight $\omega(v_k)$ depends on the k-th areal datum $d(v_k)$ not on the prediction location $\mathbf{u}_p$ so that the set of DOK weights $\boldsymbol{\omega}$ does not need to be recomputed for every prediction location, as is the case with the primal Kriging weights $\boldsymbol{\lambda}_p$. This implies that the DOK prediction in Equation (16) appears as a deterministic function of the prediction location $\mathbf{u}_p$, since the predicted value at a location is obtained by plugging the distance vector $\mathbf{h} = |\mathbf{u}_p - \mathbf{u}|, \mathbf{u} \in v_k$ into the prediction function whose coefficients are already determined. In similar context, a_0 is also a location-free DOK coefficient, corresponding to the constant but unknown mean in the primary form of OK. From this point of view, DOK formalism in Equation (16) can be easily generalized to Kriging with a trend model [‡], where the mean component is modeled as a linear combination of drift functions.

The coefficients of DOK interpolator, denoted as $\boldsymbol{\omega}$ and a_0, are derived per solution of the following DOK system, which ensures exact data reproduction and unbiasedness [§].

$$\begin{bmatrix} \mathbf{C}_D & \mathbf{G1} \\ \mathbf{1}'\mathbf{G}' & 0 \end{bmatrix}\begin{bmatrix} \boldsymbol{\omega} \\ a_0 \end{bmatrix} = \begin{bmatrix} \mathbf{d} \\ 0 \end{bmatrix}$$

On the other hand, interpolation problems are solved by restricting the postulated functionals to meet some criteria, such as equality constraints where the interpolation function should pass

[‡] see Appendix II for the dual Kriging with a trend model

[§] Consider the case where all the K areal data are equally correlated with the unknown value, written as $C_{DZ}(v_k, \mathbf{u}_p) = C_{DZ}(v_{k'}, \mathbf{u}_p) = c$, $\forall\ k = 1, \ldots, K$. Then, $\hat{z}(\mathbf{u}_p) = c\sum_{k=1}^{K}\omega(v_k) + a_0$, and the unbiasedness condition, $\hat{z}(\mathbf{u}_p) = m_Z = a_0$, leads to the following relationship: $\sum_{k=1}^{K}\omega(v_k) = 0$ (Goovaerts, 1997).

through the data or inequality constraints where the postulated function should be within a constraint interval. For example, the “thin-plate splines”, one of the most commonly used interpolators, are derived by coefficients which minimize the bending energy of a thin plate, while honoring all the data values. The extended version of this smoothing spline, what is called “constrained spline problems”, can be also solved in the framework of spline theory by minimizing various objective functions (Dubrule and Kostov, 1986; Galli et al., 1984; Kostov and Dubrule, 1986; Wong, 1980).

2.2.2 Constrained thin-plate splines

We review solutions for thin-plate splines and their generalized form subject to inequality constraints with regards to area-to-point Kriging. Since dual Kriging is identical in form to the equations used for obtaining smoothing splines (Dubrule, 1983; Matheron, 1981; Wahba, 1990*b*), the solution to the problem of constrained smoothing splines can be applied to the constrained area-to-point Kriging prediction. Consider the following common data model for thin-plate splines and Kriging:

$$z(\mathbf{u}_k) = \theta(\mathbf{u}_k) + \varepsilon(\mathbf{u}_k)$$

where $z(\mathbf{u}_k)$ denotes the k-th point support datum at location $\mathbf{u}_k$ among K supports, denoted as $[z(\mathbf{u}_k),\ k = 1,\ldots,K]$, and $\theta(\mathbf{u}_k)$ represents an unknown smooth function in the case of splines (Hutchinson and Gessler, 1994) and the noise-filtered linear model of regionalization in Kriging (Wackernagel, 2003). Spatially uncorrelated errors in the data model are denoted as $\varepsilon(\mathbf{u}_k)$. In spline theory, the smooth function θ is estimated by minimizing the following expression (Wahba, 1990*b*) :

$$\text{Min} \quad \sum_{k=1}^{K} \{z(\mathbf{u}_k) - \hat{\theta}(\mathbf{u}_k)\}^2 + \rho J_m^d(\hat{\theta}) \tag{17}$$

with a smoothing parameter, denoted as ρ ($\rho > 0$), and a penalty function $J_m^d(\hat{\theta})$, which measures roughness of the m-th degree derivatives of θ in a d-dimensional space R^d. For example, the penalty function for the thin-plate spline in one-dimensional space R^1 is $J_2^1(\hat{\theta}) = \int_a^b [\hat{\theta}''(\mathbf{u})]^2 d\mathbf{u}$, and in two-dimensional space R^2, becomes $J_2^2(\hat{\theta}) = \int\int [\hat{\theta}''(\mathbf{u})]^2 d\mathbf{u}$, where θ'' denotes the second derivative of a smooth function θ.

Duchon (1977) showed that the unique solution to the minimization problem in Equation (17) exists for a set of M low-order monomials with the following explicit expression:

$$\hat{\theta}(\mathbf{u}) = \sum_{m=0}^{M} a_m f_m(\mathbf{u}) + \sum_{k=1}^{K} b_k \Upsilon(\mathbf{u}, \mathbf{u}_k) \tag{18}$$

where $f_m(\mathbf{u})$ are low-order polynomials corresponding to the drift functions in Kriging, and $\Upsilon(:)$ is a set of parametric functionals, which is equivalent to the variogram when it is replaced with a generalized covariance of an intrinsic RF (Chilès and Delfiner, 1999; Cressie, 1993; Dubrule and Kostov, 1986). For example, in one-dimensional space R^1, $\Upsilon(\mathbf{u}, \mathbf{u}_k) = \Upsilon(|\mathbf{h}|) = |\mathbf{h}|^3$, in two-dimensional space R^2, $\Upsilon(\mathbf{u}, \mathbf{u}_k) = \Upsilon(|\mathbf{h}|) = |\mathbf{h}|^2 log(|\mathbf{h}|)$, where $|\mathbf{h}|$ denotes the distance between $\mathbf{u}$ and $\mathbf{u}_k$. The thin-plate spline coefficients, denoted by $\{a_m, m = 0, \ldots, M\}$ and

$\{b_k, k = 1,\ldots,K\}$ in Equation (18), are obtained by solving a linear system of order $(M+1)$ with highly optimized computation procedures (Hutchinson and Gessler, 1994).
Let's extend the scope of the discussion from the standard thin-plate splines to the constrained thin-plate splines. By using an alternative mathematical formalism, we can arrive at the same result for the minimization, as well as incorporate additional constraints, such as inequality conditions or boundary value conditions, to the basic spline formulation.
Consider two normed linear spaces [¶] of X and Y, denoted as $\| X \|^2$ and $\| Y \|^2$ with a *bounded linear operator* [‖] T on X to Y, where $\| \;\|$ is the ordinary Euclidean *n*-space norm (Dubrule and Kostov, 1986; Luenberger, 1969; Prenter, 1975; Wahba, 1990*b*). Then, a smooth spline function, denoted as $\hat{\theta}(\mathbf{u})$, can be defined by minimizing the norm of $\hat{\theta}$ with a bounded linear operator, denoted as $\| T\hat{\theta} \|^2$, subject to given constraints, as shown in the following formulation:

$$\text{Min } J_m^d(\hat{\theta}) = \| T\hat{\theta} \|^2 \tag{19}$$

$$\begin{aligned} \hat{\theta}(\mathbf{u}_k) &= z_k, \qquad k = 1,\ldots,K \\ \hat{\theta}(\mathbf{u}_k) &\geq z_l, \qquad k = (K+1),\ldots,(K+Q) \end{aligned}$$

Note that the penalty function in Equation (19) is characterized by the minimum norm problem in Hilbert space (Luenberger, 1969; Prenter, 1975; Wahba, 1990*b*):

$$\| T\hat{\theta} \|^2 = < T\hat{\theta}, T\hat{\theta} > = \sum_{k=1}^{K+Q} b_k \hat{\theta}(\mathbf{u}_k) \quad \forall \theta \in X \tag{20}$$

where $< \cdot\,,\cdot >$ denotes the inner product of two vectors.
Therefore, the penalty function in the Equation (19) is reformulated, by plugging the extended definition of splines in Equation (18) into the Equation (20), as follows:

$$\| T\hat{\theta} \|^2 = \sum_{k=1}^{K+Q} b_k \hat{\theta}(\mathbf{u}_k) = \sum_{k=1}^{K+Q} b_k \left\{ \sum_{m=0}^{M} a_m f_m(\mathbf{u}_k) + \sum_{k'=1}^{K+Q} b_{k'} \Upsilon(\mathbf{u}_k, \mathbf{u}_{k'}) \right\} \tag{21}$$

where the set of M low-order monomials, denoted f_m, is selected among the basis function of the kernel of the bounded linear operator T (Wahba, 1990*b*). Therefore, the inner product between $T\hat{\theta}$ and the basis functions always becomes zero. For example, the inner product with a polynomial of degree 2 in R^2 results in the following relationships:

$$\sum_{k=1}^{K+Q} b_k 1 = \sum_{k=1}^{K+Q} b_k x = \sum_{k=1}^{K+Q} b_k y = 0$$

In summary, a constrained thin-plate smoothing spline is characterized by the coefficients $\{b_k, k = 1,\ldots,(K+Q)\}$ which minimize the following quadratic penalty function:

[¶] a linear space in which we can assign a notion of length $||x||$ to each vector x in X, where the usual norm is assumed $||x||^2 = \int_a^b x^2 dx$

[‖] a bounded linear operator implies that there exists a constant $c > 0$ with the property that $\| Tx \| \leq c \| x \|$ for each $x \in X$, which maps a variable X to Y, (Prenter, 1975)

$$\text{Min} \sum_{k=1}^{K+Q} \sum_{k'=1}^{K+Q} b_k b_{k'} \Upsilon(\mathbf{u}_k, \mathbf{u}_{k'}) \tag{22}$$

subject to

$$\begin{aligned} \hat{\theta}(\mathbf{u}_k) &= z_k, \qquad k = 1, \ldots, K \\ \hat{\theta}(\mathbf{u}_k) &\geq z_l, \qquad k = (K+1), \ldots, (K+Q) \end{aligned}$$

2.2.3 Constrained area-to-point prediction

In the previous two sections, we considered Kriging as an interpolator and reviewed a solution to a constrained thin-plate spline interpolator. The coefficients of constrained thin-plate splines are determined by finding a unique minimizer of the quadratic form of Equation (22), and then constraining the minimization problem via equality and inequality constraints. In two papers Dubrule and Kostov (1986), Kostov and Dubrule (1986), they developed a method for constrained point-to-point Kriging prediction on the analogy of dual Kriging formalism with splines. In essence, the algorithm replaces the strongest inequality data, which violate some inequality constraints when the initial prediction using only areal data is evaluated against bound values, so that all the other constraints are automatically satisfied (Kostov and Dubrule, 1986). This is very similar to one of the widely used non-linear optimization techniques, termed “active set methods” (Chilès and Delfiner, 1999).
In this section, we generalize Kostov and Dubrule’s approach to area-to-point Kriging, which minimizes the clamping effect of prediction and handles various inequality constraints, while taking into consideration support differences between available data and sought-after predictions. To illustrate this method, consider the same task of predicting the attribute value at location $\mathbf{u}_p$ with a constant mean, m_Z, over the study domain A. The prediction is written as in Equation (16) by substituting the coefficient a_0 with the constant mean m_Z. Without considering inequality constraints, the coefficients of dual simple Kriging (DSK) interpolator are derived by solving the following system:

$$\mathbf{C}_D \boldsymbol{\omega} = \mathbf{r}_D$$

where $\mathbf{r}_D$ denotes a $(K \times 1)$ vector of the residuals of the areal data from their mean, written as $\mathbf{r}_D = (\mathbf{d} - \mathbf{m}_D)$. In the case of constrained DSK, however, the coefficients can be chosen to minimize the following penalty function subject to both equality and inequality constraints, based on the analogy with the spline formalism. Consider K equality constraints and Q inequality constraints consisting of Q_l lower bounds and Q_u upper bounds, where the quadratic minimizer is written as:

$$\text{Min} \ \frac{1}{2} \left[\sum_{k=1}^{K} \sum_{k'=1}^{K} \omega(v_k)\omega(v_{k'}) C_D(v_k, v_{k'}) + \sum_{k=K+1}^{K+Q} \sum_{k'=K+1}^{K+Q} \omega(\mathbf{u}_k)\omega(\mathbf{u}_{k'}) C_Z(\mathbf{u}_k, \mathbf{u}_{k'}) \right] \tag{23}$$

subject to

$$
\begin{aligned}
\sum_{p=1}^{P_k} g_k(\mathbf{u}_p)\hat{z}(\mathbf{u}_p) &= d(v_k), \quad k = 1, \ldots, K \\
\hat{z}(\mathbf{u}_k) &\geq z_l, \qquad k = (K+1), \ldots, (K+Q_l) \\
\hat{z}(\mathbf{u}_k) &\leq z_u, \qquad k = (K+Q_l+1), \ldots, (K+Q)
\end{aligned}
$$

In the constrained area-to-point prediction, the quadratic minimizer in Equation (22) is extended to the form of Equation (23) to take into account support differences by distinguishing the K DSK weights for the areal data from the Q DSK weights for the inequality constraints, respectively. The same concept applies to the constraints, where the K equality constraints pertain to the reproduction of areal data, whereas the Q inequality constraints pertain to point support predictions directly.

In matrix notation, the objective function of Equation (23) is rewritten as:

$$
\begin{aligned}
& \text{Min } \frac{1}{2}\boldsymbol{\omega}_+' \mathbf{C}_+ \boldsymbol{\omega}_+ \\
= & \text{ Min } \frac{1}{2} [\boldsymbol{\omega}_D' \quad \boldsymbol{\omega}_{Z_l}' \quad \boldsymbol{\omega}_{Z_u}'] \begin{bmatrix} \mathbf{C}_D & \mathbf{C}_{DZ_l} & \mathbf{C}_{DZ_u} \\ \mathbf{C}_{Z_lD} & \mathbf{C}_{Z_l} & \mathbf{C}_{Z_lZ_u} \\ \mathbf{C}_{Z_uD} & \mathbf{C}_{Z_uZ_l} & \mathbf{C}_{Z_u} \end{bmatrix} \begin{bmatrix} \boldsymbol{\omega}_D \\ \boldsymbol{\omega}_{Z_l} \\ \boldsymbol{\omega}_{Z_u} \end{bmatrix}
\end{aligned} \tag{24}
$$

subject to

(a) equality condition

$$
\mathbf{G} \left[\; \mathbf{C}_{ZD} \; \mathbf{C}_{ZZ_l} \; \mathbf{C}_{ZZ_u} \; \right] \begin{bmatrix} \boldsymbol{\omega}_D \\ \boldsymbol{\omega}_{Z_l} \\ \boldsymbol{\omega}_{Z_u} \end{bmatrix} = \left[\; \mathbf{C}_D \; \mathbf{C}_{DZ_l} \; \mathbf{C}_{DZ_u} \; \right] \begin{bmatrix} \boldsymbol{\omega}_D \\ \boldsymbol{\omega}_{Z_l} \\ \boldsymbol{\omega}_{Z_u} \end{bmatrix} = \mathbf{r}_D \tag{25}
$$

(b) inequality condition

$$
\begin{bmatrix} -\mathbf{C}_{Z_lD} & -\mathbf{C}_{Z_l} & -\mathbf{C}_{Z_lZ_u} \\ \mathbf{C}_{Z_uD} & \mathbf{C}_{Z_uZ_l} & \mathbf{C}_{Z_u} \end{bmatrix} \begin{bmatrix} \boldsymbol{\omega}_D \\ \boldsymbol{\omega}_{Z_l} \\ \boldsymbol{\omega}_{Z_u} \end{bmatrix} \leq \begin{bmatrix} \mathbf{r}_l \\ \mathbf{r}_u \end{bmatrix} \tag{26}
$$

where the $(K+Q) \times 1$ vector of unknown coefficients, denoted as $\boldsymbol{\omega}_+$, consists of the weights for the K areal data and the weights for the Q inequality constraints, and is written as: $\boldsymbol{\omega}_+ = [\boldsymbol{\omega}_D' \quad \boldsymbol{\omega}_{Z_l}' \quad \boldsymbol{\omega}_{Z_u}']'$. The $(K+Q) \times (K+Q)$ covariance matrix, denoted as $\mathbf{C}_+$, is a matrix consisting of the areal data covariances $\mathbf{C}_D$, and cross-covariances between any areal datum and the lower and upper bound inequality constraints, denoted as, $\mathbf{C}_{DZ_l}$ and $\mathbf{C}_{DZ_u}$, respectively. Note that the quadratic minimization problem in Equation (23) has a unique solution only if $\mathbf{C}_+$ is symmetric, and positive-definite.

Recall that the areal data are equivalent to the convolution of unknown (or predicted) values via a sampling function $g_k(\mathbf{u}_p)$, which is noted by the multiplication of the $(K \times P)$ matrix of sampling functions $\mathbf{G}$ and the $(P \times 1)$ vector of DSK predictions $\hat{\mathbf{z}}$ in matrix notation: $\mathbf{d} = \mathbf{G}\hat{\mathbf{z}}$. The exactitude property of area-to-point Kriging is treated as K equality constraints as shown in

Equation (25). Alternatively, this relation can be represented by the weighted combination of cross-covariances between the areal data and all constraints $[\mathbf{C}_D\ \mathbf{C}_{DZ_l}\ \mathbf{C}_{DZ_u}]$, based upon the associativity of matrix multiplication, i.e. $\mathbf{GC}_{ZD} = \mathbf{C}_D,\ \mathbf{GC}_{ZZ_l} = \mathbf{C}_{DZ_l},\ \mathbf{GC}_{ZZ_u} = \mathbf{C}_{DZ_u}$. Inequality constraints can be summarized in a matrix, as shown in Equation (26), by alternating the sign of Q_l lower bound and Q_u upper bound constraints.

In terms of representation, on the other hand, the constrained area-to-point prediction is considered as a special case of linearly constrained optimization problems, in particular, "Quadratic Programming (QP) problems", which occurs when the objective function is the quadratic form of a constant vector and a constant symmetric matrix (Gill, Murray and Wright, 1981). For efficient computation, we can apply the concept of duality to quadratic problems, since it is known that if a solution to either the primal or dual form of an optimization problem exists and is finite, then a solution to the other problem also exists (Dorn, 1960). These relationships prove to be extremely useful in a variety of ways. For example, the restatement of the primary form of Equation (24) into the dual form can be directly solved to identify an optimal solution to the primary problem, because it is the number of constraints rather than the number of variables of the QP that affect the computational effort (Hillier and Lieberman, 2001). The dual form of the objective function in Equation (23) is reexpressed as follows (see Appendix III for a detailed derivation):

$$\text{Min}\left[\sum_{k=1}^{K}\sum_{k'=1}^{K}[\omega(v_k)\omega(v_{k'})C_D(v_k,v_{k'}) - r_D(v_k)\omega(v_k)] + \sum_{k=K+1}^{K+Q}\sum_{k'=K+1}^{K+Q}[\omega(\mathbf{u}_k)\omega(\mathbf{u}_{k'})C_Z(\mathbf{u}_k,\mathbf{u}_{k'}) - r(\mathbf{u}_k)\omega(\mathbf{u}_k)]\right] \tag{27}$$

subject to

$$\begin{aligned} \omega(\mathbf{u}_k) &\geq 0, \quad k=(K+1),\ldots,(K+Q_l) \\ \omega(\mathbf{u}_k) &\leq 0, \quad k=(K+Q_l+1),\ldots,(K+Q_l+Q_u) \end{aligned}$$

where the k-th equality constraint applies to the k-th areal datum residual, written as $r_D(v_k) = d(v_k) - m_D(v_k)$, whereas the q-th inequality constraint applies to the location $\mathbf{u}_k$, where $\{r(\mathbf{u}_q) = z_l - m_Z,\ k=(K+1),\ldots,(K+Q_l)\}$ for lower bounds and $\{r(\mathbf{u}_k) = z_u - m_Z,\ k = (K+Q_l+1),\ldots,(K+Q)\}$ for upper bounds. Note that this simple conversion of the primary form of constrained area-to-point prediction problem into the dual form reduces the number of constraints from $(K+Q)$ to Q. In matrix notation, the dual form of the minimization problem can be simplified in terms of constraints, which is written as:

$$\text{Min } \boldsymbol{\omega}_+'\mathbf{C}_+\boldsymbol{\omega}_+ - \mathbf{r}_t'\boldsymbol{\omega}_+ \tag{28}$$

$$= [\boldsymbol{\omega}_D'\ \boldsymbol{\omega}_{Z_l}'\ \boldsymbol{\omega}_{Z_u}'] \begin{bmatrix} \mathbf{C}_D & \mathbf{C}_{DZ_l} & \mathbf{C}_{DZ_u} \\ \mathbf{C}_{Z_lD} & \mathbf{C}_{Z_l} & \mathbf{C}_{Z_lZ_u} \\ \mathbf{C}_{Z_uD} & \mathbf{C}_{Z_uZ_l} & \mathbf{C}_{Z_u} \end{bmatrix} \begin{bmatrix} \boldsymbol{\omega}_D \\ \boldsymbol{\omega}_{Z_l} \\ \boldsymbol{\omega}_{Z_u} \end{bmatrix} - [\mathbf{r}_D'\ \mathbf{r}_l'\ \mathbf{r}_u'] \begin{bmatrix} \boldsymbol{\omega}_D \\ \boldsymbol{\omega}_{Z_l} \\ \boldsymbol{\omega}_{Z_u} \end{bmatrix} \tag{29}$$

subject to

$$\begin{bmatrix} -\boldsymbol{\omega}_{Z_l} \\ \boldsymbol{\omega}_{Z_u} \end{bmatrix} \leq \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

where the bound values for equality and inequality constraints are $\mathbf{r}_t = [\mathbf{d}' \; \mathbf{z}_1' \; \mathbf{z}_2']' - m_Z$.
Although a number of interesting optimization problems can be resolved analytically, as shown in the BOK system using the Kuhn-Tucker theorem, almost all major algorithms for QP problems are iterative methods, such as "primal-dual active set methods" (Luenberger, 1969). Active set methods, or "projection methods", are very common for handling constraints via a descent method, in which the direction of descent is chosen to decrease the cost functional and to remain within the constraint region. These methods find an initial feasible solution by first solving a linear programming problem. If the initial guess is not feasible, a new active set is determined, and the process is iterated until the optimal solution is obtained. Therefore, in the application of constrained area-to-point prediction, these algorithms pick a subset of inequality constraints which satisfies all constraints automatically (Gill et al., 1981).
Based on the understanding of active set methods, we derive the associated constrained area-to-point prediction error variance. Recall that the algorithms for QP problems pick a subset of inequality constraints, while automatically satisfying the entire set of inequality constraints. The selected inequality constraint points are considered as data for the second phase of the area-to-point prediction, while the other inequality constraints do not affect prediction. If we can distinguish which inequality constraints are picked, the prediction error variance can be obtained using the areal data and the subset of inequality constraints chosen by the QP solution, since the variance is a sole function of support configuration. Let $\{\mathbf{u}_k, \; k = 1, \ldots, S\}$ be a subset of inequality constraints chosen from the QP solution ($S \leq Q$). Then, the prediction error variance $\sigma^2(\mathbf{u}_p)$ at location $\mathbf{u}_p$ is written as:

$$\sigma^2(\mathbf{u}_p) = C_Z(\mathbf{0}) - [\mathbf{c}_{DZ}^p]' \mathbf{C}_D^{-1} \mathbf{c}_{DZ}^p - [\mathbf{c}_{SZ}^p]' \mathbf{C}_S^{-1} \mathbf{c}_{SZ}^p \tag{30}$$

where $\mathbf{C}_S$ is a $(S \times S)$ matrix of covariances among the subset of inequality constraint points, and $\mathbf{c}_{SZ}^p$ denotes $(S \times 1)$ vector of cross-covariances between the subset of inequality constraints and the prediction location $\mathbf{u}_p$.
Consider the case where global constraints are imposed at all the prediction locations over the study domain A, due to the very definition of the variable; for example, population density or mineral grades, are positive over the entire study region. Unfortunately, it is very difficult to handle global constraints in area-to-point prediction, because the QP solution should be sought after for all the areal data and the entire set of prediction locations, which requires a large amount of computation. Therefore, we can conclude that global constraints can be imposed almost everywhere in theory, but not everywhere in practice. As an alternative, we propose a two phase approach to handle global constraints, whereby inequality constraints are enforced only to those point locations for which the initial predictions do not honor the lower or upper bound constraints. The two stage prediction is performed through the following three steps: (i) initial area-to-point Kriging prediction using only areal data, (ii) predicted point values are evaluated against bound values; if there exists any areal support in which any point prediction violates inequality constraints, all the points inside the selected areal support are placed to the subset of inequality constraints, (iii) a QP problem subject to equality and inequality constraints is sought after to get the DSK weights; if a solution exists, constrained area-to-point prediction is performed, and the predicted values of all the points inside the violating areal support are updated by new predictions. Otherwise, the inequality constraints can not be imposed to the current configuration, and the initial unconstrained prediction is returned.

3 Case study

In the case study, we demonstrate the application of various local and global constraints to synthetic data sets. Local constraints mixed with upper, lower, and two-sided interval constraints are assigned to a line transect data set (see Dubrule and Kostov (1986)), whereas global constraints, i.e., non-negativity, are enforced to entire simulated 2D surface. In both cases, we suppose that areal data are averages of the unknown point support values in any support.

3.1 Area-to-point prediction subject to local constraints

Consider a one-dimensional data set with 9 area-averaged data $\{d(v_k), k = 1, \ldots, 9\}$ with support size $|v_k| = 4$, which are assumed to be the convolution results of point values via a simple normalized indicator function ($g_k(\mathbf{u}_p) = 1/|v_k|$). From now on, the 9 areal data are treated as equality constraints, whereas 12 lower bounds and 6 upper bounds are considered as inequality constraints. Also, note that two points at locations $\{\mathbf{u} = 8, 12\}$ have two-sided inequality constraints. The expected value of point support is assumed to be 5.0556 and, a generalized Cauchy model, which is very regular near the origin, is used for the point support semivariogram:

$$\gamma_Z(\mathbf{h}) = \gamma_Z(|\mathbf{h}|) = C_Z(\mathbf{0}) \left[1 - (1 + (\frac{|\mathbf{h}|}{4})^2)^{-1} \right] \tag{31}$$

with variance $C_Z(\mathbf{0}) = 1$, and practical range (distance at which 95% of the model sill is reached) 4 distance units.

In the case study of local constraints, 4 different area-to-point prediction results are compared: (i) unconstrained prediction using only areal data, (ii) unconstrained prediction with areal data and a subset of inequality constraints whose initial predictions violate inequality constraints, (iii) bounded simple Kriging (BSK) prediction (see section 2.1.2), (iv) constrained DSK prediction via the QP solution (see section 2.2.3).

The unconstrained area-to-point prediction using only areal data is shown in Figure 1A, and the associated prediction error variance in Figure 1B. The initial unconstrained predictions violate 11 inequality constraints including one interval constraint at location $\mathbf{u} = 8$. The prediction error variances in Figure 1B show the minimum variance at the center of areal supports, and the maximum error variance at the point location $\mathbf{u} = 16$ whose distance is far from the areal data. In Figure 1C, unconstrained area-to-point predictions using the 9 areal data and the 11 violated inequality constraints are shown, where the violated inequality constraints are replaced by their bound values, and they are treated as known data in the second stage of prediction. In the case of two-sided interval constraints, for example, at location $\mathbf{u} = 8$, an arbitrary decision is made to replace the inequality constraint by one of the bound values. By treating inequality constraints as known data, the overall prediction error variances in Figure 1D are significantly reduced, and the predictions look realistic. However, note that the two stage unconstrained area-to-point predictions in Figure 1C no longer satisfy some inequality constraints whose initial prediction in Figure 1A was met, e.g., the two-sided interval constraints at location $\mathbf{u} = 12$, lower bounds at $\mathbf{u} = 16, 18, 46$, and upper bounds at $\mathbf{u} = 36$. As an alternative approach to the two stage method, BSK is used and the resulting predictions are shown in Figure 1E. The BSK predictions are almost identical to the unconstrained predictions in Figure 1A except the locations where inequality constraints are not met in the unconstrained prediction. The Kuhn-Tucker conditions are applied only to this subset of inequality constraints, by binding the violated initial predictions to bound values. The associated BSK variance is also shown in Figure 1F, where the variance jumps only at locations where the initial prediction violates inequality constraints. As a result, all the inequality constraints are satisfied and the uncertainties of prediction are reported. In the BSK variance, locations where bound values are fixed to the BSK predictions get increased prediction error variance than the unconstrained variance as a sort of penalty for assuming the bound value as known data. However, it should be noted that this approach does not guarantee the reproduction of areal data, since the BSK approach is based on "the point by point formalism" of Kriging, there is no compensation for the change of initial prediction in the point-to-point formalism. Since other discretized points inside the areal data are not affected by the change of predictions to bounds. In addition, the abrupt changes or discontinuities in both prediction and prediction error variance are unavoidable in the BSK approach. Figure 1G and H shows the constrained area-to-point predictions and associated prediction error variances obtained via the QP solution. The predictions in Figure 1G honor both equality and inequality constraints, while providing a smooth profile. As we can see from the prediction result in Figure 1G, only a subset of inequality constraints rather than the entire inequality constraints, which is strong enough to meet all other constraints automatically must be picked from the QP algorithm. For the application of local constraints, the solution needs to be obtained only once subject to both equality and inequality constraints, although the prediction variances need to be computed separately after the QP solutions are obtained.

3.2 Area-to-point prediction subject to global constraints

Global constraints, such as "non-negativity" of population density surface or "sum to a constant" of data expressed as fractions or percentages, are often enforced due to the very definition of variables. In the following section, we consider area-to-point predictions subject to the non-negativity global constraints imposed over all the prediction locations in the study area A. The areal data are defined on a (7x7) grid with cell size (3x3), and they are derived from the simulated point support reference data whose expected value is 1, and the following generalized Cauchy point semivariogram model with anisotropy is used for the spatial structure:

$$\gamma_Z(\mathbf{h}) = C_Z(\mathbf{0}) \left[1 - (1 + (\frac{|h|_{45}}{3})^2 + (\frac{|h|_{135}}{5})^2)^{-0.5} \right] \tag{32}$$

with variance $C_Z(\mathbf{0})$, range $\gamma_{45} = 3$ distance units along azimuth 45°, and $\gamma_{135} = 5$ distance units along 135°.

The areal data are derived from the simulated point reference values via a simple normalization indicator function, and the values are in the range between 0 and 3. Figure 2A shows the minimum values of areal data are clustered in the northern east of the study area. The unconstrained predictions at locations $\{\mathbf{u}_p, p = 1, \ldots, 441\}$ using the 49 areal data are shown in Figure 2B, which contains 13 negative values. When the BSK approach is applied to the global constrained area-to-point prediction, we select areal data whose discretized points have at least one negative predicted values from unconstrained area-to-point prediction. The Kuhn-Tucker conditions are applied to all the discretized points by replacing their predictions by bound values (see Figure 2C), whereas the prediction is identical to the initial point prediction where the initial prediction meets between inequality constraints. The BSK prediction error variance, which increases proportional to the difference between the bound value and the initial prediction, where inequality constraints are violated, are shown in Figure 2D. Slight increase in the northern east of study area are noticeable. In 2D example, the clamping effect is more evident, which prevents the use of BSK approach from general mapping applications. Also, it is easily inferred that the areal support data where Kuhn-Tucker conditions are applied do not reproduce the areal data anymore by the correction. Figure 2E and 2F shows the results when a two phase approach is applied to handle global inequality constraint via a constrained DSK. The unconstrained area-to-point prediction is evaluated against global non-negativity constraints, and determine the areal supports which contain more than one negative point prediction. In the second phase, the QP problem is constructed using the 49 areal data and the discretized point supports which belong to the selected areal supports. In the formalism of the QP problem, the support differences are explicitly accounted for, and the optimal solution yields the DSK weights as well as identifies the subset of inequality constraints which are active, if a solution exists. The results are shown in Figure 2E. The associated variance are obtained by the areal supports and the active inequality point supports, which were identified by the QP solution.

4 Conclusion

"Inequality-type data", or "inequality constraints", often provide useful information to refine area-to-point prediction. Due to the incompatible properties of measurements and inequality-type data, special care should be taken to incorporate inequality constraints into area-to-point prediction. In this paper, we have adopted two geostatistical methods to deal with inequality constraints within the context of area-to-point prediction. In both approaches, constraints are imposed on the predictions rather than on the Kriging weights, but the approaches differ in terms of the choice of Kriging formalism. The first approach, called "bounded Kriging", is based on non-linear optimization solution, whereby Kriging is solved by minimizing prediction error variance, whose objective function is represented via a non-linear form. The second approach is developed based on the analogy of dual Kriging formalism with spline. Bounded Kriging is intuitive and easy to implement, although abrupt discontinuities to locations where unconstrained predictions violate inequality constraints are unavoidable. In addition, the critical drawback in the application of bounded Kriging to area-to-point prediction is that it does not reproduce the areal data. Hence, this approach is not recommended for the task of generating a smooth prediction surface nor general area-to-point prediction problems, where the coherence property of predictions is important. We also pointed out that this approach is limited only to the case where one-side inequalities are imposed, whereas the other approach is more flexible in many aspects. It is well-known that dual formalism of Kriging allows connecting Kriging to splines or other deterministic interpolation functions so that constrained thin-plate splines can be generalized to constrained area-to-point prediction. Based on the fact that the objective function of constrained splines or constrained area-to-point prediction takes the quadratic form subject to equality and inequality constraints, we demonstrated that QP algorithms resolve constrained area-to-point prediction in some cases. In contrast to other approaches, the constrained area-to-point prediction approach takes into account support differences explicitly and produces a smooth surface which

satisfies inequality constraints at the same time. The assessment of the uncertainty in prediction is provided via prediction error variances, which were derived from “active set methods”.
In the case studies, we demonstrated the application of local constraints where general constraints on minimum or maximum bounds, including two-sided interval constraints, are forced to some locations. We also demonstrated how to derive point predictions when global constraints, in particular, non-negativity of the predicted surface, are imposed to synthetic 2D data. In both cases, we compare the resulting predicted point values using several criteria, such as the consideration of support differences, the smoothness of the predicted surface, and the satisfaction of inequality constraints. In summary, the constrained area-to-point approach via the QP algorithm is the most flexible and precise method to handle inequality constraints, although there are potential problems of clamping effects, the existence of feasible solutions, and computing capacity. In particular, when global constraints are imposed, more careful consideration for large data sets is required. In this paper, we propose a selective application of the QP algorithm, which satisfies all the inequality constraints, while reducing computing effort. Furthermore, the feasible solution is not always guaranteed to be obtained, since the QP solution is sensitive to the choice of the point support covariance model.
In future work, we will focus on: (i) exploring alternative methods to relax the limitation of the finite number of constraints, i.e., a two-phase approach, (ii) developing more efficient computational algorithms to characterize uncertainty of predictions, (iii) accounting for non-Gaussian data, (iv) considering mixed constraints with several variables, and (v) applying the proposed methods to real data sets.

5 Acknowledgements

Financial support for this study was provided by the National Science Foundation Grant BCS ♯ 0352879.

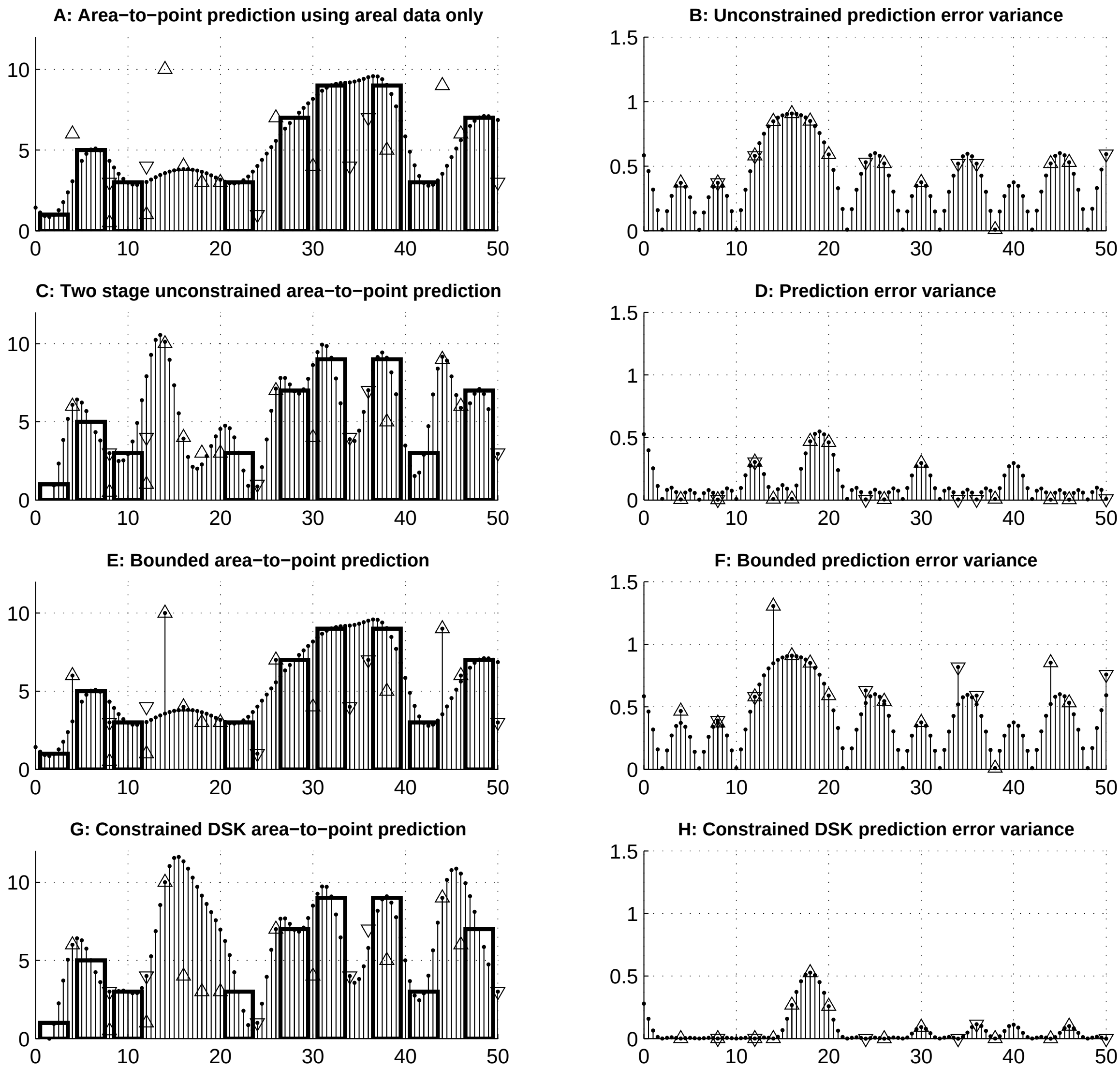

Figure 1: **A:** Unconstrained area-to-point prediction, where △ denotes a lower bound and ∇ denotes an upper bound. The width of rectangle represents the support of areal data, whereas their height represents their attribute value. **B:** Prediction error variance of initial unconstrained predictions. **C:** Two stage area-to-point predictions using areal data and a subset of inequality data whose initial predictions violate inequality constraints. **D:** Two stage area-to-point prediction error variance. **E:** Bounded simple Kriging prediction. **F:** Bounded simple Kriging prediction error variances. **G:** Constrained area-to-point DSK prediction. **H:** Constrained area-to-point DSK prediction error variances.

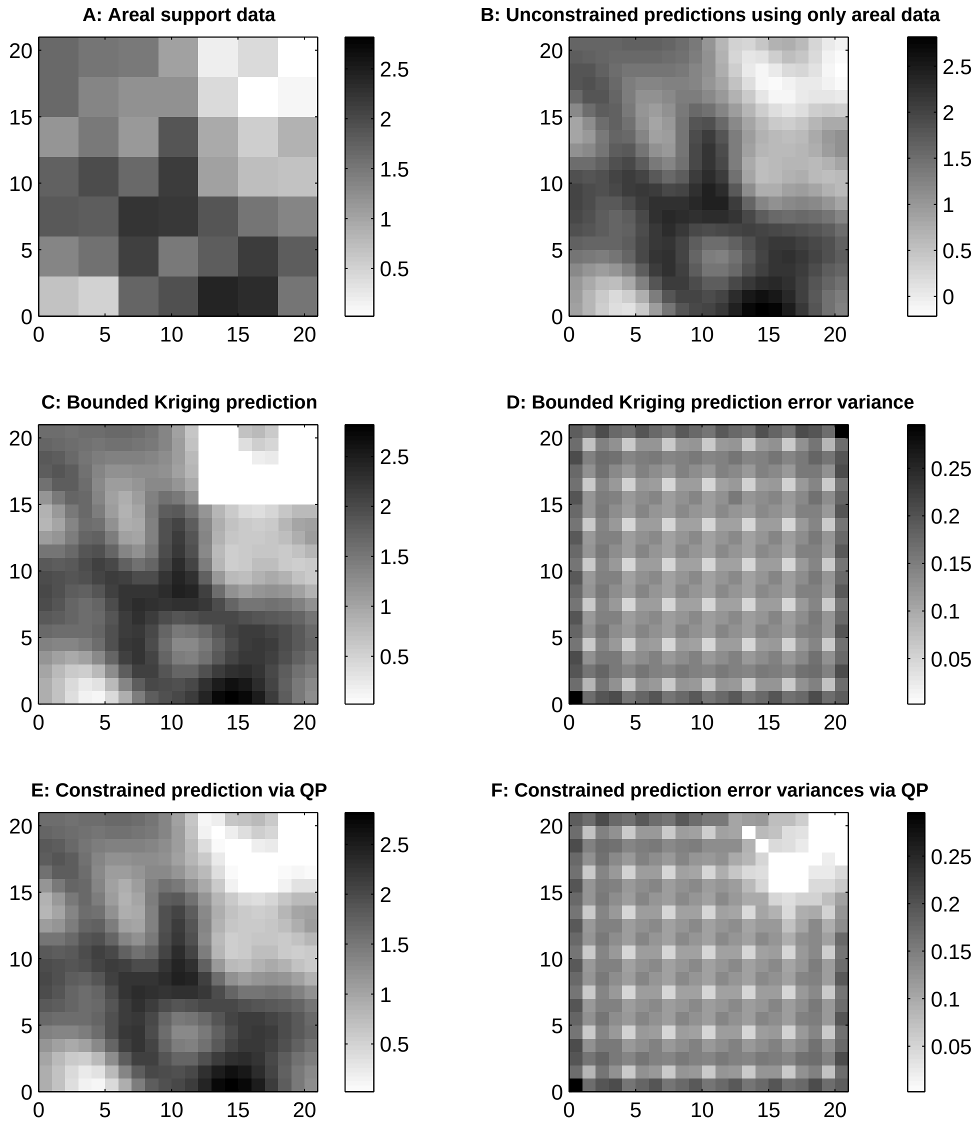

Figure 2: **A:** Synthetic areal support data. **B:** Unconstrained area-to-point predictions using only areal data. **C:** Bounded simple Kriging prediction. **D:** Bounded simple Kriging prediction error variance. **E:** Two-phase constrained DSK area-to-point prediction. **F:** Constrained DSK area-to-point prediction error variance

APPENDIX I. Derivation of bounded ordinary Kriging (BOK) variance

Let $\mathbf{C}$ be an $(K+2)\times(K+2)$ matrix that is partitioned as follows:

$$\mathbf{C} = \left[\begin{array}{cc|c} \mathbf{C}_D & \mathbf{1} & \mathbf{d} \\ \mathbf{1}' & 0 & 0 \\ \hline \mathbf{d}' & 0 & 0 \end{array}\right] = \left[\begin{array}{cc} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{array}\right] \tag{33}$$

where $\mathbf{C}_{11} = \left[\begin{array}{cc} \mathbf{C}_D & \mathbf{1} \\ \mathbf{1}' & 0 \end{array}\right]$, $\mathbf{C}_{12} = \left[\begin{array}{c} \mathbf{d} \\ 0 \end{array}\right]$, $\mathbf{C}_{21} = [\mathbf{d}'\ 0]$, and $\mathbf{C}_{22} = [0]$.

We assume that the inverse of $\mathbf{C}$ exists, denoted as $\mathbf{C}^{-1} = \mathbf{B} = \left[\begin{array}{cc} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{array}\right]$.

Let $\mathbf{K}$ be the $(K+2)\times 1$ vector, consisting of the covariance between areal data and unknowns, the sum of Lagrange multipliers, and the lower bound value, which can be partitioned as:

$$\mathbf{V} = \left[\begin{array}{c} \mathbf{c}_{DZ}^{p} \\ 1 \\ \hline z_l \end{array}\right] = \left[\begin{array}{c} \mathbf{V}_0 \\ z_l \end{array}\right] \tag{34}$$

Then, the BOK variance in Equation (13) is rewritten as

$$\hat{\sigma}^2_{BOK}(\mathbf{u}_p) = C_Z(\mathbf{0}) - [\mathbf{V}_0'\ z_l]\left[\begin{array}{cc} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{array}\right]\left[\begin{array}{c} \mathbf{V}_0 \\ z_l \end{array}\right] = C_Z(\mathbf{0}) - [\mathbf{V}_0'\mathbf{B}_{11}\mathbf{V}_0 + z_l\mathbf{B}_{21}\mathbf{V}_0 + \mathbf{V}_0'\mathbf{B}_{12}z_l + z_l\mathbf{B}_{22}z_l] \tag{35}$$

According to the *Partitioned Matrix Lemma* (Meyer, 2000), the inverse matrix of $\mathbf{C}$, denoted as $\mathbf{B}$, can be summarized as follows, since $\mathbf{C}_{11}$ and $\mathbf{B}_{22}$ are both nonsingular:

- $\mathbf{B}_{22} = [\mathbf{C}_{22} - \mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{12}]^{-1} = [0 - \mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{12}]^{-1} = -[\mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{12}]^{-1}$
- $\mathbf{B}_{12} = -\mathbf{C}_{11}^{-1}\mathbf{C}_{12}[\mathbf{C}_{22} - \mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{12}]^{-1} = -\mathbf{C}_{11}^{-1}\mathbf{C}_{12}\mathbf{B}_{22}$
- $\mathbf{B}_{21} = -\mathbf{B}_{22}\mathbf{C}_{21}\mathbf{C}_{11}^{-1}$
- $\mathbf{B}_{11} = \mathbf{C}_{11}^{-1} + \mathbf{C}_{11}^{-1}\mathbf{C}_{12}\mathbf{B}_{22}\mathbf{C}_{21}\mathbf{C}_{11}^{-1}$

Let $\eta = [\mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{12}]$, and recall that predicted value at the location $\mathbf{u}_p$ is represented by $\hat{z}(\mathbf{u}_p) = \mathbf{V}_0'\mathbf{C}_{11}^{-1}\mathbf{C}_{12}$. The BOK variance in Equation (35) is rewritten as:

$$\begin{aligned}
\hat{\sigma}^2_{BOK}(\mathbf{u}_p) &= C_Z(\mathbf{0}) - \mathbf{V}_0'\mathbf{B}_{11}^{-1}\mathbf{V}_0 + z_l\mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{V}_0\eta^{-1} + \mathbf{V}_0'\mathbf{C}_{11}^{-1}\mathbf{C}_{12}z_l\eta^{-1} - z_l^2\eta^{-1} \\
&= C_Z(\mathbf{0}) - \mathbf{V}_0'\mathbf{B}_{11}^{-1}\mathbf{V}_0 + [2z_l\hat{z}(\mathbf{u}_p) - z_l^2]\eta^{-1} \\
&= C_Z(\mathbf{0}) - \mathbf{V}_0'\mathbf{C}_{11}^{-1}\mathbf{V}_0 + [\mathbf{V}_0'\{\mathbf{C}_{11}^{-1}\mathbf{C}_{12}\mathbf{C}_{21}\mathbf{C}_{11}^{-1}\}\mathbf{V}_0 + 2z_l\hat{z}(\mathbf{u}_p) - z_l^2]\eta^{-1} \\
&= \hat{\sigma}^2_{OK}(\mathbf{u}_p) + [\hat{z}(\mathbf{u}_p)^2 - 2z_l\hat{z}(\mathbf{u}_p) + z_l^2]\eta^{-1} = \hat{\sigma}^2_{OK}(\mathbf{u}_p) + \{\hat{z}(\mathbf{u}_p) - z_l\}^2\eta^{-1}
\end{aligned}$$

where $\hat{\sigma}^2_{OK}(\mathbf{u}_p)$ denotes the unconstrained prediction error variance at location $\mathbf{u}_p$.

APPENDIX II. Dual Kriging with a trend model

Recall the general form of dual Kriging, which is represented by a linear combination of the cross-covariances between prediction locations and areal data, denoted as $\mathbf{c}_{ZD}$, and the drift functionals $\mathbf{f}_m^p$. The point prediction at location $\mathbf{u}_p$ using K areal data is written as:

$$\begin{aligned} \hat{z}(\mathbf{u}_p) &= \sum_{k=1}^{K} \omega(v_k) C_{DZ}(v_k, \mathbf{u}_p) + \sum_{m=0}^{M} a_m f_m(\mathbf{u}_p) \\ &= [\boldsymbol{\omega}' \quad \mathbf{a}'_m][\begin{array}{c} \mathbf{c}_{DZ}^p \\ \mathbf{f}_m^p \end{array}] = \boldsymbol{\omega}' \mathbf{c}_{DZ}^p + \mathbf{a}'_m \mathbf{f}_m^p, \quad \mathbf{u}_p \in A \end{aligned} \tag{36}$$

with

$$\sum_{k=1}^{K} \omega(v_k) f_m(\mathbf{u}_k) = 0, \qquad m = 0, \ldots, M$$

where $\boldsymbol{\omega}$ denotes a $(K \times 1)$ vector of the coefficients of the stochastic interpolator, whereas $\mathbf{a}_m$ is a $(M+1) \times 1$ vector of trend coefficients. Note that OK is a special case of Kriging with a trend, where the drift function is a constant as 1, i.e., $f_0^p = 1$.

$$\hat{z}(\mathbf{u}_p) = [\boldsymbol{\omega}' \quad a'_0][\begin{array}{c} \mathbf{c}_{DZ}^p \\ 1 \end{array}] = \boldsymbol{\omega}' \mathbf{c}_{DZ}^p + a_0 \tag{37}$$

In the perspective of interpolation problems, dual Kriging with a trend model can be viewed as the following problem: “find the coefficients $\boldsymbol{\omega}$ and $\mathbf{a}_m$ of the linear combination of $\mathbf{c}_{DZ}^p$ and $\mathbf{f}_m^p$ in Equation (36), satisfying the K data identification and $(M+1)$ unbiasedness constraints”, on which are written as:

$$\begin{aligned} \sum_{k=1}^{K} g_k(\mathbf{u}_p) \hat{z}(\mathbf{u}_p) &= d(v_k), \quad k = 1, \ldots, K \\ \sum_{k=1}^{K} \omega(v_k) f_m(\mathbf{u}_k) &= 0, \quad m = 0, \ldots, M \end{aligned}$$

The above weights $\boldsymbol{\omega}$ and coefficients $\mathbf{a}_m$ are derived by solving the following system of normal equations:

$$\begin{bmatrix} \mathbf{C}_D & \mathbf{F} \\ \mathbf{F}' & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\omega} \\ \mathbf{a}_m \end{bmatrix} = \begin{bmatrix} \mathbf{d} \\ \mathbf{0} \end{bmatrix} \tag{38}$$

where $\mathbf{F}$ is the $K \times (M+1)$ matrix, composed of $(M+1)$ low-order drift functionals, denoted as $\mathbf{F} = [\mathbf{f}_0 \quad \ldots \quad \mathbf{f}_M]$, where $\mathbf{f}_m = [f_m(\mathbf{u}_k),\ k = 1, \ldots, K]'$.

APPENDIX III. Kriging as a Quadratic Programming problem

1. A classical QP problem

QP problems involve the optimization of a quadratic objective function, subject to a number of constraints, and can involve either a maximization or a minimization problem. Consider a quadratic function subject to the linear inequality constraints, written as:

$$\begin{aligned} \text{Min } \frac{1}{2}\mathbf{x}'\mathbf{C}\mathbf{x} + \mathbf{f}'\mathbf{x} \\ \mathbf{A}\mathbf{x} \geq \mathbf{b} \end{aligned} \tag{39}$$

where $\mathbf{x}$ is a $(K \times 1)$ vector of coefficients that minimize the objective function subject to the inequality constraints $\mathbf{Ax} \geq \mathbf{b}$, and $\mathbf{C}$ is a $(K \times K)$ symmetric, positive semi-definite matrix. $\mathbf{f}$ is a $(K \times 1)$ vector of coefficients, whereas $\mathbf{A}$ denotes a $(Q_l \times K)$ matrix of Q_l constraints, and $\mathbf{b}$ is a $(Q_l \times 1)$ vector of lower bound values, so that the solution is greater than lower bound values. The consideration of both the dual and primal forms often provide insights about alternative methods to derive a solution to reduce computational burden. The dual form of Equation (39) can be written as:

$$\begin{aligned} \text{Max } -\frac{1}{2}\mathbf{x}'\mathbf{C}\mathbf{x} + \mathbf{b}'\mathbf{t} \\ \mathbf{A}'\mathbf{t} - \mathbf{C}\mathbf{x} = \mathbf{f} \\ \mathbf{t} \geq \mathbf{0} \end{aligned} \tag{40}$$

where $\mathbf{t}$ is a $(Q_l \times 1)$ vector of Lagrange multipliers associated with Q_l inequality constraints. Let's consider a simplified version of the dual QP problem in Equation (40). Note that $\mathbf{x}$ is a linear combination of $\mathbf{t}$, as in $\mathbf{x} = \mathbf{C}^{-\mathbf{1}}(\mathbf{A}'\mathbf{t} - \mathbf{f})$. Simply replacing $\mathbf{x}$ by the function of $\mathbf{t}$, the objective function of the dual QP problem in Equation (40) becomes the maximization problem expressed by:

$$\begin{aligned} \text{Max } -\frac{1}{2}(\mathbf{A}'\mathbf{t} - \mathbf{f})'\mathbf{C}^{-\mathbf{1}}(\mathbf{A}'\mathbf{t} - \mathbf{f}) + \mathbf{b}'\mathbf{t} \\ \mathbf{t} \geq \mathbf{0} \end{aligned}$$

which is equivalent to the following minimization problem:

$$\begin{aligned} \text{Min } \frac{1}{2}(\mathbf{A}'\mathbf{t} - \mathbf{f})'\mathbf{C}^{-\mathbf{1}}(\mathbf{A}'\mathbf{t} - \mathbf{f}) - \mathbf{b}'\mathbf{t} \\ \mathbf{t} \geq \mathbf{0} \end{aligned} \tag{41}$$

The dual form of QP problems with equality constraints only also leads to the the same minimization problem in Equation (41). The only difference between QP problems with equality and with inequality constraints is that there is no positive constraints on the Lagrange multipliers associated with Q_l constraints in the case of equality constraints.

2. Constrained dual simple Kriging (DSK) as a QP problem

In the chapter 2.2.3, we showed the constrained DSK is equivalent to a constrained interpolator, whose coefficients are determined so as to minimize the following quadratic objective function subject to both equality and inequality constraints. To keep the problem simple, consider a task of predicting the value at the location $\mathbf{u}_p$ using K areal data $\mathbf{d} = [d(v_1), \ldots, d(v_K)]'$ under lower bound inequality constraints only. We assume that the mean of RV Z is known and constant as m_Z. By the coherence of prediction and Q_l lower bound inequality constraints for locations $\{\mathbf{u}_q, q = 1, \ldots, Q_l\}$, the QP problem is written as:

$$\begin{array}{ccc} & \text{Min} \quad \frac{1}{2}\boldsymbol{\omega}'_+\mathbf{C}_+\boldsymbol{\omega}_+ & \\ \mathbf{A}_1 \cdot \boldsymbol{\omega}_+ & = & \mathbf{r}_D \\ \mathbf{A}_2 \cdot \boldsymbol{\omega}_+ & \geq & \mathbf{r}_{Z_l}, \end{array} \tag{42}$$

where $\boldsymbol{\omega}_+$ is a $(K+Q_l)\times 1$ vector of weights applied to K equality and Q_l inequality constraints, and $\mathbf{C}_+$ is a $(K+Q_l)\times(K+Q_l)$ matrix of covariance among all the constraints. $\mathbf{A}_1$ denotes the $K\times(K+Q_l)$ matrix of equality constraints, and $\mathbf{r}_D$ denotes the residuals of the areal data from their mean, written as $\mathbf{r}_D = (\mathbf{d} - \mathbf{m}_D)$. The $Q_l \times (K+Q_l)$ matrix of inequality constraints is denoted as $\mathbf{A}_2$, and $\mathbf{r}_{Z_l}$ represents the corresponding the residuals for the lower bounds, denoted as $\mathbf{r}_{Z_l} = (z_l\mathbf{1} - m_Z)$.
The dual form of constrained DSK problem in Equation (42) is written as:

$$\text{Max} \; -\frac{1}{2}\boldsymbol{\omega}'_+\mathbf{C}_+\boldsymbol{\omega}_+ + \mathbf{r}'\boldsymbol{\psi} \tag{43}$$

$$\mathbf{A}'\boldsymbol{\psi} - \mathbf{C}_+\boldsymbol{\omega}_+ = \mathbf{0} \tag{44}$$

$$\boldsymbol{\psi}_2 \geq 0 \tag{45}$$

where the objective function involves a maximization problem with $(K+Q_l)\times 1$ vector of Lagrange multipliers, denoted as $\boldsymbol{\psi} = [\boldsymbol{\psi}'_1 \; \boldsymbol{\psi}'_2]'$, and bound values, denoted as $\mathbf{r} = [\mathbf{r}'_D \; \mathbf{r}'_{Z_l}]'$. The new constraints for the dual form of the quadratic function in Equation (44) combines both equality and inequality constraints in the primary form of (42). The constraints matrix $\mathbf{A}$ is a combination of two sub-matrices $\mathbf{A} = [\mathbf{A}'_\mathbf{1} \; \mathbf{A}'_\mathbf{2}]'$, where $\mathbf{A}_1$ is a $K\times(K+Q_l)$ matrix of equality constraints and $\mathbf{A}_2$ is a $Q_l\times(K+Q_l)$ matrix of inequality constraints. The corresponding Lagrange multipliers for each constraint are denoted by $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$, respectively. Thus, Equation (44) is a combination of the following two equations:

$$\mathbf{A}'_\mathbf{1}\boldsymbol{\psi}_\mathbf{1} - \mathbf{C}_+\boldsymbol{\omega}_+ = \mathbf{0}$$
$$\mathbf{A}'_\mathbf{2}\boldsymbol{\psi}_\mathbf{2} - \mathbf{C}_+\boldsymbol{\omega}_+ = \mathbf{0}$$

As shown in the case of the classical QP problem, the constrained DSK problem can be simplified by the following steps. First, switch the objective function to a maximization problem, and note that $\boldsymbol{\omega}_+$ is a function of $\boldsymbol{\psi}$, written as $\boldsymbol{\omega}_+ = \mathbf{C}_+^{-1}\mathbf{A}'\boldsymbol{\psi}$ from Equation (44). Therefore, the simplified version of the dual form of QP problems for constrained DSK prediction can be reformulated as:

$$\text{Max} \; -\frac{1}{2}[\boldsymbol{\psi}'\mathbf{A}\mathbf{C}_+^{-\mathbf{1}}]\mathbf{C}_+[\mathbf{C}_+^{-\mathbf{1}}\mathbf{A}'\boldsymbol{\psi}] + \mathbf{r}'\boldsymbol{\psi} \tag{46}$$
$$\boldsymbol{\psi}_\mathbf{2} \geq \mathbf{0}$$

Let $\mathbf{Q}^* = (\mathbf{AC}_+^{-1}\mathbf{A}')$, then the simplified dual form of SK maximization problem in Equation (46) can be converted to the minimization problem:

$$\text{Min } \frac{1}{2}\boldsymbol{\psi}'\mathbf{Q}^*\boldsymbol{\psi} - \mathbf{r}'\boldsymbol{\psi} \qquad (47)$$
$$\boldsymbol{\psi}_2 \geq \mathbf{0}$$

where $\mathbf{Q}^*$ is a $(K+Q_l) \times (K+Q_l)$ matrix of covariances involving K area-to-area covariances where equality constraints are assigned, and Q_l point covariances where inequality constraints are imposed.

References

Arbia, G. (1989), *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems*, Kluwer Academic Publisher, Dordrecht, Holland.

Barnes, R. J. and Johnson, T. B. (1984), Positive Kriging, *in* G. Verly et al., eds, '*Geostatistics for Natural Resources Characterization*', Vol. 2, Reidel, Dordrecht, Holland, pp. 231–244.

Barnes, R. J. and You, K. (1992), 'Adding bounds to Kriging', *Mathematical Geology* **24**(2), 171–176.

Chilès, J. P. and Delfiner, P. (1999), *Geostatistics: Modeling Spatial Uncertainty*, John Wiley, New York.

Cressie, N. A. C. (1990), 'Reply to comment on Cressies' Geostatistics', *The American Statistician* **44**(3), 256–258.

Cressie, N. A. C. (1993), *Statistics for Spatial Data*, revised edn, John Wiley & Sons, New York.

Deutsch, C. V. (1996), 'Correcting for negative weights in Ordinary Kriging', *Computers & Geosciences* **22**(7), 765–773.

Dorn, W. (1960), 'Duality in Quadratic Programming', *Quarterly of Applied Mathematics* **18**(2), 152–162.

Dubrule, O. (1983), 'Two Methods with different objectives: Splines and Kriging', *Mathematical Geology* **15**(2), 245–257.

Dubrule, O. and Kostov, C. (1986), 'An interpolation method taking into account inequality constraints: I. Methodology', *Mathematical Geology* **18**(1), 33–51.

Duchon, J. (1977), Splines minimizing rotation-invariant semi-norms in sobolev spaces, *in* 'Constructive Theory of Functions of Several Variables', Springer-Verlag, Berlin, pp. 85–100.

Dyn, N. and Wong, W. H. (1987), 'On the characterization of non-negative volume-matching surface splines', *Journal of Approximation* **51**(1), 1–10.

Galli, S., Murillo, E. and Thomann, J. (1984), Dual Kriging-It's properties and its uses in direct contouring, *in* G. Verly et al., eds, '*Geostatistics for Natural Resources Characterization*', Vol. 2, Reidel, Dordrecht, The Holland, pp. 621–634.

Gill, P., Murray, W. and Wright, M. (1981), *Practical Optimization*, Academic Press, London.

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

Hillier, F. S. and Lieberman, G. J. (2001), *Introduction to Operations Research*, 7th edn, McGraw-Hill, New York, NY.

Hutchinson, M. F. and Gessler, P. E. (1994), 'Splines-More than just a smooth interpolator', *Geoderma* **62**, 45–67.

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

Journel, A. G. (1986), 'Constrained interpolation and qualitative information', *Mathematical Geology* **18**(3), 269–286.

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

Kostov, C. and Dubrule, O. (1986), 'An interpolation method taking into account inequality constraints: II. Practical Approach', *Mathematical Geology* **18**(1), 53–73.

Künzi, H. P., Tzschach, H. G. and Zehnder, C. A. (1971), *Numerical Methods of Mathematical Optimization*, Academic press, Dordrecht/ Boston/ London.

Kyriakidis, P. C. (2004), 'A geostatistical framework for the area-to-point spatial interpolation', *Geographical Analysis* **36**(3), 259–289.

Kyriakidis, P. C. and Yoo, E.-H. (2005), 'Geostatistical prediction and simulation of point values from areal data', *Geographical Analysis* **37**(2), 124–151.

Limic, N. and Mikelic, A. (1984), 'Constrained Kriging using Quadratic Programming', *Mathematical Geology* **16**(4), 423–429.

Luenberger, D. G. (1969), *Optimization by Vector Space Methods*, John Wiley & Sons, New York.

Matheron, G. (1981), Splines and Kriging: Their formal equivalence, *in* D. F. Merrian, ed., '*Down-to-Earth Statistics:Solutions Looking for Geological Problems*', Syracuse University, Geology Contributions, pp. 77–95.

Meyer, C. D. (2000), *Matrix Analysis and Applied Linear Algebra*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia.

Olea, R. A., ed. (1999), *Geostatistics for Engineers and Earth Scientists*, Kluwer Academic Publishers, Dordrecht, Holland.

Prenter, P. M. (1975), *Spline and Variational Methods*, John Wiley & Sons, New York.

Royer, J. J. and Vieira, P. C. (1984), Dual formalism of Kriging, *in* G. Verly et al., eds, '*Geostatistics for Natural Resources Characterization*', Vol. 2, Reidel, Dordrecht, Holland, pp. 691–702.

Szidarovszky, F., Baafi, E. Y. and Kim, Y. C. (1987), 'Kriging without negative weights', *Mathematical Geology* **19**(6), 549–560.

Tobler, W. R. (1979), 'Smooth pychnophylactic interpolation for geographical regions', *Journal of the American Statistical Association* **74**(367), 519–530.

Wackernagel, H. (2003), *Multivariate Geostatistics: An introduction with Applications*, Springer-Verlag, Berlin.

Wahba, G. (1990*a*), 'Comment on Cressies' Geostatistics', *The American Statistician* **44**(3), 255–256.

Wahba, G. (1990*b*), *Spline Models for Observational Data*, CBMS-Regional conference series in applied mathematics, Society for Industrial and Applied Math, Philadelphia, Pennsylvania.

Walvoort, D. J. J. and de Gruijter, J. J. (2001), 'Compositional Kriging: A spatial interpolation method for compositional data', *Mathematical Geology* **33**(8), 951–966.

Wong, W. H. (1980), *An analysis of the volume-matching problem and related topics in smooth density estimation*, Ph.d. dissertation, dept of statistics, university of wisconsin-madison, Wisconsin-Madison.