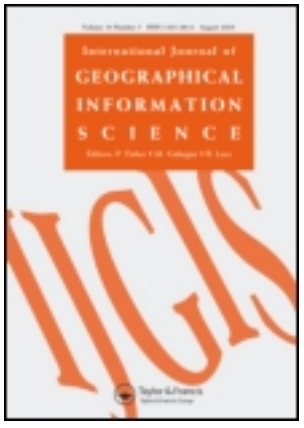

International Journal of Geographical Information Science

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tgis20

On the prediction error variance of three common spatial interpolation schemes

Phaedon C. Kyriakidis [a] & Michael F. Goodchild [a]

[a] Department of Geography, University of California, Santa Barbara, CA 93106, USA

Available online: 20 Feb 2007

To cite this article: Phaedon C. Kyriakidis & Michael F. Goodchild (2006): On the prediction error variance of three common spatial interpolation schemes, International Journal of Geographical Information Science, 20:8, 823-855

To link to this article: http://dx.doi.org/10.1080/13658810600711279

International Journal of Geographical Information Science
Vol. 20, No. 8, September 2006, 823–855

Research Article

On the prediction error variance of three common spatial interpolation schemes

PHAEDON C. KYRIAKIDIS* and MICHAEL F. GOODCHILD
Department of Geography, University of California, Santa Barbara, CA 93106, USA

(*Received 4 June 2005; in final form 23 February 2006*)

Three forms of linear interpolation are routinely implemented in geographical information science, by interpolating between measurements made at the endpoints of a line, the vertices of a triangle, and the vertices of a rectangle (bilinear interpolation). Assuming the linear form of interpolation to be correct, we study the propagation of error when measurement error variances and covariances are known for the samples at the vertices of these geometric objects. We derive prediction error variances associated with interpolated values at generic points in the above objects, as well as expected (average) prediction error variances over random locations in these objects. We also place all the three variants of linear interpolation mentioned above within a geostatistical framework, and illustrate that they can be seen as particular cases of Universal Kriging (UK). We demonstrate that different definitions of measurement error in UK lead to different UK variants that, for particular expected profiles or surfaces (drift models), yield weights and predictions identical with the interpolation methods considered above, but produce fundamentally different (yet equally plausible from a pure data standpoint) prediction error variances.

Keywords: Spatial accuracy assessment; Error propagation; Linear interpolation; Bilinear interpolation; Geostatistics; Trend surface models

1. Introduction

Spatial interpolation can be defined as the prediction of the unknown value of an attribute at some point where no measurement is available, from known measurements obtained at a set of sample locations. Many methods of spatial interpolation are commonly implemented in geographical information systems (GIS), including inverse-distance weighting, spline interpolation, and the various forms of Kriging that are generally lumped under the heading of geostatistics; see for example Ripley (1981) or Cressie (1993) and the references therein. Spatial interpolation occurs both explicitly, as a procedure for estimating values at specified points, and also implicitly in the execution of other standard operations, such as geocoding (the interpolation of street address from known values at block endpoints), image resampling (the estimation of values in a raster from values in another, geometrically incompatible raster), or contouring (the interpolation of isolines among point observations).

It is common for GIS designers to implement comparatively simple and straightforward forms of interpolation, particularly when these are needed by other spatial analysis operations. Such interpolation methods are rapid in execution, and

*Corresponding author. Email: phaedon@geog.ucsb.edu

International Journal of Geographical Information Science
ISSN 1365-8816 print/ISSN 1362-3087 online
http://www.tandf.co.uk/journals
DOI: 10.1080/13658810600711279

involve no explicit decisions on the part of the user. They are thus suited for embedding in procedures such as contour interpolation or resampling, when they can be invoked without further user intervention. In this paper, we focus on one such class of methods, generally known as linear interpolation, and on the effects of measurement error on predictions and associated error variances obtained by such methods.

First, linear interpolation is used in surveying to determine the locations of boundary lines between monuments or endpoints. In such cases, it is common to have estimates of the errors inherent in measurements at endpoints, and we examine the propagation of such errors to intermediate points: what errors result along the boundary line as a result of measurement errors at the endpoints? Linear interpolation is also used in geocoding to determine the locations of specific addresses at intermediate points along blocks, given the locations and addresses of block endpoints; again, we provide estimates of error at specified points along the block.

Secondly, linear interpolation is used in the TIN model to predict elevation at points within triangles, based on values at triangle vertices. We assume knowledge of measurement errors at vertices, and study the effect of measurement error at predictions made at points within each triangle. Finally, bilinear interpolation is commonly used in resampling to determine the value of a variable at some point within a rectangle from values at the rectangle's vertices. Such methods are often used in image processing, when a raster of radiance measurements must be adjusted to some second, geometrically offset raster. Again, we study the propagation of measurement errors from the vertices to points within the rectangle.

In summary, our objective is to derive results concerning the propagation of measurement error in simple interpolation operations in GIS; this paper thus falls within the literature of error propagation (Heuvelink 1998). Assuming that: (i) measurement error variances and covariances are known for sample points, and (ii) the linear form of interpolation is correct, we provide results for prediction error at any interpolated point within line segments, triangles, and rectangles, as well as results for average prediction error over randomly chosen points. Similar results to those presented in this paper, but only for the special cases of exponential and Gaussian covariance models (with or without a nugget-effect component), can be found in papers by Kubik and Botman (1976) and Botman and Kubik (1979). Sections 2, 3 and 5 of the paper consider the line, triangle and rectangle cases respectively, obtaining general results and then simplifying them for the special case of unit variances and zero covariances. As with many problems in geometric probability (Ball and Coxeter 1960), it is possible to approach the three cases from several distinct perspectives: in Section 4, we present an alternative formulation to the triangle problem that yields identical results.

In Section 6, we present the geostatistical formulation of Universal Kriging (UK), whereby spatial prediction accounts for an arbitrary parametric expected profile or surface (drift model), and the number of available samples is generally greater than the number of the parameters characterizing that expected surface. We also derive the predictions and associated error variances for different variants of UK, and illustrate how they are linked to different assumptions regarding the magnitude of measurement error.

In Section 7, we demonstrate that the three linear interpolation methods considered in this work can be regarded as particular cases of UK under appropriate definitions of drift models. More precisely, the weights, and consequently the associated predictions, used in 1D linear interpolation, as well as TIN and bilinear interpolation in 2D, are identical with those derived via the different variants of UK.

The associated prediction error variances, however, differ fundamentally owing to the different definition of measurement error adopted for each variant of UK. It is shown that the prediction error variances reported in previous studies correspond to those obtained by UK of the drift component, but other variants of UK yield equally plausible (yet fundamentally different) error variances, too. In essence, it is demonstrated that the terms 1D linear interpolation, TIN and bilinear interpolation do not suffice to determine the nature of measurement error, a fact that can lead to different, yet equally plausible from a pure data standpoint, prediction error variances.

2. Linear interpolation in 1D

Consider two locations along a 1D transect with coordinates x_1 and x_2, and measured values $z_1=z(x_1)$ and $z_2=z(x_2)$, respectively, shown in figure 1.

Using (piecewise) linear interpolation, the predicted value $\hat{z}=\hat{z}(x)$ at an arbitrary (but fixed) location with coordinate $x_1 \leqslant x \leqslant x_2$ is written as

$$\hat{z}=w_1 z_1+w_2 z_2=\left(\frac{x_2-x}{x_2-x_1}\right)z_1+\left(\frac{x-x_1}{x_2-x_1}\right)z_2=wz_1+(1-w)z_2,$$

where $w=(x_2-x)/(x_2-x_1) \in [0, 1]$ represents the weight assigned to datum z_1 for prediction at location x. In other words, the weight assigned to z_1 is the proportion of the segment length between point x_2 and the prediction location x, to the total segment length (x_2-x_1). In this work, we do not use any index to explicate the dependence of w on the prediction location.

When the observations z_1 and z_2 are treated as outcomes of spatially correlated random variables (RVs) Z_1 and Z_2, one defines the predictor RV $\hat{Z}=\hat{Z}(x)$ at location x as

$$\hat{Z}=wZ_1+(1-w)Z_2.$$

The variance of the predictor RV $\hat{Z}$ is then given by

$$\mathrm{Var}\left\{\hat{Z}\right\}=\mathrm{Var}\{wZ_1+(1-w)Z_2\}=w^2\sigma_{11}+(1-w)^2\sigma_{22}+2w(1-w)\sigma_{12}, \qquad (1)$$

where σ_{11} is the variance of the RV Z_1, σ_{22} is the variance of the RV Z_2, and σ_{12} is the covariance between Z_1 and Z_2.

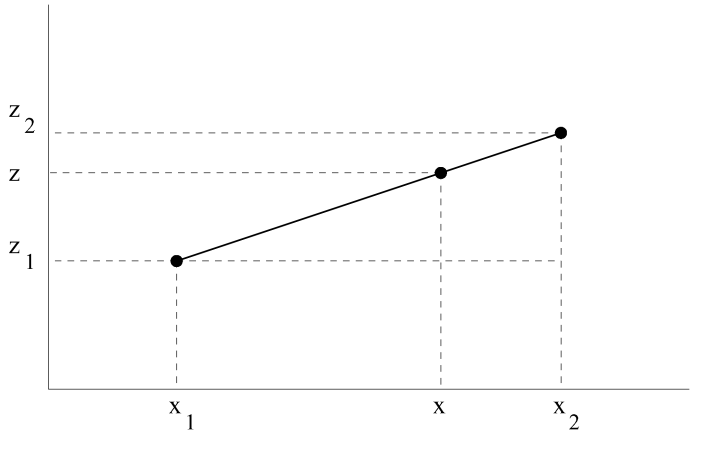

Figure 1. Example of linear interpolation in 1D.

Consider now the case whereby the prediction location x is randomly chosen on the line segment between x_1 and x_2, i.e. x is drawn from a uniform distribution in $[x_1, x_2]$. This also entails that the weight parameter w is an outcome (realization) from a uniform distribution in [0, 1].

The average variance of the predictor RV $\widehat{Z}$ over randomly chosen locations is obtained by taking the expectation of equation (1) with respect to the (now random) weight parameter W:

$$E\left\{\mathrm{Var}\left\{\widehat{Z}\right\}\right\} = E\{W^2\}\sigma_{11} + E\left\{(1-W)^2\right\}\sigma_{22} + 2E\{W(1-W)\}\sigma_{12}. \qquad (2)$$

The term $E\{W^2\}$ of equation (2) is expanded as

$$E\{W^2\} = \mathrm{Var}\{W\} + [E\{W\}]^2 = \frac{1}{12} + \left(\frac{1}{2}\right)^2 = \boxed{\frac{1}{3}} = E\left\{(1-W)^2\right\},$$

since the mean and variance of an RV uniformly distributed in [0, 1] are 1/2 and 1/12, respectively. In addition, both the RV W and its complement RV $(1-W)$ are uniformly distributed in [0, 1], hence $E\{W^2\} = E\{(1-W)^2\}$.

The term $2E\{W(1-W)\}$ of equation (2) is expanded as

$$2E\{W(1-W)\} = 2E\{W - W^2\} = 2\left[E\{W\} - E\{W^2\}\right] = 2\left[\frac{1}{2} - \frac{1}{3}\right] = \frac{1}{3}. \qquad (3)$$

By accounting for the above results, equation (2) becomes

$$\boxed{E\left\{\mathrm{Var}\left\{\widehat{Z}\right\}\right\} = \frac{1}{3}(\sigma_{11} + \sigma_{22} + \sigma_{12}).} \qquad (4)$$

Special cases of equation (4):

(i) Both RVs have unit variance: $\sigma_{11} = \sigma_{22} = 1$, but arbitrary correlation $\rho_{12} = \rho$:

$$E\left\{\mathrm{Var}\left\{\widehat{Z}\right\}\right\} = \frac{1}{3}(1 + 1 + \rho) = \frac{2+\rho}{3}.$$

(ii) Both RVs have unit variance, and zero correlation $\rho = 0$:

$$E\left\{\mathrm{Var}\left\{\widehat{Z}\right\}\right\} = \frac{1}{3}(1 + 1 + 0) = \frac{2}{3}.$$

(iii) Both RVs have unit variance, and perfect *positive* correlation $\rho = 1$:

$$E\left\{\mathrm{Var}\left\{\widehat{Z}\right\}\right\} = \frac{1}{3}(1 + 1 + 1) = 1.$$

(iv) Both RVs have unit variance, and perfect *negative* correlation $\rho = -1$:

$$E\left\{\mathrm{Var}\left\{\widehat{Z}\right\}\right\} = \frac{1}{3}(1 + 1 - 1) = \frac{1}{3}.$$

It should be noted that cases (iii) and (iv) are pathological, in the sense that they yield singular correlation matrices.

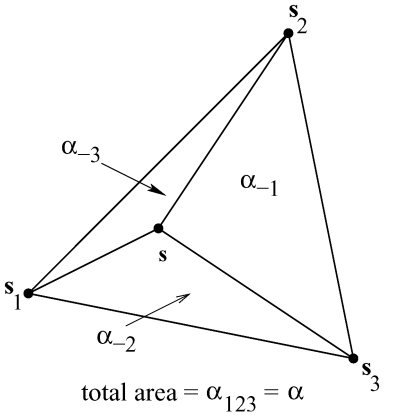

Figure 2. Example of linear interpolation in a TIN element.

3. TIN interpolation in 2D

Consider a TIN element as shown in figure 2, whose vertices have coordinate vectors $\mathbf{s}_1=(x_1, y_1)$, $\mathbf{s}_2=(x_2, y_2)$, and $\mathbf{s}_3=(x_3, y_3)$. Let the attribute measurements at these three locations be denoted as $z_1=z(\mathbf{s}_1)$, $z_2=z(\mathbf{s}_2)$ and $z_3=z(\mathbf{s}_3)$. In addition, the area of the TIN element is denoted as a.

Using TIN interpolation, the predicted value $\hat{z}=\hat{z}(\mathbf{s})$ at an arbitrary (but fixed) location in the TIN element with coordinate vector $\mathbf{s}=(x, y)$ is written as

$$\hat{z}=\sum_{i=1}^{3} w_i z_i=\sum_{i=1}^{3}\frac{a_{-i}}{a}z_i=w_1z_1+w_2z_2+w_3z_3,$$

where w_i denotes the weight assigned to datum z_i for prediction at location $\mathbf{s}$; again, the dependence of w_i on the prediction location is not explicated. The weight w_i is the proportion a_{-i}/a of the area a_{-i} of the sub-triangle formed by the two vertices opposite $\mathbf{s}_i$ (excluding $\mathbf{s}_i$, and thus the notation a_{-i}) and the prediction location $\mathbf{s}$, to the total area a of the TIN element; see Section 7.1 for a proof. Note that $\Sigma_{i=1}^{3} w_i=1$, by construction.

When the attribute measurements z_1, z_2 and z_3 are treated as outcomes of spatially correlated RVs Z_1, Z_2 and Z_3, one defines the predictor RV $\hat{Z}=\hat{Z}(s)$ at location $\mathbf{s}$ as

$$\hat{Z}=w_1Z_1+w_2Z_2+w_3Z_3.$$

The variance of the predictor RV $\hat{Z}$ is given by

$$\begin{aligned}\mathrm{Var}\left\{\hat{Z}\right\}&=\mathrm{Var}\{w_1Z_1+w_2Z_2+w_3Z_3\}=\sum_{i=1}^{3} w_i\sigma_{ii}+2\sum_{i}\sum_{j>i} w_iw_j\sigma_{ij}\\&=w_1^2\sigma_{11}+w_2^2\sigma_{22}+w_3^2\sigma_{33}+2w_1w_2\sigma_{12}+2w_1w_3\sigma_{13}+2w_2w_3\sigma_{23},\end{aligned}\tag{5}$$

where σ_{ii} is the variance of the ith RV Z_i, and σ_{ij} is the covariance between RVs Z_i and Z_j.

Consider now the case where the prediction location $\mathbf{s}$ is randomly chosen within the TIN element, which entails that the areas a_{-1}, a_{-2}, a_{-3} of the resulting

sub-triangles of figure 2, and hence the corresponding weights w_1, w_2, w_3, are also random. The area a_{-i} of each sub-triangle is completely determined by the height of that triangle, since its base is fixed (being the corresponding side of the TIN element). The proportion of the randomly picked height to the total height of the TIN element (from that same base) is a random variable with triangular distribution Tri [0, 1] and mode at 0. This is due to the probability of drawing a small height being larger than that of drawing a large one (large heights correspond to smaller sub-areas of the TIN element). By the same token, the proportion of the area of the resulting sub-triangle to the area a of the TIN element is an RV with the same distribution, since area is proportional to height. Consequently, the weight RVs W_1, W_2, W_3 follow the same triangular distribution in [0, 1] with mode at 0.

Let us now determine the covariance and correlation between the weight RVs W_1, W_2 and W_3. The covariance of any two RVs W_i and W_j can be expressed as

$$\mathrm{Cov}\{W_i, W_j\} = E\{W_i W_j\} - E\{W_i\}E\{W_j\} = E\{W_i W_j\} - \frac{1}{9}$$

since the mean of an RV with triangular distribution in [0, 1] with mode 0 is 1/3 (Evans *et al.* 2000).

The only unknown in the above equation is the term $E\{W_i W_j\}$, which can be expanded for all three RVs into the following system of equations:

$$E\{W_1 W_2\} = E\{W_1(1 - W_1 - W_3)\} = E\{W_1\} - E\{W_1^2\} - E\{W_1 W_3\}$$
$$E\{W_1 W_2\} = E\{(1 - W_2 - W_3) W_2\} = E\{W_2\} - E\{W_2^2\} - E\{W_2 W_3\}$$
$$E\{W_1 W_3\} = E\{(1 - W_2 - W_3) W_3\} = E\{W_3\} - E\{W_3^2\} - E\{W_2 W_3\}.$$

Since any RV W_i has a triangular distribution, it follows that

$$E\{W_i\} - E\{W_i^2\} = E\{W_i\} - \mathrm{Var}\{W_i\} - [E\{W_i\}]^2 = \frac{1}{3} - \frac{1}{18} - \frac{1}{9} = \frac{1}{6}$$

because the variance of an RV with triangular distribution in [0, 1] with mode 0 is 1/18 (Evans *et al.* 2000).

The above system of equations can therefore be written as

$$E\{W_1 W_2\} + E\{W_1 W_3\} = \frac{1}{6}$$
$$E\{W_1 W_2\} + E\{W_2 W_3\} = \frac{1}{6}$$
$$E\{W_1 W_3\} + E\{W_2 W_3\} = \frac{1}{6}.$$

The solution of this system of three equations in three unknowns yields

$$E\{W_i W_j\} = \frac{1}{12}, \quad \forall i \neq j, \tag{6}$$

which entails

$$\boxed{\mathrm{Cov}\{W_i W_j\} = \frac{1}{12} - \frac{1}{9} = -\frac{1}{36} \text{ and } \mathrm{Corr}\{W_i W_j\} = -\frac{1}{36}\Big/\frac{1}{18} = -\frac{1}{2}.} \tag{7}$$

In other words, *the areal proportions (to the total TIN element area) of sub-triangles formed between the TIN element vertices and randomly located points in that element, follow triangular distributions* Tri [0, 1]*, with mode* 0 *and pairwise correlations* -0.5.

The average variance of the predictor RV $\widehat{Z}$ over randomly chosen locations is obtained by taking the expectation of equation (5) with respect to the (now random) weight parameters W_1, W_2 and W_3:

$$E\left\{\mathrm{Var}\left\{\widehat{Z}\right\}\right\} = E\{W_1^2\}\sigma_{11} + E\{W_2^2\}\sigma_{22} + E\{W_3^2\}\sigma_{33} + 2E\{W_1W_2\}\sigma_{12} + 2E\{W_1W_3\}\sigma_{13} + 2E\{W_2W_3\}\sigma_{23}. \tag{8}$$

The terms $E\{W_i^2\}$ of the above equation are

$$E\{W_i^2\} = \mathrm{Var}\{W_i\} + [E\{W_i\}]^2 = \frac{1}{18} + \left(\frac{1}{3}\right)^2 = \frac{1}{9}\left[\frac{1}{2} + 1\right] = \frac{1}{6}, \tag{9}$$

By accounting for equations (9) and (6), equation (8) becomes

$$\boxed{E\left\{\mathrm{Var}\left\{\widehat{Z}\right\}\right\} = \frac{1}{6}(\sigma_{11} + \sigma_{22} + \sigma_{33} + \sigma_{12} + \sigma_{13} + \sigma_{23}).} \tag{10}$$

Special cases of equation (10):

(i) All RVs have unit variance: $\sigma_{ii}=1$, $\forall i$, and arbitrary but equal pairwise correlations: $\rho_{ij}=\rho$, $\forall i, j$:

$$E\left\{\mathrm{Var}\left\{\widehat{Z}\right\}\right\} = \frac{1}{6}(3 + 3\rho) = \frac{1+\rho}{2}.$$

(ii) All RVs have unit variance, and zero pairwise correlations $\rho_{ij}=0$, $\forall i, j$:

$$E\left\{\mathrm{Var}\left\{\widehat{Z}\right\}\right\} = \frac{1}{6}(3 + 0) = \frac{1}{2}.$$

(iii) All RVs have unit variance, and perfect *positive* pairwise correlations $\rho_{ij}=1$, $\forall i, j$:

$$E\left\{\mathrm{Var}\left\{\widehat{Z}\right\}\right\} = \frac{1}{6}(3 + 3) = 1,$$

where, again, this case is pathological, in the sense that it yields a singular correlation matrix.

It should be noted here that, in a recent paper, Zhu *et al.* (2005) obtained the same result as above for the special case of zero covariances and equal variances, and demonstrated its validity using simulation. Our derivation is both simpler and more general (see also the following sections).

4. An alternative formulation for the error of TIN interpolation

Consider a TIN element as shown in figure 3, and a randomly drawn (but fixed) line segment parallel to one base (here the line segment connecting $\mathbf{s}_1$ and $\mathbf{s}_3$). The left

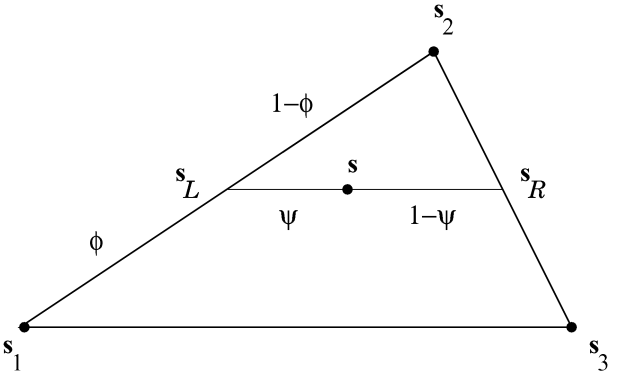

Figure 3. Alternative example of linear interpolation in a TIN element.

endpoint of this segment (on the side connecting $\mathbf{s}_1$ and $\mathbf{s}_2$) is denoted as $\mathbf{s}_L$, and the right endpoint of this segment (on the side connecting $\mathbf{s}_3$ and $\mathbf{s}_2$) is denoted as $\mathbf{s}_R$. The proportion, to the respective lengths of the two sides of the triangle, of the distance between each endpoint and the vertices $\mathbf{s}_1$ and $\mathbf{s}_3$ is ϕ.

Linear interpolation at these two endpoints $\mathbf{s}_L$ and $\mathbf{s}_R$ yields predicted values $\hat{z}_L = \hat{z}(s_L)$ and $\hat{z}_R = \hat{z}(s_R)$:

$$\hat{z}_L = (1-\phi)z_1 + \phi z_2 \text{ and } \hat{z}_R = (1-\phi)z_3 + \phi z_2.$$

Consider now a randomly located (but fixed) point $\mathbf{s}=(x, y)$ on the line segment connecting $\mathbf{s}_L$ and $\mathbf{s}_R$, whose distance from the left endpoint $\mathbf{s}_L$ is a proportion ψ of the length of that line. Linear interpolation between $\mathbf{s}_L$ and $\mathbf{s}_R$ yields the predicted value $\hat{z} = \hat{z}(s)$ at $\mathbf{s}$:

$$\begin{aligned}\hat{z} &= (1-\psi)\hat{z}_L + \psi\hat{z}_R \\ &= (1-\phi)(1-\psi)z_1 + \phi(1-\psi)z_2 + (1-\phi)\psi z_3 + \phi\psi z_2 \\ &= (1-\phi)(1-\psi)z_1 + \phi z_2 + (1-\phi)\psi z_3.\end{aligned}$$

When the attribute measurements z_1, z_2 and z_3 are treated as outcomes of spatially correlated RVs Z_1, Z_2 and Z_3, one defines the predictor RV $\hat{Z} = \hat{Z}(s)$ at location $\mathbf{s}$ as

$$\hat{Z} = (1-\phi)(1-\psi)Z_1 + \phi Z_2 + (1-\phi)\psi Z_3.$$

The variance of the predictor RV $\hat{Z}$ is given by

$$\begin{aligned}\mathrm{Var}\left\{\hat{Z}\right\} = {} & (1-\phi)^2(1-\psi)^2\sigma_{11} + \phi^2\sigma_{22} + (1-\phi)^2\psi^2\sigma_{33} \\ & + 2(1-\phi)(1-\psi)\phi\sigma_{12} + 2(1-\phi)^2(1-\psi)\psi\sigma_{13} + 2\phi(1-\phi)\psi\sigma_{23}.\end{aligned} \tag{11}$$

Consider now the case where the prediction location $\mathbf{s}$ is randomly chosen within the TIN triangle, which entails that the parameters ϕ and ψ are themselves randomly generated from a triangular and a uniform distribution, respectively.

More precisely, $\Phi \sim \text{Tri}\,[0, 1]$, with mode 0, and $\Psi \sim \text{Unif}\,[0, 1]$; in addition, the two RVs Φ and Ψ are uncorrelated, i.e. $\text{Cov}\{\Phi, \Psi\} = 0$. The mean of the triangular RV Φ is $E\{\Phi\} = 1/3$, and its variance is $\text{Var}\{\Phi\} = 1/18$. The expected value of Φ^2 is

$$E\{\Phi^2\} = \text{Var}\{\Phi\} + [E\{\Phi\}]^2 = \frac{1}{18} + \frac{1}{9} = \frac{1}{6}.$$

The complement RV $(1-\Phi)$ also has a triangular distribution in [0, 1], with mode 1; its mean is $E\{(1-\Phi)\} = 2/3$, and its variance is $\text{Var}\{(1-\Phi)\} = 1/18$ (Evans *et al.* 2000). The expected value of $(1-\Phi)^2$ is

$$E\left\{(1-\Phi)^2\right\} = \text{Var}\{(1-\Phi)\} + [E\{(1-\Phi)\}]^2 = \frac{1}{18} + \frac{4}{9} = \frac{1}{2}.$$

The mean and variance of the uniform RVs Ψ and $(1-\Psi)$ are $E\{\Psi\} = 1/2 = E\{(1-\Psi)\}$ and $\text{Var}\{\Psi\} = 1/12 = \text{Var}\{(1-\Psi)\}$. The expected value of Ψ^2 and $(1-\Psi)^2$ is

$$E\{\Psi^2\} = \text{Var}\{\Psi\} + [E\{\Psi\}]^2 = \frac{1}{12} + \frac{1}{4} = \frac{4}{12} = \boxed{\frac{1}{3}} = E\left\{(1-\Psi)^2\right\}.$$

The average variance of the predictor RV $\widehat{Z}$ over randomly chosen locations is obtained by taking the expectation of equation (11) with respect to the (now random) parameters Φ and Ψ:

$$\begin{aligned} E\left\{\text{Var}\left\{\widehat{Z}\right\}\right\} = {} & E\left\{(1-\Phi)^2(1-\Psi)^2\right\}\sigma_{11} + E\{\Phi^2\}\sigma_{22} \\ & + E\left\{(1-\Phi)^2\Psi^2\right\}\sigma_{33} + 2E\{(1-\Phi)(1-\Psi)\Phi\}\sigma_{12} \\ & + 2E\left\{(1-\Phi)^2(1-\Psi)\Psi\right\}\sigma_{13} + 2E\{\Phi(1-\Phi)\Psi\}\sigma_{23}. \end{aligned} \tag{12}$$

Let us now focus on the individual terms of equation (12). In the equations that follow, we will use the fact that for two independent RVs X and Y, and for two arbitrary functions $f(X)$ and $g(Y)$: $E\{f(X)g(Y)\} = E\{f(X)\}E\{g(Y)\}$. In other words, the expected value of the product of two RVs that are defined as functions of other independent RVs can be factored into the product of the individual expectations:

$$E\left\{(1-\Phi)^2(1-\Psi)^2\right\} = E\left\{(1-\Phi)^2\right\}E\left\{(1-\Psi)^2\right\} = \frac{1}{2}\frac{1}{3} = \frac{1}{6}$$

$$E\left\{(1-\Phi)^2\Psi^2\right\} = E\left\{(1-\Phi)^2\right\}E\{\Psi^2\} = \frac{1}{2}\frac{1}{3} = \frac{1}{6}$$

$$\begin{aligned} 2E\{(1-\Phi)(1-\Psi)\Phi\} &= 2E\{(1-\Phi)\Phi\}E\{(1-\Psi)\} \\ &= 2\left[E\{\Phi\} - E\{\Phi^2\}\right]E\{(1-\Psi)\} = 2\left[\frac{1}{3} - \frac{1}{6}\right]\frac{1}{2} = \frac{1}{6} \end{aligned}$$

$$2E\left\{(1-\Phi)^2(1-\Psi)\Psi\right\} = 2E\left\{(1-\Phi)^2\right\}E\{(1-\Psi)\Psi\}$$
$$= 2E\left\{(1-\Phi)^2\right\}\left[E\{\Psi\} - E\{\Psi^2\}\right] = 2\frac{1}{2}\left[\frac{1}{2} - \frac{1}{3}\right] = \frac{1}{6}$$

$$2E\{\Phi(1-\Phi)\Psi\} = 2E\{\Phi(1-\Phi)\}E\{\Psi\}$$
$$= 2\left[E\{\Phi\} - E\left\{\Phi^2\right\}\right]E\{\Psi\} = 2\left[\frac{1}{3} - \frac{1}{6}\right]\frac{1}{2} = \frac{1}{6}.$$

Based on the above results, equation (12) becomes

$$\boxed{E\left\{\mathrm{Var}\left\{\hat{Z}\right\}\right\} = \frac{1}{6}(\sigma_{11} + \sigma_{22} + \sigma_{33} + \sigma_{12} + \sigma_{13} + \sigma_{23}),} \tag{13}$$

which is identical with equation (10) of Section 3.

5. Bilinear interpolation in a rectangle

Consider the rectangle of figure 4 with vertices $\mathbf{s}_1=(x_1, y_1)$, $\mathbf{s}_2=(x_1, y_2)$, $\mathbf{s}_3=(x_2, y_2)$ and $\mathbf{s}_4=(x_2, y_1)$, as well as a randomly chosen (but fixed) location in that rectangle with coordinate vector $\mathbf{s}=(x, y)$. The sample measurements at the four vertices are denoted as $z_1=z(\mathbf{s}_1)$, $z_2=z(\mathbf{s}_2)$, $z_3=z(\mathbf{s}_3)$ and $z_4=z(\mathbf{s}_4)$. Let $\mathbf{s}_T=(x, y_2)$ and $\mathbf{s}_B=(x, y_1)$ denote the coordinates of the top and bottom endpoints of a line segment drawn from $\mathbf{s}$ parallel to the rectangle side $\mathbf{s}_1$, $\mathbf{s}_2$. Let ϕ denote the proportion of distance between $\mathbf{s}_B$ and $\mathbf{s}_4$ to the total distance of side $\mathbf{s}_1$, $\mathbf{s}_4$, and let ψ denote the proportion of distance between $\mathbf{s}$ and $\mathbf{s}_T$ to the total distance of side $\mathbf{s}_1$, $\mathbf{s}_2$.

Bilinear interpolation at location $\mathbf{s}$ proceeds by first performing linear interpolation at points $\mathbf{s}_T$ and $\mathbf{s}_B$ from the measurement pairs z_2, z_3 and z_1, z_4, respectively, to obtain predicted values $\hat{z}_B=\hat{z}(\mathbf{s}_B)$ and $\hat{z}_T=\hat{z}(\mathbf{s}_T)$ as

$$\hat{z}_B = \phi z_1 + (1-\phi)z_4 \text{ and } \hat{z}_T = \phi z_2 + (1-\phi)z_3.$$

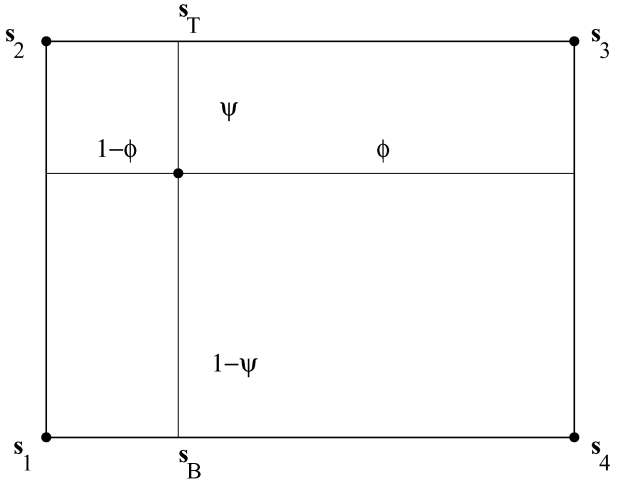

Figure 4. Example of bilinear interpolation in a rectangle.

The final prediction $\hat{z}=\hat{z}(\mathbf{s})$ at location $\mathbf{s}$ is obtained via linear interpolation between $\mathbf{s}_T$ and $\mathbf{s}_B$ using the previously derived predictions $\hat{z}_T$ and $\hat{z}_B$:

$$\begin{aligned}\hat{z}&=\psi\hat{z}_B+(1-\psi)\hat{z}_T=\psi[\phi z_1+(1-\phi)z_4]+(1-\psi)[\phi z_2+(1-\phi)z_3]\\&=\phi\psi z_1+\phi(1-\psi)z_2+(1-\phi)(1-\psi)z_3+(1-\phi)\psi z_4\\&=w_1z_1+w_2z_2+w_3z_3+w_4z_4,\end{aligned}$$

where $w_1=\phi\psi$, $w_2=\phi(1-\psi)$, $w_3=(1-\phi)(1-\psi)$ and $w_4=(1-\phi)\psi$.

It is also evident that the weights are proportional to the areas of the sub-rectangles formed by the prediction location $\mathbf{s}$, the endpoints $\mathbf{s}_T$, $\mathbf{s}_B$, $\mathbf{s}_L$, $\mathbf{s}_R$ of the line segments in the rectangle of figure 4, and the rectangle vertices. For example,

$$w_1=\phi\psi=\frac{\|s_R-s\|}{\|s_4-s_1\|}\frac{\|s_T-s\|}{\|s_2-s_1\|}=\frac{a_{-1}}{a},$$

where $\|\cdot\|$ denotes a vector norm and a_{-1} denotes the area of the sub-rectangle formed by vertex $\mathbf{s}_3$ (opposite to $\mathbf{s}_1$, hence the notation a_{-1}), points $\mathbf{s}_R$, $\mathbf{s}_T$ and the prediction location $\mathbf{s}$. Note also that $\Sigma_{i=1}^4 w_i=1$, by construction.

When the attribute measurements z_1, z_2, z_3 and z_4 are treated as outcomes of spatially correlated RVs Z_1, Z_2, Z_3 and Z_4, one defines the predictor RV $\hat{Z}=\hat{Z}(\mathbf{s})$ at location $\mathbf{s}$ as

$$\hat{Z}=w_1Z_1+w_2Z_2+w_3Z_3+w_4Z_4.$$

The variance of the predictor RV $\hat{Z}$ is expressed as

$$\begin{aligned}\mathrm{Var}\{\hat{Z}\}&=\mathrm{Var}\{w_1Z_1+w_2Z_2+w_3Z_3+w_4Z_4\}=\sum_{i=1}^4 w_i\sigma_{ii}+2\sum_i\sum_{j>i}w_iw_j\sigma_{ij}\\&=\phi^2\psi^2\sigma_{11}+\phi^2(1-\psi)^2\sigma_{22}+(1-\phi)^2(1-\psi)^2\sigma_{33}+(1-\phi)^2\psi^2\sigma_{44}\\&\quad+2\phi^2\psi(1-\psi)\sigma_{12}+2\phi(1-\phi)\psi(1-\psi)\sigma_{13}+2\phi(1-\phi)\psi^2\sigma_{14}\\&\quad+2\phi(1-\phi)(1-\psi)^2\sigma_{23}+2\phi(1-\phi)\psi(1-\psi)\sigma_{24}+2(1-\phi)^2\psi(1-\psi)\sigma_{34}.\end{aligned}\tag{14}$$

Consider now the case where the prediction location $\mathbf{s}$ is randomly chosen within the rectangle. This entails that the parameters ϕ and ψ are realizations of uncorrelated RVs Φ and Ψ that follow a uniform distribution in [0, 1]. In addition, the weights w_i are also random, since they are functions of ϕ and ψ.

The average variance of the predictor RV $\hat{Z}$ over randomly chosen locations is obtained by taking the expectation of equation (14) with respect to the (now random) weights:

$$\begin{aligned}E\{\mathrm{Var}\{\hat{Z}\}\}&=E\left\{\sum_{i=1}^4 W_i\sigma_{ii}+2\sum_i\sum_{j>i}W_iW_j\sigma_{ij}\right\}=\sum_{i=1}^4 E\{W_i\}\sigma_{ii}+2\sum_i\sum_{j>i}E\{W_iW_j\}\sigma_{ij}\\&=E\{\Phi^2\Psi^2\}\sigma_{11}+E\{\Phi^2(1-\Psi)^2\}\sigma_{22}+E\{(1-\Phi)^2(1-\Psi)^2\}\sigma_{33}\\&\quad+E\{(1-\Phi)^2\Psi^2\}\sigma_{44}+2E\{\Phi^2\Psi(1-\Psi)\}\sigma_{12}+2E\{\Phi(1-\Phi)\Psi(1-\Psi)\}\sigma_{13}\\&\quad+2E\{\Phi(1-\Phi)\Psi^2\}\sigma_{14}+2E\{\Phi(1-\Phi)(1-\Psi)^2\}\sigma_{23}\\&\quad+2E\{\Phi(1-\Phi)\Psi(1-\Psi)\}\sigma_{24}+2E\{(1-\Phi)^2\Psi(1-\Psi)\}\sigma_{34}.\end{aligned}\tag{15}$$

For an RV with uniform distribution in [0, 1],

$$E\{\Phi^2\} = \mathrm{Var}\{\Phi\} + [E\{\Phi\}]^2 = \frac{1}{12} + \frac{1}{4} = \boxed{\frac{1}{3}} = E\left\{(1-\Phi)^2\right\} = E\left\{(1-\Psi)^2\right\} = E\{\Psi^2\},$$

since the mean and variance of Φ are 1/2 and 1/12, respectively.

Similarly, for an RV with uniform distribution in [0, 1]:

$$E\{\Phi(1-\Phi)\} = E\{\Phi\} - E\{\Phi^2\} = \frac{1}{2} - \frac{1}{3} = \boxed{\frac{1}{6}} = E\{\Psi(1-\Psi)\}.$$

For two independent RVs Φ and Ψ, each uniformly distributed in [0, 1]:

$$E\{\Phi^2\Psi^2\} = E\{\Phi^2\}E\{\Psi^2\} = \frac{1}{3}\frac{1}{3} = \boxed{\frac{1}{9}}$$
$$= E\left\{(1-\Phi)^2\Psi^2\right\} = E\left\{\Phi^2(1-\Psi)^2\right\} = E\left\{(1-\Phi)^2(1-\Psi)^2\right\}.$$

By accounting for the above results, equation (15) becomes

$$E\left\{\mathrm{Var}\left\{\hat{Z}\right\}\right\} = \frac{1}{9}[\sigma_{11} + \sigma_{22} + \sigma_{33} + \sigma_{44}]$$
$$+ 2\frac{1}{3}\frac{1}{6}\sigma_{12} + 2\frac{1}{6}\frac{1}{6}\sigma_{13} + 2\frac{1}{6}\frac{1}{3}\sigma_{14} + 2\frac{1}{6}\frac{1}{3}\sigma_{23} + 2\frac{1}{6}\frac{1}{6}\sigma_{24} + 2\frac{1}{3}\frac{1}{6}\sigma_{34} \quad (16)$$
$$= \frac{1}{9}\left[\sigma_{11} + \sigma_{22} + \sigma_{33} + \sigma_{44} + \sigma_{12} + \sigma_{14} + \sigma_{23} + \sigma_{34} + \frac{1}{2}(\sigma_{13} + \sigma_{24})\right].$$

For the special case of unit variances $\sigma_{ii}=1$, $\forall i$ and zero covariances $\sigma_{ij}=0$, $\forall i, j$, equation (16) simplifies to

$$E\left\{\mathrm{Var}\left\{\hat{Z}\right\}\right\} = \frac{4}{9}.$$

6. General geostatistical formulation

In this section, we place spatial interpolation within a general geostatistical framework; the results given hereafter are then simplified in Section 7 to yield the three interpolation methods considered in this work as particular cases. It is demonstrated that different assumptions regarding the origin of measurement error lead to different prediction objectives and associated predicted values and error variances. Note that, in what follows, we loosely use the term surface to refer to a profile in 1D or an actual surface in 2D.

Consider the following linear regression model:

$$\mathbf{z} = \mathbf{m} + \mathbf{r} = \mathbf{Fb} + (\mathbf{q} + \mathbf{e}) = \mathbf{p} + \mathbf{e}, \quad (17)$$

where $\mathbf{z}=[z(\mathbf{s}_i),\ i=1,\ \ldots,\ n]'$ denotes an $(n \times 1)$ vector of measurements (sample data), $\mathbf{m}=[m(\mathbf{s}_i),\ i=1,\ \ldots,\ n]'$ denotes an $(n \times 1)$ vector of mean values with $\mathbf{m}=\mathbf{Fb}$, where $\mathbf{F}=[f_k(\mathbf{s}_i),\ i=1,\ \ldots,\ n,\ k=1,\ \ldots,\ K]$ is an $(n \times K)$ design matrix with $f_k(\mathbf{s}_i)$ being the value of the kth predictor variable at the ith sample location $\mathbf{s}_i$ (by

convention, $f_1(\mathbf{s}_i)=1$, $\forall i$), and $\mathbf{b}=[b_k,\ k=1,\ \ldots,\ K]'$ is a $(K\times 1)$ vector of regression coefficients; $\mathbf{r}=[r(\mathbf{s}_i),\ i=1,\ \ldots,\ n]'$ denotes an $(n\times 1)$ vector of regression residuals, which is composed of an $(n\times 1)$ vector $\mathbf{e}=[e(\mathbf{s}_i),\ i=1,\ \ldots,\ n]'$ associated with measurement error and an $(n\times 1)$ vector $\mathbf{q}=[q(\mathbf{s}_i),\ i=1,\ \ldots,\ n]'$ not associated with such error; $\mathbf{p}=\mathbf{m}+\mathbf{q}=\mathbf{z}-\mathbf{e}$ denotes the $(n\times 1)$ vector of 'signal' values, i.e. $\mathbf{p}$ is the composite of all components of z-variability not associated with measurement error.

In this section we consider: (i) an arbitrary design matrix $\mathbf{F}$, the only requirement being that it be of full rank, and (ii) the general over-determined case, whereby the number of observations is larger than the number of parameters involved in the regression model of equation (17), i.e. $n>K$. In Section 7, we consider: (i) particular design matrices populated with the coordinate values of the sample locations, and (ii) the completely determined case of $n=K$; the latter items (i) and (ii) jointly comprise the setting of the three spatial interpolation methods considered in this work.

The linear model of equation (17) decomposes the spatial variability of the data vector $\mathbf{z}$ into a deterministic mean component (drift) $\mathbf{m}=E\{\mathbf{z}|\mathbf{F}\}=\mathbf{Fb}$, and a stochastic residual component $\mathbf{r}$; such a decomposition is unavoidably subjective, since no data exist on either $\mathbf{m}$ or $\mathbf{r}$ (Ripley 1981, Cressie 1993, Chilès and Delfiner 1999). The mean component $\mathbf{m}$ accounts for spatial variability due to the predictor data stored in vector $\mathbf{F}$ (first-order effects), thus modelling the expected (average) surface, say a line in 1D. It is assumed here that the drift component $\mathbf{Fb}$ is correctly specified in terms of its order (the number K of predictors dictating the number of columns in $\mathbf{F}$) and its quality (the particular predictors considered to populate $\mathbf{F}$). The stochastic component $\mathbf{r}$ accounts for spatial variability stemming from the residuals (second-order effects), thus modelling stochastic deviations from the expected surface. That residual component $\mathbf{r}$ is assumed to be composed of two sub-components $\mathbf{q}$ and $\mathbf{e}$. In other words, deviations of the observations $\mathbf{z}$ from the expected surface $\mathbf{m}$ arise owing to: (i) the inadequacy of that expected surface to capture the true surface, and (ii) measurement error. Sub-component $\mathbf{q}$ thus models spatial variability in $\mathbf{r}$ due to stochastic deviations from $\mathbf{m}$ not attributed to measurement error, whereas sub-component $\mathbf{e}$ models spatial variability in $\mathbf{r}$ due precisely to measurement error.

Typical assumptions of the model of equation (17) include (Ripley 1981, Cressie 1993, Chilès and Delfiner 1999): (i) $\mathrm{Cov}\{\mathbf{r},\ \mathbf{F}\}=0$ – that is, residuals are uncorrelated with the predictor variables; (ii) $\mathbf{r}\sim N(\mathbf{0},\ \mathbf{\Sigma}_\mathrm{R})$ – that is, residuals follow a multivariate Gaussian distribution with zero mean, $\mathbf{0}$ being an $(n\times 1)$ vector of zeros, and an $(n\times n)$ covariance matrix $\mathbf{\Sigma}_\mathrm{R}=\left[\sigma^\mathrm{R}_{ij}=\sigma_\mathrm{R}(\mathbf{s}_i,\ \mathbf{s}_j),\ i=1,\ \ldots,\ n,\ j=1,\ \ldots,\ n\right]$, where $\sigma_\mathrm{R}(\mathbf{s}_i,\ \mathbf{s}_j)$ is the residual covariance between sample locations $\mathbf{s}_i$ and $\mathbf{s}_j$; (iii) $\mathbf{q}\sim N(\mathbf{0},\ \mathbf{\Sigma}_Q)$ and $\mathbf{e}\sim N(\mathbf{0},\ \mathbf{\Sigma}_E)$ – that is, both sub-components $\mathbf{q}$ and $\mathbf{e}$ follow a multivariate Gaussian distribution with zero mean and $(n\times n)$ covariance matrices $\mathbf{\Sigma}_Q$ and $\mathbf{\Sigma}_E$, respectively; and (iv) $\mathrm{Cov}\{\mathbf{e},\ \mathbf{q}\}=\mathrm{Cov}\{\mathbf{q},\ \mathbf{F}\}=\mathrm{Cov}\{\mathbf{e},\ \mathbf{F}\}=\mathbf{0}$ – that is, both sub-components are uncorrelated with each other, and also uncorrelated with the predictor variables; since $\mathrm{Cov}\{\mathbf{e},\ \mathbf{q}\}=0$, the residual covariance matrix $\mathbf{\Sigma}_\mathrm{R}$ can be decomposed as: $\mathbf{\Sigma}_\mathrm{R}=\mathbf{\Sigma}_Q+\mathbf{\Sigma}_E$.

Under the general decomposition of equation (17), the objective is to predict the noise-free component $p(\mathbf{s}_0)$ of the unavailable measurement $z(\mathbf{s}_0)$ at location $\mathbf{s}_0$. This general prediction objective falls in the realm of Factorial Universal Kriging (Chilès and Delfiner 1999), abbreviated here as UKF. The signal component $\mathbf{p}$ in

equation (17), however, can attain two other limit values (which in turn affects the definition of the sought-after signal), depending on the relative magnitude of the two residual sub-components **q** and **e** (Ripley 1981, Cressie 1993, Chilès and Delfiner 1999).

On one hand, the residual component **r** could be entirely attributed to measurement error. In this case, $\mathbf{q}=\mathbf{0}$, and consequently $\mathbf{r}=\mathbf{e}$ and $\boldsymbol{\Sigma}_R=\boldsymbol{\Sigma}_E$. Equation (17) thus becomes

$$\mathbf{z}=\mathbf{m}+\mathbf{r}=\mathbf{F}\mathbf{b}+\mathbf{e} \tag{18}$$

and the objective here is to obtain $\widehat{m}(\mathbf{s}_0)$, i.e. a prediction of the unknown mean component $m(\mathbf{s}_0)$ at location $\mathbf{s}_0$. This prediction objective falls in the realm of Universal Kriging of the drift (mean) component (Chilès and Delfiner 1999), abbreviated here as UKM, which coincides with Generalized Least Squares (GLS) prediction – the letter 'M' in the term 'UKM' explicates that the prediction objective pertains to the mean component of the z-attribute.

On the other hand, the contribution of measurement error on the residual component **r** could be assumed to be null. In this case $\mathbf{e}=\mathbf{0}$, and consequently $\mathbf{r}=\mathbf{q}$ and $\boldsymbol{\Sigma}_R=\boldsymbol{\Sigma}_Q$. Equation (17) thus becomes

$$\mathbf{z}=\mathbf{m}+\mathbf{r}=\mathbf{F}\mathbf{b}+\mathbf{q} \tag{19}$$

and the objective is to obtain $\widehat{z}(\mathbf{s}_0)$, i.e. a prediction of the unknown attribute value $z(\mathbf{s}_0)$ at location $\mathbf{s}_0$, since there is no measurement error. This prediction objective falls in the realm of Universal Kriging of the attribute itself (Chilès and Delfiner 1999), abbreviated here as UKZ – the letter 'Z' in term 'UKZ' explicates that the prediction objective pertains to the z-attribute itself.

In what follows, we label as *Model I* the case of equation (18) corresponding to complete domination of measurement error, as *Model II* the case of equation (19) corresponding to complete absence of measurement error, and as *Model III* the general case of equation (17). We derive the predictions and associated prediction error variances for these three models for the over-determined case ($n>K$) and for arbitrary sampling configurations. The completely determined case ($n=K$) for the particular sampling configurations considered in 1D linear interpolation, TIN and bilinear interpolation is treated in Section 7.

6.1 *Spatial prediction under Model I*

The UKM prediction $\widehat{m}(\mathbf{s}_0)$ of the unknown mean component $m(\mathbf{s}_0)$ at location $\mathbf{s}_0$ is written as a weighted linear combination of the entries of the measurement vector **z**:

$$\widehat{m}(\mathbf{s}_0)=\left(\mathbf{w}_0^{\mathrm{M}}\right)'\mathbf{z}, \tag{20}$$

where $\mathbf{w}_0^{\mathrm{M}}=\left[w_i^{\mathrm{M}}(\mathbf{s}_0),\ i=1,\ \ldots,\ n\right]'$ denotes an $(n\times 1)$ vector of UKM weights.

Unbiasedness of prediction, i.e. $E\left\{\widehat{M}(\mathbf{s}_0)\right\}=E\{M(\mathbf{s}_0)\}=m(\mathbf{s}_0)$, is ensured by imposing the constraints $\mathbf{F}'\mathbf{w}_0^{\mathrm{M}}=\mathbf{f}_0$ on the UKM weights for every prediction location $\mathbf{s}_0$ (Chilès and Delfiner 1999), where $\mathbf{f}_0=[f_k(\mathbf{s}_0),\ k=1,\ \ldots,\ K]'$ denotes the $(K\times 1)$ vector of predictor variables at location $\mathbf{s}_0$, with $f_k(\mathbf{s}_0)$ being the value of the kth variable at that location; by convention, $f_1(\mathbf{s}_0)=1$.

The UKM weights vector $\mathbf{w}_0^{\mathrm{M}}$ is obtained by solving the following system of equations (Chilès and Delfiner 1999):

$$\begin{bmatrix} \boldsymbol{\Sigma}_{\mathrm{R}} & \mathbf{F} \\ \mathbf{F}' & \mathbf{O} \end{bmatrix} \begin{bmatrix} \mathbf{w}_0^{\mathrm{M}} \\ -\mathbf{t}_0^{\mathrm{M}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{f}_0 \end{bmatrix},$$

where $\mathbf{O}$ denotes a $(K \times K)$ matrix of zeros, and $\mathbf{t}_0^{\mathrm{M}} = \left[t_k^{\mathrm{M}}(\mathbf{s}_0),\, k=1,\, \ldots,\, K\right]'$ denotes a $(K \times 1)$ vector of Lagrange multipliers (due to the constraints on the weights). Note again that, under *Model I*, the residual covariance is identified with the covariance of the measurement error component, i.e. $\boldsymbol{\Sigma}_{\mathrm{R}} = \boldsymbol{\Sigma}_{\mathbf{E}}$.

The solution to the above system of equations can be analytically derived using the inverse of a partitioned matrix as (Searle 1982):

$$\begin{bmatrix} \mathbf{w}_0^{\mathrm{M}} \\ -\mathbf{t}_0^{\mathrm{M}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathrm{R}}^{-1} - \boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{F}\mathbf{A}\mathbf{F}'\boldsymbol{\Sigma}_{\mathrm{R}}^{-1} & \boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{F}\mathbf{A} \\ \mathbf{A}\mathbf{F}'\boldsymbol{\Sigma}_{\mathrm{R}}^{-1} & -\mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \mathbf{f}_0 \end{bmatrix}, \tag{21}$$

where $\mathbf{A} = \left(\mathbf{F}'\boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{F}\right)^{-1}$ is a $(K \times K)$ symmetric matrix.

From equation (21), the UKM weights vector $\boldsymbol{w}_{\mathbf{0}}^{\mathbf{M}}$ can be expressed as

$$\mathbf{w}_0^{\mathrm{M}} = \boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{F}\mathbf{A}\mathbf{f}_0 = \boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{F}\left(\mathbf{F}'\boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{F}\right)^{-1}\mathbf{f}_0, \tag{22}$$

and consequently the UKM prediction $\widehat{m}(\mathbf{s}_0)$ of equation (20) becomes

$$\widehat{m}(\mathbf{s}_0) = \left(\mathbf{w}_0^{\mathrm{M}}\right)'\mathbf{z} = \left[\mathbf{f}_0'\mathbf{A}\mathbf{F}'\boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\right]\mathbf{z} = \left[\mathbf{f}_0'\left(\mathbf{F}'\boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{F}\right)^{-1}\mathbf{F}'\boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\right]\mathbf{z} = \mathbf{f}_0'\widehat{\mathbf{b}}, \tag{23}$$

where $\mathbf{A}' = \mathbf{A}$ and $\left(\boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\right)' = \boldsymbol{\Sigma}_{\mathrm{R}}^{-1}$, because both $\mathbf{A}$ and $\boldsymbol{\Sigma}_{\mathrm{R}}^{-1}$ are symmetric matrices; $\widehat{\mathbf{b}}$ is the $(K \times 1)$ vector of GLS regression coefficients:

$$\widehat{\mathbf{b}} = \mathbf{A}\mathbf{F}'\boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{z} = \left(\mathbf{F}'\boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{F}\right)^{-1}\mathbf{F}'\boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{z}, \tag{24}$$

and it is a function of the covariance $\boldsymbol{\Sigma}_{\mathrm{R}}$ of the measurement error component adopted in *Model I*.

Note that, in the over-determined case ($n > K$), the UKM predictions derived from equation (23) *do not* reproduce the data values at their locations; in other words, *UKM is generally a smoother, not an interpolator, no matter what the covariance matrix* $\boldsymbol{\Sigma}_{\mathrm{R}}$ *is*.

From equation (21), the vector $\mathbf{t}_0^{\mathrm{M}}$ of UKM Lagrange multipliers can be written as $\mathbf{t}_0^{\mathrm{M}} = \mathbf{A}\mathbf{f}_0$, leading to the following expression for the UKM prediction error variance $\widehat{\sigma}_{\mathrm{M}}(\mathbf{s}_0)$ for the unknown local mean component $m(\mathbf{s}_0)$ (Chilès and Delfiner 1999):

$$\widehat{\sigma}_{\mathrm{M}}(\mathbf{s}_0) = \left(\mathbf{t}_0^{\mathrm{M}}\right)'\mathbf{f}_0 = \mathbf{f}_0'\mathbf{A}\mathbf{f}_0 = \left[\mathbf{f}_0'\left(\mathbf{F}'\boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{F}\right)^{-1}\right]\mathbf{f}_0, \tag{25}$$

where $\mathbf{A}$ can be seen as the $(K \times K)$ covariance matrix of the estimated regression coefficients in vector $\widehat{\mathbf{b}}$.

For a given sample configuration and a given residual covariogram model, i.e. for a given residual covariance matrix $\boldsymbol{\Sigma}_{\mathrm{R}}$, the spatial pattern of the UKM error variance $\widehat{\sigma}_{\mathrm{M}}(\mathbf{s}_0)$ is dictated by the relative magnitude of $\mathbf{f}_0$ with respect to the columns of the design matrix $\mathbf{F}$ appearing in $\mathbf{A}$. It is well known from classical regression theory (Searle 1971), that (when residual variances are equal) the prediction variance for the mean component $\widehat{\sigma}_{\mathrm{M}}(\mathbf{s}_0)$ attains its minimum for that $\mathbf{f}_0$ vector whose entries are the mean values of each predictor variable. In our case,

whereby the predictor variables are (functions of) location coordinates, the UKM prediction error variance $\widehat{\sigma}_{\mathrm{M}}(\mathbf{s}_0)$ attains larger values at the sample locations, since at those locations the corresponding coordinates differ more from their mean.

When an arbitrary (yet positive-definite) residual covariance matrix $\boldsymbol{\Sigma}_{\mathrm{R}}=\boldsymbol{\Sigma}_{\mathrm{E}}$ is used for the measurement errors in the above equations, the UKM prediction of the mean component $\widehat{\mathbf{m}}(\mathbf{s}_0)$ and the associated error variance $\widehat{\sigma}_{\mathrm{M}}(\boldsymbol{s_0})$ coincide with those obtained via GLS. When that covariance matrix $\boldsymbol{\Sigma}_{\mathrm{R}}$ is diagonal, with the entries along its main diagonal taken from vector $[\sigma_{\mathrm{R}}(\mathbf{s}_i),\ i=1,\ \ldots,\ n]'$, where $\sigma_{\mathrm{R}}(\mathbf{s}_i)$ denotes the measurement error variance at location $\mathbf{s}_i$, then the UKM prediction and associated error variance coincide with those obtained via Weighted Least Squares (WLS). Lastly, when $\boldsymbol{\Sigma}_{\mathrm{R}}$ is still diagonal, but with constant entries along its main diagonal, i.e. $\boldsymbol{\Sigma}_{\mathrm{R}}=\sigma_{\mathrm{R}}\mathbf{I}$, where σ_{R} denotes the overall variance of the measurement errors and $\mathbf{I}$ denotes the $(n\times n)$ identity matrix, the UKM prediction and associated error variance coincide with Ordinary Least Squares (OLS). *When the elements of vector* $\mathbf{f}_0$ *and matrix* $\mathbf{F}$ *are functions of coordinates, this latter OLS case corresponds to classical trend surface analysis*. It should be noted again that, in the over-determined case ($n>K$), trend surface fitting does not reproduce the data values at their sample locations; in other words, *trend surface models are generally smoothers, not interpolators*.

6.2 *Spatial prediction under Model II*

The UKZ prediction $\widehat{z}(\mathbf{s}_0)$ of the unknown, noise-free, attribute value $z(\mathbf{s}_0)$ at location $\mathbf{s}_0$ is again written as a weighted linear combination of the entries of the data vector $\mathbf{z}$:

$$\widehat{z}(\mathbf{s}_0)=\left(\mathbf{w}_0^{\mathrm{Z}}\right)'\mathbf{z}, \tag{26}$$

where $\mathbf{w}_0^{\mathrm{Z}}=\left[\mathrm{w}_i^{\mathrm{Z}}(\mathbf{s}_0),\ i=1,\ \ldots,\ n\right]'$ denotes the $(n\times 1)$ vector of UKZ weights for prediction at $\mathbf{s}_0$.

Unbiasedness of prediction, i.e. $E\left\{\widehat{Z}(\mathbf{s}_0)\right\}=E\{Z(\mathbf{s}_0)\}=m(\mathbf{s}_0)$, is ensured by imposing the constraints $\mathbf{F}'\mathbf{w}_0^{\mathrm{Z}}=\mathbf{f}_0$ on the UKZ weights for every prediction location $\mathbf{s}_0$ (Chilès and Delfiner 1999).

The UKZ weights vector $\mathbf{w}_0^{\mathrm{Z}}$ is obtained by solving the following system of equations (Chilès and Delfiner 1999):

$$\begin{bmatrix}\boldsymbol{\Sigma}_{\mathrm{R}} & \mathbf{F}\\ \mathbf{F}' & \mathbf{O}\end{bmatrix}\begin{bmatrix}\mathbf{w}_0^{\mathrm{Z}}\\ -\mathbf{t}_0^{\mathrm{Z}}\end{bmatrix}=\begin{bmatrix}\boldsymbol{\sigma}_0^{\mathrm{R}}\\ \mathbf{f}_0\end{bmatrix},$$

where $\mathbf{t}_0^{\mathrm{Z}}=\left[t_k^{\mathrm{Z}}(\mathbf{s}_0),\ k=1,\ \ldots,\ K\right]'$ denotes a $(K\times 1)$ vector of Lagrange multipliers, and $\sigma_0^{\mathrm{R}}=[\sigma_{\mathrm{R}}(\mathbf{s}_0,\mathbf{s}_i),\ i=1,\ \ldots,\ n]'$ denotes an $(n\times 1)$ vector of residual covariance values between the prediction location $\mathbf{s}_0$ and all n data locations.

The solution to the above system of equations can again be derived analytically using the inverse of a partitioned matrix:

$$\begin{bmatrix}\mathbf{w}_0^{\mathrm{Z}}\\ -\mathbf{t}_0^{\mathrm{Z}}\end{bmatrix}=\begin{bmatrix}\boldsymbol{\Sigma}_{\mathrm{R}}^{-1}-\boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{F}\mathbf{A}\mathbf{F}'\boldsymbol{\Sigma}_{\mathrm{R}}^{-1} & \boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{F}\mathbf{A}\\ \mathbf{A}\mathbf{F}'\boldsymbol{\Sigma}_{\mathrm{R}}^{-1} & -\mathbf{A}\end{bmatrix}\begin{bmatrix}\boldsymbol{\sigma}_0^{\mathrm{R}}\\ \mathbf{f}_0\end{bmatrix}. \tag{27}$$

From equation (27), the resulting UKZ weights vector $\mathbf{w}_0^{\mathrm{Z}}$ can be written as

$$\mathbf{w}_0^{\mathrm{Z}}=\boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{F}\mathbf{A}\mathbf{f}_0+\left(\mathbf{I}-\boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{F}\mathbf{A}\mathbf{F}'\right)\boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\boldsymbol{\sigma}_0^{\mathrm{R}}, \tag{28}$$

and consequently the UKZ attribute prediction $\widehat{z}(\mathbf{s_0})$ of equation (26) becomes

$$\begin{aligned}\widehat{z}(\mathbf{s}_0) &= \left(\mathbf{w}_0^{\mathrm{Z}}\right)' \mathbf{z} = \mathbf{f}_0' \mathbf{A}\mathbf{F}' \boldsymbol{\Sigma}_{\mathrm{R}}^{-1} \mathbf{z} + \left(\boldsymbol{\sigma}_0^{\mathrm{R}}\right)' \boldsymbol{\Sigma}_{\mathrm{R}}^{-1} \left(\mathbf{I} - \mathbf{F}\mathbf{A}\mathbf{F}' \boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\right) \mathbf{z} \\ &= \mathbf{f}_0' \widehat{\mathbf{b}} + \left(\boldsymbol{\lambda}_0^{\mathrm{R}}\right)' \left(\mathbf{z} - \mathbf{F}\widehat{\mathbf{b}}\right) = \widehat{m}(\mathbf{s}_0) + \widehat{r}(\mathbf{s}_0),\end{aligned} \tag{29}$$

where $\boldsymbol{\lambda}_0^{\mathrm{R}} = \boldsymbol{\Sigma}_{\mathrm{R}}^{-1} \boldsymbol{\sigma}_0^{\mathrm{R}}$ is the $(n \times 1)$ vector of Simple Kriging (SK) weights for predicting the unknown residual $r(\mathbf{s}_0)$ at location $\mathbf{s}_0$ from vector $\widehat{\mathbf{r}} = \mathbf{z} - \mathbf{F}\widehat{\mathbf{b}}$ containing 'data' residuals at the n sample locations.

Equation (29) entails that the UKZ attribute prediction $\widehat{z}(\mathbf{s}_0)$ at location $\mathbf{s}_0$ is composed of two distinct contributions: (i) a contribution $\widehat{m}(\mathbf{s}_0) = \mathbf{f}_0' \widehat{\mathbf{b}}$ linked to the predictor variables, and thus incorporating first-order effects in the form of an expected surface – see equation (23); and (ii) a contribution $\widehat{r}(\mathbf{s}_0)$ linked to the the covariogram model used to populate matrix $\boldsymbol{\Sigma}_{\mathrm{R}}$ and vector $\boldsymbol{\sigma}_0^{\mathrm{R}}$. When no drift is considered (or, better stated, when the the mean component $\mathbf{m}$ is assumed constant), the smoothness of the predicted UKZ attribute surface is dictated only by the covariogram model of the residual component $\mathbf{r}$. In the over-determined case ($n > K$), the UKZ predictions derived from equation (29) reproduce the data values at their locations; in other words, *UKZ is an interpolator not a smoother, no matter what the residual covariance matrix* $\boldsymbol{\Sigma}_{\mathrm{R}}$ *is.*

From equation (27), the vector of UKZ Lagrange multipliers can be written as $\mathbf{t}_0^{\mathrm{Z}} = \mathbf{A}\mathbf{f}_0 - \mathbf{A}\mathbf{F}'\boldsymbol{\lambda}_0^{\mathrm{R}} = \mathbf{A}\left(\mathbf{f}_0 - \mathbf{F}'\boldsymbol{\lambda}_0^{\mathrm{R}}\right)$, and consequently the UKZ prediction error variance $\widehat{\sigma}_{\mathrm{Z}}(\mathbf{s_0})$ at location $\mathbf{s}_0$ becomes

$$\begin{aligned}\widehat{\sigma}_{\mathrm{Z}}(\mathbf{s}_0) &= \sigma_{\mathrm{R}}(\mathbf{s}_0) - \left(\mathbf{w}_0^{\mathrm{Z}}\right)' \boldsymbol{\sigma}_0^{\mathrm{R}} + \left(\mathbf{t}_0^{\mathrm{Z}}\right)' \mathbf{f}_0 \\ &= \left[\sigma_{\mathrm{R}}(\mathbf{s}_0) - \left(\boldsymbol{\lambda}_0^{\mathrm{R}}\right)' \boldsymbol{\sigma}_0^{\mathrm{R}}\right] + \left[\mathbf{f}_0 - \mathbf{F}'\boldsymbol{\lambda}_0^{\mathrm{R}}\right]' \mathbf{A} \left[\mathbf{f}_0 - \mathbf{F}'\boldsymbol{\lambda}_0^{\mathrm{R}}\right],\end{aligned} \tag{30}$$

where $\sigma_{\mathrm{R}}(\mathbf{s}_0) - \left(\boldsymbol{\lambda}_0^{\mathrm{R}}\right)' \boldsymbol{\sigma}_0^{\mathrm{R}}$ is the SK error variance for predicting the unknown residual $r(\mathbf{s}_0)$, and $\sigma_{\mathrm{R}}(\mathbf{s}_0)$ is the variance of that residual at $\mathbf{s}_0$ prior to accounting for nearby data.

Note that, in the last term of equation (30), one needs to subtract the product $\mathbf{F}'\boldsymbol{\lambda}_0^{\mathrm{R}}$ from vector $\mathbf{f}_0$. If this adjustment was (incorrectly) not performed, then the second term would identify the UKM prediction error variance $\widehat{\sigma}_{\mathrm{M}}(\mathbf{s}_0) = \mathbf{f}_0' \mathbf{A} \mathbf{f}_0$ of the mean component $m(\mathbf{s}_0)$, and the corresponding UKZ variance $\widehat{\sigma}_{\mathrm{Z}}(\mathbf{s}_0)$ would (incorrectly) be different from zero at the sample locations. Instead, the adjustment term $\mathbf{F}'\boldsymbol{\lambda}_0^{\mathrm{R}}$ reduces the contribution of $\mathbf{f}_0$ in the term $\left[\mathbf{f}_0 - \mathbf{F}'\boldsymbol{\lambda}_0^{\mathrm{R}}\right] \boldsymbol{A} \left[\mathbf{f}_0 - \mathbf{F}'\boldsymbol{\lambda}_0^{\mathrm{R}}\right]$ as a function of the distance of the prediction location $\mathbf{s}_0$ from the sample locations; that distance influence is embedded in the SK weights vector $\boldsymbol{\lambda}_0^{\mathrm{R}}$. In general, when the prediction location $\mathbf{s}_0$ is close to sample locations, the entries of $\boldsymbol{\lambda}_0^{\mathrm{R}}$ tend to be larger than when $\mathbf{s}_0$ is far away from sample locations. Consequently, $\mathbf{F}'\boldsymbol{\lambda}_0^{\mathrm{R}}$ tends to increase in the former case, which entails that $\left[\mathbf{f}_0 - \mathbf{F}'\boldsymbol{\lambda}_0^{\mathrm{R}}\right]$ decreases, thus lowering the contribution of term $\left[\mathbf{f}_0 - \mathbf{F}'\boldsymbol{\lambda}_0^{\mathrm{R}}\right]' \mathbf{A} \left[\mathbf{f}_0 - \mathbf{F}'\boldsymbol{\lambda}_0^{\mathrm{R}}\right]$ in the final value of $\widehat{\sigma}_{\mathrm{Z}}(\mathbf{s}_0)$. If the prediction location $\mathbf{s}_0$ coincides with a sample location $\mathbf{s}_i$, then vector $\boldsymbol{\lambda}_0^{\mathrm{R}}$ becomes $\boldsymbol{\lambda}_i^{\mathrm{R}}$ with entries $\lambda_i^{\mathrm{R}}(\mathbf{s}_i) = 1$, and $\lambda_j^{\mathrm{R}}(\mathbf{s}_i) = 0, \forall j \neq i$, in which case $\mathbf{F}'\boldsymbol{\lambda}_i^{\mathrm{R}} = \mathbf{f}_i$. Since the SK variance is always zero at the sample locations, because $\left(\boldsymbol{\lambda}_i^{\mathrm{R}}\right)' \boldsymbol{\sigma}_i^{\mathrm{R}} = \sigma_{\mathrm{R}}(\mathbf{s}_i)$, the composite UKZ variance $\widehat{\sigma}_{\mathrm{Z}}(\mathbf{s}_i)$ also becomes zero. As opposed to the UKM prediction error variance $\widehat{\sigma}_{\mathrm{M}}(\mathbf{s}_0)$ for the mean component, the UKZ prediction error variance $\widehat{\sigma}_{\mathrm{Z}}(\mathbf{s_0})$ for the attribute itself attains its maximum away from sample locations, and decreases as the prediction location $\mathbf{s}_0$ lies closer to sample locations.

6.3 *Spatial prediction under Model III*

Lastly, the UKF signal prediction $\widehat{p}(\mathbf{s}_0)$ at location $\mathbf{s}_0$ is again expressed as a weighted linear combination of the entries of the data vector $\mathbf{z}$:

$$\widehat{p}(\mathbf{s}_0) = \left(\mathbf{w}_0^P\right)' \mathbf{z}, \tag{31}$$

where $\mathbf{w}_0^P = \left[\mathbf{w}_i^P(\mathbf{s}_0),\, i=1,\, \ldots,\, n\right]'$ denotes the $(n \times 1)$ vector of UKF weights.

Unbiasedness of prediction, i.e. $E\left\{\widehat{P}(\mathbf{s}_0)\right\} = E\{P(\mathbf{s}_0)\} = m(\mathbf{s}_0)$, is ensured by imposing the constraints $\mathbf{F}'\mathbf{w}_0^P = \mathbf{f}_0$ on the UKF weights for every prediction location $\mathbf{s}_0$ (Chilès and Delfiner 1999).

The UKF weights vector $\mathbf{w}_0^P$ is obtained by solving the following system of equations (Chilès and Delfiner 1999):

$$\begin{bmatrix} \boldsymbol{\Sigma}_{\mathrm{R}} & \mathbf{F} \\ \mathbf{F}' & \mathbf{O} \end{bmatrix} \begin{bmatrix} \mathbf{w}_0^P \\ -\mathbf{t}_0^P \end{bmatrix} = \begin{bmatrix} \boldsymbol{\sigma}_0^Q \\ \mathbf{f}_0 \end{bmatrix},$$

where $\mathbf{t}_0^P = \left[t_k^P(\mathbf{s}_0),\, k=1,\, \ldots,\, K\right]'$ denotes a $(K \times 1)$ vector of UKF Lagrange multipliers, and $\boldsymbol{\sigma}_0^Q = \left[\sigma_Q(\mathbf{s}_0, \mathbf{s}_i),\, i=1,\, \ldots,\, n\right]'$ is an $(n \times 1)$ vector containing q-residual covariance values between $\mathbf{s}_0$ and all n sample locations. Note that $\boldsymbol{\sigma}_0^Q = \boldsymbol{\sigma}_0^{\mathrm{R}} - \boldsymbol{\sigma}_0^E$, with $\boldsymbol{\sigma}_0^E = [\sigma_E(\mathbf{s}_0, \mathbf{s}_i),\, i=1,\, \ldots,\, n]'$ being an $(n \times 1)$ vector of e-residual covariance values between $\mathbf{s}_0$ and all sample locations. This relationship holds because the two residual components $\mathbf{q}$ and $\mathbf{e}$ are assumed uncorrelated (Chilès and Delfiner 1999).

The solution to the above system of equations can again be derived analytically using the inverse of a partitioned matrix:

$$\begin{bmatrix} \mathbf{w}_0^P \\ -\mathbf{t}_0^P \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathrm{R}}^{-1} - \boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{F}\mathbf{A}\mathbf{F}'\boldsymbol{\Sigma}_{\mathrm{R}}^{-1} & \boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{F}\mathbf{A} \\ \mathbf{A}\mathbf{F}'\boldsymbol{\Sigma}_{\mathrm{R}}^{-1} & -\mathbf{A} \end{bmatrix} \begin{bmatrix} \boldsymbol{\sigma}_0^Q \\ \mathbf{f}_0 \end{bmatrix}. \tag{32}$$

From equation (32), the resulting UKF weights vector $\mathbf{w}_0^P$ can be written as

$$\mathbf{w}_0^P = \boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{F}\mathbf{A}\mathbf{f}_0 + \left(\mathbf{I} - \boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{F}\mathbf{A}\mathbf{F}'\right)\boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\boldsymbol{\sigma}_0^Q, \tag{33}$$

entailing that the UKF prediction $\widehat{p}(\mathbf{s}_0)$ of equation (31) becomes

$$\begin{aligned} \widehat{p}(\mathbf{s}_0) &= \left(\mathbf{w}_0^P\right)'\mathbf{z} = \mathbf{f}'_0\mathbf{A}\mathbf{F}'\boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{z} + \left(\boldsymbol{\sigma}_0^Q\right)'\boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\left(\mathbf{I} - \mathbf{F}\mathbf{A}\mathbf{F}'\boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\right)\mathbf{z} \\ &= \mathbf{f}'_0\widehat{\mathbf{b}} + \left(\boldsymbol{\lambda}_0^Q\right)'\left(\mathbf{z} - \mathbf{F}\widehat{\mathbf{b}}\right) = \widehat{m}(\mathbf{s}_0) + \widehat{q}(\mathbf{s}_0) \\ &= \mathbf{f}'_0\widehat{\mathbf{b}} + \left(\boldsymbol{\lambda}_0^{\mathrm{R}} - \boldsymbol{\lambda}_0^E\right)'\left(\mathbf{z} - \mathbf{F}\widehat{\mathbf{b}}\right) = \mathbf{f}'_0\widehat{\mathbf{b}} + \left(\boldsymbol{\lambda}_0^{\mathrm{R}}\right)'\left(\mathbf{z} - \mathbf{F}\widehat{\mathbf{b}}\right) - \left(\boldsymbol{\lambda}_0^E\right)'\left(\mathbf{z} - \mathbf{F}\widehat{\mathbf{b}}\right) \\ &= \widehat{m}(\mathbf{s}_0) + \widehat{z}(\mathbf{s}_0) - \widehat{e}(\mathbf{s}_0), \end{aligned} \tag{34}$$

where $\boldsymbol{\lambda}_0^Q = \boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\boldsymbol{\sigma}_0^Q$ and $\boldsymbol{\lambda}_0^E = \boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\boldsymbol{\sigma}_0^E$ denote $(n \times 1)$ vectors of Simple coKriging (SCK) weights for predicting the unknown q-residual $q(\mathbf{s}_0)$ and measurement error $e(\mathbf{s}_0)$, respectively, at location $\mathbf{s}_0$ from vector $\widehat{\mathbf{r}} = \mathbf{z} - \mathbf{F}\widehat{\mathbf{b}}$ containing r-residuals at the n sample locations. Relationship $\boldsymbol{\lambda}_0^Q = \boldsymbol{\lambda}_0^{\mathrm{R}} - \boldsymbol{\lambda}_0^E$ is obtained from $\boldsymbol{\sigma}_0^Q = \boldsymbol{\sigma}_0^{\mathrm{R}} - \boldsymbol{\sigma}_0^E$.

Equation (34) entails that the UKF prediction can be seen as a UKZ prediction $\widehat{m}(\mathbf{s}_0)+\widehat{r}(\mathbf{s}_0)$ minus a prediction $\widehat{e}(\mathbf{s}_0)$ of the measurement error. In other words, one performs UKZ as if the objective were to predict the noisy attribute value $z(\mathbf{s}_0)$, and then subtracts from that prediction $\widehat{z}(\mathbf{s}_0)$ a prediction of the measurement error at the same location. This latter prediction $\widehat{e}(\mathbf{s}_0)$ is again derived from the residual vector $\mathbf{r}=\mathbf{z}-\mathbf{Fb}$, since the only data available are the measurements in vector $\mathbf{z}$. As opposed to the UKZ prediction $\widehat{z}(\mathbf{s}_0)$, the UKF prediction $\widehat{p}(bfs_0)$ does not reproduce the measurements at their sampling locations. More precisely, when the prediction location $\mathbf{s}_0$ coincides with a sample location $\mathbf{s}_i$, then $\widehat{z}(\mathbf{s}_i)=\mathbf{z}(\mathbf{s}_i)$ but $\widehat{p}(\mathbf{s}_i)=\mathbf{z}(\mathbf{s}_i)-\widehat{e}(\mathbf{s}_i)$. This smoothing (not interpolating) property of UKF at the sample locations is precisely a consequence of filtering out the measurement error from the corresponding UKZ prediction. In the over-determined case ($n>K$), the UKF predictions derived from equation (34) *do not* reproduce the data values at their locations; in other words, *UKF is generally a smoother, not an interpolator*.

From equation (32), the vector of UKF Lagrange multipliers can be written as $\mathbf{t}_0^P=\mathbf{A}\mathbf{f}_0-\mathbf{A}\mathbf{F}'\boldsymbol{\lambda}_0^Q=\mathbf{A}\left[\mathbf{f}_0-\mathbf{F}'\boldsymbol{\lambda}_0^Q\right]$, and consequently the UKF prediction error variance $\widehat{\sigma}_P(\mathbf{s}_0)$ at location $\mathbf{s}_0$ becomes

$$\begin{aligned}\widehat{\sigma}_P(\mathbf{s}_0)&=\sigma_Q(\mathbf{s}_0)-\left(\mathbf{w}_0^P\right)'\boldsymbol{\sigma}_0^Q+\left(\mathbf{t}_0^P\right)'\mathbf{f}_0\\&=\left[\sigma_Q(\mathbf{s}_0)-\left(\boldsymbol{\lambda}_0^Q\right)'\boldsymbol{\sigma}_0^Q\right]+\left[\mathbf{f}_0-\mathbf{F}'\boldsymbol{\lambda}_0^Q\right]'\mathbf{A}\left[\mathbf{f}_0-\mathbf{F}'\boldsymbol{\lambda}_0^Q\right],\end{aligned}\tag{35}$$

where $\sigma_Q(\mathbf{s}_0)$ is the variance of the residual q-component at location $\mathbf{s}_0$ prior to accounting for nearby data.

As opposed to the UKM prediction error variance $\widehat{\sigma}_M(\mathbf{s}_0)$ given in equation (25), which increases towards the sample locations, the UKF variance $\widehat{\sigma}_P(\mathbf{s}_0)$ decreases as the prediction location $\mathbf{s}_0$ approaches the sample locations; this is in agreement with the general spatial pattern of the UKZ prediction error variance $\widehat{\sigma}_Z(\mathbf{s}_0)$ given in equation (30). The UKF variance, however, is generally smaller than the UKZ variance, i.e. $\widehat{\sigma}_P(\mathbf{s_0})<\widehat{\sigma}_Z(\mathbf{s}_0)$, because the signal is less variable than the attribute (since the latter has the added variability of measurement error). This is not true at the sample locations, where UKF variance $\widehat{\sigma}_P(\mathbf{s}_0)$ *does not* become zero, precisely owing to the variance of the measurement error. In other words, since the data are corrupted by measurement error, the prediction error variance at sample locations is greater than zero, indicating a lack of confidence in the measurements.

7. Particular cases of the geostatistical formulation

In the completely determined case, i.e. when the number of parameters equals the number of observations ($K=n$), all equations given in the previous section attain simplified versions. With these simplifications, and under appropriate definitions of matrix $\mathbf{F}$ and vector $\mathbf{f}_0$ (i.e. using appropriate predictors), 1D linear interpolation, TIN and bilinear interpolation, can be regarded as particular cases of the general geostatistical framework.

More precisely, in the case of 1D piecewise linear interpolation (see figure 1), $n=2$, $K=2$, and equation (17) becomes

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}.$$

Similarly, in the case of linear interpolation based on three TIN vertices (see figure 2), $n=3$, $K=3$, and equation (17) becomes

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} + \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix}.$$

Lastly, in the case of bilinear interpolation in a rectangle (see figure 4), $n=4$, $K=4$, and equation (17) becomes

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} = \begin{bmatrix} 1 & x_1 & y_1 & x_1y_1 \\ 1 & x_1 & y_2 & x_1y_2 \\ 1 & x_2 & y_2 & x_2y_2 \\ 1 & x_2 & y_1 & x_2y_1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} + \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{bmatrix}.$$

In what follows, we prove that in the completely determined case and when the predictors are the particular functions of coordinates given above: (i) the weights derived via UKM, UKZ, and UKF, are all *identical*; and (ii) these geostatistically derived weights *coincide* with the weights traditionally used for 1D linear interpolation, TIN and bilinear interpolation. Naturally, if different predictors were adopted, the above statement would not hold, because the definition of the expected surface would be different. It is beyond the scope of this paper to derive prediction error variances for such cases.

7.1 *Kriging weights*

Consider the UKM weights vector $\mathbf{w}_0^{\mathrm{M}}$ given explicitly in equation (22). In the completely determined case ($n=K$), matrix $\mathbf{F}$ is not any more of size $(n \times K)$ but of size $(K \times K)$, and thus can be inverted directly since it is square (provided it is of full rank). Consequently, equation (22) can be simplified to

$$\mathbf{w}_0^{\mathrm{M}} = \boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{F}\mathbf{F}^{-1}\boldsymbol{\Sigma}_{\mathrm{R}}(\mathbf{F}')^{-1}\mathbf{f}_0 = (\mathbf{F}')^{-1}\mathbf{f}_0,$$

which entails that the UKM weights *do not* depend on the measurement error (or residual) covariance matrix $\boldsymbol{\Sigma}_{\mathrm{R}}=\boldsymbol{\Sigma}_E$.

Similarly, equation (28) giving the UKZ weights vector $\mathbf{w}_0^{\mathrm{Z}}$ simplifies to

$$\mathbf{w}_0^{\mathrm{Z}} = \mathbf{w}_0^{\mathrm{M}} + \left[\mathbf{I} - \boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{F}\mathbf{F}^{-1}\boldsymbol{\Sigma}_{\mathrm{R}}(\mathbf{F}')^{-1}\mathbf{F}'\right]\boldsymbol{\lambda}_0^{\mathrm{R}} = \mathbf{w}_0^{\mathrm{M}} + [\mathbf{I}-\mathbf{I}]\boldsymbol{\lambda}_0^{\mathrm{R}} = \mathbf{w}_0^{\mathrm{M}},$$

which entails that the UKZ weights also *do not* depend on the residual covariance matrix $\boldsymbol{\Sigma}_{\mathrm{R}}$, and are equal to the UKM weights.

Lastly, equation (33) giving the UKF weights vector $\mathbf{w}_0^{P}$ simplifies to

$$\mathbf{w}_0^{P} = \mathbf{w}_0^{\mathrm{M}} + \left[\mathbf{I} - \boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{F}\mathbf{F}^{-1}\boldsymbol{\Sigma}_{\mathrm{R}}(\mathbf{F}')^{-1}\mathbf{F}'\right]\boldsymbol{\lambda}_0^{Q} = [\mathbf{I}-\mathbf{I}]\boldsymbol{\lambda}_0^{Q} = \mathbf{w}_0^{\mathrm{M}},$$

which entails that the UKF weights also *do not* depend on the residual covariance matrix $\boldsymbol{\Sigma}_R$, and are equal to the UKM and UKZ weights.

We hereafter demonstrate that, in the completely determined case and for the particular predictors used above, the Universal Kriging weights coincide with the weights used in classical 1D linear interpolation, TIN interpolation and bilinear interpolation. In what follows, we denote as $\mathbf{w}=[w_i,\ i=1,\ \ldots,\ n]'$ the generic weights vector obtained by UKM, UKZ or UKF. For notational simplicity, we denote the prediction location as $\mathbf{s}$, and we do not explicate the dependence of vectors $\mathbf{w}$ and $\mathbf{f}$ on $\mathbf{s}$. Our demonstration is based on Cramer's rule, which entails that $\mathbf{w}$ can be generally computed as

$$\begin{bmatrix} w_1 \\ \vdots \\ w_i \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} \frac{|\mathbf{F}'_{-1}|}{|\mathbf{F}'|} \\ \vdots \\ \frac{|\mathbf{F}'_{-i}|}{|\mathbf{F}'|} \\ \vdots \\ \frac{|\mathbf{F}'_{-n}|}{|\mathbf{F}'|} \end{bmatrix}, \tag{36}$$

where $|\cdot|$ denotes a matrix determinant, and $\mathbf{F}'_{-i}$ denotes a matrix formed by replacing the ith column of $\mathbf{F}'$ by vector $\mathbf{f}$.

For 1D linear interpolation, the vector $\mathbf{w}$ of UK weights is computed from equation (36) as

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \frac{1}{|\mathbf{F}'|} \begin{bmatrix} |\mathbf{F}'_{-1}| \\ |\mathbf{F}'_{-2}| \end{bmatrix} = \frac{1}{\begin{vmatrix} 1 & 1 \\ x_1 & x_2 \end{vmatrix}} \begin{bmatrix} \begin{vmatrix} 1 & 1 \\ x & x_2 \end{vmatrix} \\ \begin{vmatrix} 1 & 1 \\ x_1 & x \end{vmatrix} \end{bmatrix} = \frac{1}{(x_2 - x_1)} \begin{bmatrix} x_2 - x \\ x - x_1 \end{bmatrix},$$

which is precisely the vector of weights used for 1D linear interpolation. Note that, in this case, the determinant $|\mathbf{F}'|$ is equal to the length of segment $\mathbf{s}_1$, $\mathbf{s}_2$. Similarly, $|\mathbf{F}'_{-1}|$ is the length of segment $\mathbf{s}_2$, $\mathbf{s}$, and $|\mathbf{F}'_{-2}|$ is the length of segment $\mathbf{s}$, $\mathbf{s}_1$.

For TIN interpolation, the vector $\mathbf{w}$ of UK weights is computed from equation (36) as

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \frac{1}{|\mathbf{F}'|} \begin{bmatrix} |\mathbf{F}'_{-1}| \\ |\mathbf{F}'_{-2}| \\ |\mathbf{F}'_{-3}| \end{bmatrix} = \frac{1}{\begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}} \begin{bmatrix} \begin{vmatrix} 1 & 1 & 1 \\ x & x_2 & x_3 \\ y & y_2 & y_3 \end{vmatrix} \\ \begin{vmatrix} 1 & 1 & 1 \\ x_1 & x & x_3 \\ y_1 & y & y_3 \end{vmatrix} \\ \begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x \\ y_1 & y_2 & y \end{vmatrix} \end{bmatrix} = \frac{1}{a} \begin{bmatrix} a_{-1} \\ a_{-2} \\ a_{-3} \end{bmatrix},$$

where $a=|\mathbf{F}'|/2$ denotes the area of the TIN element, and $a_{-i}=|\mathbf{F}'_{-i}|/2$ denotes the area of the sub-triangle formed by the prediction location $\mathbf{s}$ and the two TIN vertices excluding $\mathbf{s}_i$; these are precisely the weights used for TIN interpolation.

For bilinear interpolation, the vector $\mathbf{w}$ of UK weights is computed from equation (36):

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} = \frac{1}{|\mathbf{F}'|} \begin{bmatrix} |\mathbf{F}'_{-1}| \\ |\mathbf{F}'_{-2}| \\ |\mathbf{F}'_{-3}| \\ |\mathbf{F}'_{-4}| \end{bmatrix} = \frac{1}{\begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_1 & x_2 & x_2 \\ y_1 & y_2 & y_2 & y_1 \\ x_1y_1 & x_1y_2 & x_2y_2 & x_2y_1 \end{vmatrix}} \begin{bmatrix} \begin{vmatrix} 1 & 1 & 1 & 1 \\ x & x_1 & x_2 & x_2 \\ y & y_2 & y_2 & y_1 \\ xy & x_1y_2 & x_2y_2 & x_2y_1 \end{vmatrix} \\ \begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 & x & x_2 & x_2 \\ y_1 & y & y_2 & y_1 \\ x_1y_1 & xy & x_2y_2 & x_2y_1 \end{vmatrix} \\ \begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_1 & x & x_2 \\ y_1 & y_2 & y & y_1 \\ x_1y_1 & x_1y_2 & xy & x_2y_1 \end{vmatrix} \\ \begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_1 & x_2 & x \\ y_1 & y_2 & y_2 & y \\ x_1y_1 & x_1y_2 & x_2y_2 & xy \end{vmatrix} \end{bmatrix}.$$

Let us focus, without loss of generality, on the first element w_1 of the solution weights vector. After calculating the determinants $|\mathbf{F}'|$ and $|\mathbf{F}'_{-1}|$, and simplifying the resulting expressions, one obtains

$$w_1 = \frac{x_2y_2 - x_2y - xy_2 + xy}{x_1y_1 + x_2y_2 - x_1y_2 - x_2y_1} = \frac{a_{-1}}{a}. \tag{37}$$

Indeed, the numerator of equation (37) is the area a_{-1} of the sub-rectangle formed by the prediction location $\mathbf{s}=(x, y)$, the two points $\mathbf{s}_T=(x, y_2)$, $\mathbf{s}_R=(x_2, y)$, and vertex $\mathbf{s}_3=(x_2, y_2)$ opposite to $\mathbf{s}_1$ (see figure 4):

$$a_{-1} = \|\mathbf{s}_R - \mathbf{s}\| \|\mathbf{s}_T - \mathbf{s}\| = (x_2 - x)(y_2 - y) = x_2y_2 - x_2y - xy_2 + xy.$$

Similarly, the denominator of equation (37) is the total area a of the rectangle element (see figure 4):

$$a = \|\mathbf{s}_4 - \mathbf{s}_1\| \|\mathbf{s}_2 - \mathbf{s}_1\| = (x_2 - x_1)(y_2 - y_1) = x_2y_2 - x_2y_1 - x_1y_2 + x_1y_1.$$

In other words, for bilinear interpolation, the UK weight w_1 assigned to the measurement z_1 at the rectangle vertex $\mathbf{s}_1$ is proportional to the area a_{-1} of the sub-rectangle formed by the prediction location $\mathbf{s}$, points $\mathbf{s}_T$, $\mathbf{s}_R$ and the vertex $\mathbf{s}_3$ opposite to $\mathbf{s}_1$. The proportionality constant is the total rectangle area a. Analogous results can be obtained for the other weights w_2, w_3 and w_4.

It is also straightforward to corroborate that the weights used for 1D linear interpolation, TIN interpolation and bilinear interpolation satisfy the UK

constraints $\mathbf{F}'\mathbf{w}=\mathbf{f}$, using the appropriate definitions of $\mathbf{F}$, $\mathbf{w}$ and $\mathbf{f}$. Indeed, for 1D linear interpolation

$$\mathbf{F}'\mathbf{w}=\begin{bmatrix}1 & 1\\ x_1 & x_2\end{bmatrix}\begin{bmatrix}\frac{x_2-x}{x_2-x_1}\\ \frac{x-x_1}{x_2-x_1}\end{bmatrix}=\begin{bmatrix}\frac{x_2-x+x-x_1}{x_2-x_1}\\ \frac{x_1x_2-xx_1+xx_2-x_1x_2}{x_2-x_1}\end{bmatrix}=\begin{bmatrix}1\\ x\end{bmatrix}=\mathbf{f},$$

and similar results can be obtained for the weights used for TIN and bilinear interpolation in 2D. This feature should be expected, since the weights vector $\mathbf{w}$ derived by UKM, UKZ or UKF satisfies the constraint $\mathbf{F}'\mathbf{w}=\mathbf{f}$ by construction.

In summary, we have demonstrated in this section that: *in the completely determined case and for the particular predictors adopted, the weights obtained via different variants of Universal Kriging coincide with the weights used in 1D linear interpolation, TIN interpolation and bilinear interpolation.*

7.2 *Kriging predictions*

It is trivial to deduce that, since all UK variants considered in this work yield identical weights in the completely determined case, the corresponding predictions given in equations (20), (26) and (31) will also be identical with each other. In what follows, we provide more insight into this equivalence of predictions based on their exact decompositions given in equations (23), (29) and (34).

Consider the vector $\widehat{\mathbf{b}}$ of UKM (or GLS) regression coefficients given in equation (24). In the completely determined case, that equation simplifies to

$$\widehat{\mathbf{b}}=\mathbf{F}^{-1}\boldsymbol{\Sigma}_{\mathrm{R}}(\mathbf{F}')^{-1}\mathbf{F}'\boldsymbol{\Sigma}_{\mathrm{R}}^{-1}\mathbf{z}=\mathbf{F}^{-1}\mathbf{z},$$

entailing that the estimated vector of regression coefficients $\widehat{\mathbf{b}}$ does not depend on the covariance $\boldsymbol{\Sigma}_{\mathrm{R}}=\boldsymbol{\Sigma}_E$ of the measurement error (or residual) component.

Consequently, the UKM predicted mean component $\widehat{\mathbf{m}}$ at the n sample locations (polygon vertices) can be written as

$$\widehat{\mathbf{m}}=\mathbf{F}\widehat{\mathbf{b}}=\mathbf{F}\mathbf{F}^{-1}\mathbf{z}=\mathbf{z},$$

entailing that the UKM-derived mean component reproduces (interpolates) the measurements $\mathbf{z}$ at their sampling locations, no matter what the covariance of the measurement error (or residual) component is. *This is a natural consequence of the completely determined case: in 1D, for example, a line will always pass through two points, and hence interpolate the corresponding measurements at the endpoints of a line segment*.

As stated in Section 6.1, all three forms of interpolation considered in this work can be viewed as variants of classical trend surface analysis. The sole difference is that classical trend surface fitting constitutes a global regression model with ($n>K$), and as such does not reproduce the data values at their sample locations. In other words, trend surface models are smoothers, not interpolators. On the contrary, the three forms of linear interpolation considered in this work constitute local regression models whose parameters are specific to a particular polygon. Because precisely the number of parameters (regression coefficients) of such local regression models are always equal to the number of data considered (measurements available at the vertices of each polygon), the resulting trend surfaces always identify (reproduce) the sample data at the polygon vertices. This feature has led to the (incorrect in our

opinion) classification of these three methods as interpolation methods, rather than regression methods. Perhaps a more appropriate term would be *local, interpolating, trend surface models*.

Consider now the UKZ and UKF predictions $\widehat{z}(\mathbf{s}_0)$ and $\widehat{p}(\mathbf{s}_0)$, given in equations (29) and (34). These expressions contain a common term $\widehat{m}(\mathbf{s}_0)$, and an extra term that is a function of the residual vector $\mathbf{z}-\mathbf{F}\widehat{\mathbf{b}}$. Since $\widehat{\boldsymbol{m}}=\boldsymbol{F}\widehat{\mathbf{b}}=\mathbf{z}$, that residual vector becomes zero, i.e. $\mathbf{z}-\mathbf{F}\widehat{\mathbf{b}}=\mathbf{0}$, entailing

$$\widehat{\mathbf{m}}=\widehat{\mathbf{z}}=\widehat{\mathbf{p}}=\mathbf{z},$$

i.e. *in the completely determined case, the predictions at sample locations (polygon vertices) obtained by any Universal Kriging variant reproduce the available measurements*.

Along the same lines, the UKZ and UKF predictions $\widehat{z}(\mathbf{s}_0)$ and $\widehat{p}(\mathbf{s}_0)$ at any arbitrary location $\mathbf{s}_0$ are equal to the UKM prediction $\widehat{m}(\mathbf{s}_0)$, since their respective SK and SCK contributions, the second terms of equations (29) and (34), are null. In other words, *in the completely determined case and for the particular predictors adopted, if one were to perform linear interpolation along a line segment, TIN interpolation or bilinear interpolation, one would obtain the same interpolation results as Universal Kriging (UKM, UKZ or UKF) irrespective of the covariances assumed for the residual and/or measurement error components*.

7.3 *Kriging prediction error variances*

As opposed to the predictions given by the different UK variants considered in this work, the associated prediction error variances differ from each other, simply because the prediction objectives in *Models I, II* and *III* are different owing to the different definitions of measurement error. In what follows, we explicitly derive these different prediction error variances for the completely determined case ($n=K$) and for the particular predictors adopted above.

More precisely, the UKM prediction error variance $\widehat{\sigma}_{\mathrm{M}}(\mathbf{s}_0)$ for the unknown mean component $m(\mathbf{s}_0)$ at location $\mathbf{s}_0$, given in equation (25), simplifies to

$$\widehat{\sigma}_{\mathrm{M}}(\mathbf{s}_0)=\mathbf{f}'_0\mathbf{A}\mathbf{f}_0=\mathbf{f}'_0\mathbf{F}^{-1}\boldsymbol{\Sigma}_{\mathrm{R}}(\mathbf{F}')^{-1}\mathbf{f}_0=\mathbf{w}'_0\boldsymbol{\Sigma}_{\mathrm{R}}\mathbf{w}_0 \tag{38}$$

with the latter equation being a quadratic in the weights $\mathbf{w}_0$, or in the coordinates $\mathbf{f}_0$ of the prediction location $\mathbf{s}_0$.

Equation (38) provides the basis for interpreting confidence intervals derived for points along a line segment, in a TIN element, or in a rectangle. As stated in Section 6.1, the prediction error variance for the mean component $\widehat{\sigma}_{\mathrm{M}}(\mathbf{s}_0)$ attains its minimum for that $\mathbf{f}_0$ vector whose entries are the mean values of each predictor variable. In 1D, for example, that variance is minimum for $\mathbf{f}_0=[1\ \bar{x}]'$, where $\bar{x}=(x_2-x_1)/2$ is the mean of the coordinates x_1 and x_2 of the two endpoints of a line segment. Conversely, the maximum error variance for the mean component is attained at the segment endpoints, since at those locations $\mathbf{f}_0=[1\ x_1]'$ and $\mathbf{f}_0=[1\ x_2]'$, hence the corresponding x-coordinates x_1 and x_2 are maximally different from $\bar{x}$.

Similarly, in the completely determined case, the UKZ prediction error variance $\widehat{\sigma}_{\mathrm{Z}}(\mathbf{s}_0)$ for the unknown attribute $z(\mathbf{s}_0)$ at location $\mathbf{s}_0$, given in equation (30),

becomes

$$\begin{aligned}\widehat{\sigma}_Z(\mathbf{s}_0) &= \left[\sigma_R(\mathbf{s}_0) - \left(\boldsymbol{\lambda}_0^R\right)' \boldsymbol{\sigma}_0^R\right] + \left[\mathbf{f}_0 - \mathbf{F}'\boldsymbol{\lambda}_0^R\right]' \mathbf{F}^{-1}\Sigma_R(\mathbf{F}')^{-1}\left[\mathbf{f}_0 - \mathbf{F}'\boldsymbol{\lambda}_0^R\right] \\ &= \left[\sigma_R(\mathbf{s}_0) - \left(\boldsymbol{\lambda}_0^R\right)' \boldsymbol{\sigma}_0^R\right] + \left[\mathbf{f}'_0\mathbf{F}^{-1} - \left(\boldsymbol{\lambda}_0^R\right)'\mathbf{F}\mathbf{F}^{-1}\right]\Sigma_R\left[(\mathbf{F}')^{-1}\mathbf{f}_0 - (\mathbf{F}')^{-1}\mathbf{F}'\boldsymbol{\lambda}_0^R\right] \quad (39) \\ &= \left[\sigma_R(\mathbf{s}_0) - \left(\boldsymbol{\lambda}_0^R\right)' \boldsymbol{\sigma}_0^R\right] + \left[\mathbf{w}_0 - \boldsymbol{\lambda}_0^R\right]' \Sigma_R\left[\mathbf{w}_0 - \boldsymbol{\lambda}_0^R\right].\end{aligned}$$

As with equation (30), the simplified version of UKZ variance for the completely determined case is the sum of the SK variance $\left[\sigma_R(\mathbf{s}_0) - \left(\boldsymbol{\lambda}_0^R\right)' \boldsymbol{\sigma}_0^R\right]$, plus a modified version $\left[\mathbf{w}_0 - \boldsymbol{\lambda}_0^R\right]' \Sigma_R\left[\mathbf{w} - \boldsymbol{\lambda}_0^R\right]$ of the UKM variance $\widehat{\sigma}_M(\mathbf{s}_0)$ for the mean component given in equation (38). That modification is again a function of distance between the prediction location $\mathbf{s}_0$ and the sample locations (again, that distance is embedded in the SK weights vector $\boldsymbol{\lambda}_0^R$). For example, when the prediction location $\mathbf{s}_0$ coincides with a polygon vertex $\mathbf{s}_i$ where a measurement is available, vector $\boldsymbol{\lambda}_0^R$ becomes $\boldsymbol{\lambda}_i^R$ with entries $\boldsymbol{\lambda}_i^R(\mathbf{s}_i) = 1$, and $\boldsymbol{\lambda}_j^R(\mathbf{s}_i) = 0$, $\forall j \neq i$. The corresponding weights vector $\mathbf{w}_0$ also becomes $\mathbf{w}_i$ with entries $w_i(\mathbf{s}_i)=1$, and $w_j(\mathbf{s}_i)=0$, $\forall j \neq i$, since the mean component is also interpolating the data. Consequently, $\left[\mathbf{w}_i - \boldsymbol{\lambda}_i^R\right] = 0$, thus cancelling the contribution of term $\left[\mathbf{w}_i - \boldsymbol{\lambda}_i^R\right]' \Sigma_R\left[\mathbf{w}_i - \boldsymbol{\lambda}_{i,}^R\right]$ to the UKZ variance of equation (39). Since the SK variance term $\left[\sigma_R(\mathbf{s}_i) - \left(\boldsymbol{\lambda}_i^R\right)' \boldsymbol{\sigma}_i^R\right]$ is also zero at a sample location $\mathbf{s}_i$, the final value for the UKZ variance $\widehat{\sigma}_Z(\mathbf{s}_i)$ is zero, too. Contrary to the UKM prediction error variance $\widehat{\sigma}_M(\mathbf{s}_0)$ given in equation (38), which decreases away from the polygon vertices, the UKZ prediction error variance $\widehat{\sigma}_Z(\mathbf{s}_0)$ attains its maximum away from those vertices, simply because the prediction objective in the latter case is that of the unknown attribute value $z(\mathbf{s}_0)$, not of the unknown mean component $m(\mathbf{s}_0)$.

Lastly, in the completely determined case, the UKF prediction error variance $\widehat{\sigma}_P(\mathbf{s}_0)$ for the unknown signal $p(\mathbf{s}_0) = m(\mathbf{s}_0) + q(\mathbf{s}_0)$ at location $\mathbf{s}_0$, given in equation (35), becomes

$$\widehat{\sigma}_P(\mathbf{s}_0) = \left[\sigma_Q(\mathbf{s}_0) - \left(\boldsymbol{\lambda}_0^Q\right)' \boldsymbol{\sigma}_0^Q\right] + \left[\mathbf{w}_0 - \boldsymbol{\lambda}_0^Q\right]' \Sigma_R\left[\mathbf{w}_0 - \boldsymbol{\lambda}_0^Q\right], \quad (40)$$

which is derived in a similar manner as equation (39).

The comments made for the UKZ prediction error variance $\widehat{\sigma}_Z(\mathbf{s}_0)$ of equation (39) apply here too. The only difference is that the UKF variance $\widehat{\sigma}_P(\mathbf{s}_0)$ does not attain a zero value at the polygon vertices owing to measurement error, and is lower than the UKZ variance $\widehat{\sigma}_Z(\mathbf{s}_0)$ apart from the polygon vertices.

Figure 5 gives examples of prediction error variance profiles obtained via: (i) UKM, (ii) UKZ where the measurement error component is zero (UKZ I), (iii) UKZ where the measurement error component is present (UKZ II), and (iv) UKF. In all cases, prediction is performed at points along a line segment with endpoints at $x_1 = 1$ and $x_2 = 11$. The residual covariogram model adopted for cases (i) and (ii) is a stationary exponential model $\sigma_R(h) = \sigma_R \exp(-3h/a)$, with sill (variance of residuals) $\sigma_R = 10$ and effective range $a = 10$; here h denotes the distance between any two locations along this line segment; note that for case (i), $\sigma_R(h) = \sigma_E(h)$. For cases (iii) and (iv), it is assumed that the covariogram model of the q-component is $\sigma_Q(h) = 5 \exp(-3h/10)$, and the error e-component is purely random (i.e. having a pure nugget effect as its variogram model) with stationary variance $\sigma_E = 5$. Note that the total variance (sill) of the residual r-component is still $\sigma_R = \sigma_Q + \sigma_E = 10$. The difference between cases (iii) and (iv) lies in the interpretation of the nugget effect: in

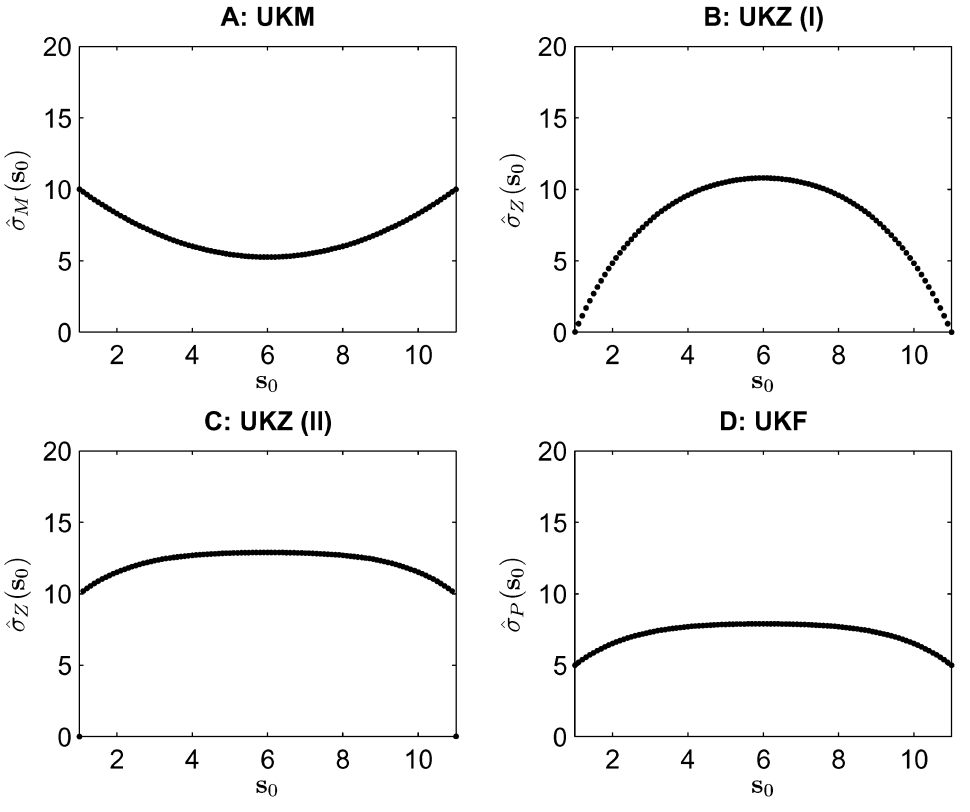

Figure 5. Examples of prediction error variance profiles along a line segment obtained using: (**A**) UKM; (**B**) UKZ (I); (**C**) UKZ (II); and (**D**) UKF. See text for details.

the latter case, that nugget effect is assumed to pertain to the error *e*-component of equation (17), whereas in the former it is assumed to pertain to the *q*-component of that equation. As stated in the previous sections, all four cases yield the *same* prediction profiles (not shown), no matter what the covariances or the actual measurements involved are.

From figure 5, one can easily appreciate the differences between the prediction error variances computed via UKM, UKZ and UKF. The UKZ and UKF variances of figures 5(B)–5(D) increase away from the sample locations (endpoints of the segment) and attain their maximum at the segment's midpoint. This behaviour is fundamentally different from that exhibited by the UKM variance profile of figure 5(A). As stated above, this is a consequence of the different prediction objectives for UKM: mean component $m(\mathbf{s}_0)$, UKZ: attribute $z(\mathbf{s}_0)$, and UKF: attribute minus measurement error $p(\mathbf{s}_0)$. Note that the difference between the two UKZ variance profiles shown in figures 5(B) (UKZ I) and 5(C) (UKZ II) is that the latter contains a nugget effect component, whereas the former does not. The UKZ II variance profile of figure 5(C) is discontinuous: it has zero values at the segment's endpoints, and immediately attains larger positive values ($\geqslant$10) at any interior point.

When compared to the UKZ variance profiles of figures 5(B)–5(C), the UKF variance profile of figure 5(D) does not reach a zero minimum at the segment's endpoints owing precisely to the explicit account of measurement error. In addition, the UKF variance profile does not reach the same maximum because the prediction

objectives are different in the two cases. In the case of UKF, the prediction objective is $\widehat{p}(\mathbf{s}_0) = \widehat{z}(\mathbf{s}_0) - \widehat{e}(\mathbf{s}_0)$, as opposed to its 'noisy' version $\widehat{z}(\mathbf{s}_0)$ predicted in UKZ. That signal component $p(\mathbf{s}_0)$ has less variance than its noisy version $z(\mathbf{s}_0)$, owing to the removal (filtering) of the measurement error. The difference between the UKZ and UKF variance profiles of figures 5(C) and 5(D), at locations other than the segment's endpoints, is precisely the variance of the measurement error $\sigma_E = 5$.

7.4 *Average prediction error variance for Model I*

We now need to correct the notation used in Sections 2 through 5 in order to explicate the prediction objective of *Model I*. More precisely, the predictor RV $\widehat{Z} = \widehat{Z}(\mathbf{s})$ used in these sections should be denoted as $\widehat{M} = \widehat{M}(\mathbf{s})$, since the prediction objective under *Model I* is the mean component $m = m(\mathbf{s})$ at any location $\mathbf{s}$, not the attribute $z(\mathbf{s})$ itself. Consequently, the predictor error variance $\widehat{\sigma}_{\mathrm{M}}(s)$ at location $\mathbf{s}$ is expanded as

$$\widehat{\sigma}_{\mathrm{M}}(\mathbf{s}) = \mathrm{Var}\left\{\widehat{M} - m\right\} = \mathrm{Var}\left\{\widehat{M}\right\} + \mathrm{Var}\{m\} - 2\,\mathrm{Cov}\left\{\widehat{M}, m\right\} = \mathrm{Var}\left\{\widehat{M}\right\}$$

since the true mean component $m(\mathbf{s})$ is assumed deterministic, hence it has zero variance and is uncorrelated with any stochastic quantities such as $\widehat{M} = \boldsymbol{\Sigma}_i w_i Z_i$. This entails that all the predictor variances given in Sections 2 through 5 are actually *prediction error* variances, and hence coincide with the UKM error variance $\widehat{\sigma}_{\mathrm{M}}(\mathbf{s})$ given in equation (38).

In what follows, we derive the average prediction error variance for the mean component $m(\mathbf{s})$ over an ensemble of random prediction locations. This is accomplished by taking the expected value of the UKM prediction error variance $\widehat{\sigma}_{\mathrm{M}}(s)$ of equation (38) with respect to the weights vector over that ensemble of prediction locations. Note that we hereafter drop the superscript 'M' and the subscript '0' from vector $\mathbf{w}_0^{\mathrm{M}}$ for notational simplicity.

More precisely, the expected value (average) of the quadratic expression in equation (38) can be computed as (Johnson and Wichern 2002)

$$E_{\mathbf{s}}\{\widehat{\sigma}_{\mathrm{M}}(\mathbf{s})\} = E\{\mathbf{w}'\boldsymbol{\Sigma}_{\mathrm{R}}\mathbf{w}\} = \mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{R}}\mathbf{V}_{\mathrm{W}}) + \mathbf{m}'_{\mathrm{W}}\boldsymbol{\Sigma}_{\mathrm{R}}\mathbf{m}_{\mathrm{W}}, \tag{41}$$

where tr $(\cdot)$ is the trace of a matrix, $\mathbf{V}_{\mathrm{W}}$ is the $(K \times K)$ variance–covariance matrix of the weights, and $\mathbf{m}_{\mathrm{W}}$ is the $(K \times 1)$ vector of expected values of the weights. Subscript $\mathbf{s}$ in the expectation indicates that averaging is performed over the ensemble of prediction locations.

In the next sections, it is also demonstrated that equation (41) includes as particular cases equation (4) for linear interpolation along a line segment, equation (10) for interpolation in a triangle, and equation (16) for bilinear interpolation in a rectangle, given appropriate definitions of $\mathbf{F}$ and $\mathbf{f}$ for each case.

7.4.1 Linear interpolation in 1D. For random prediction locations on a line segment, the resulting weights follow uniform distributions in [0, 1]. Consequently, their mean is 1/2 and their variance is 1/12. This entails

$$\mathbf{m}_{\mathrm{W}} = E\{\mathbf{w}\} = [1/2\ 1/2]'$$

and

$$\mathbf{V}_{\mathrm{W}} = \begin{bmatrix} \mathrm{Var}\{W_1\} & \mathrm{Cov}\{W_1, W_2\} \\ \mathrm{Cov}\{W_2, W_1\} & \mathrm{Var}\{W_2\} \end{bmatrix} = \begin{bmatrix} \frac{1}{12} & -\frac{1}{12} \\ -\frac{1}{12} & \frac{1}{12} \end{bmatrix} = \frac{1}{12}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix},$$

since $\mathrm{Cov}\{W, (1-W)\}$ can be computed as (see also equation 3)

$$\mathrm{Cov}\{W, (1-W)\} = E\{W(1-W)\} - E\{W\}E\{(1-W)\} = \frac{1}{6} - \frac{1}{4} = -\frac{1}{12}.$$

Therefore equation (41) becomes

$$\begin{aligned} E_s\{\widehat{\sigma}_{\mathrm{M}}(\mathbf{s})\} &= \frac{1}{12}\mathrm{tr}\left(\begin{bmatrix}\sigma^E_{11} & \sigma^E_{12}\\ \sigma^E_{21} & \sigma^E_{22}\end{bmatrix}\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}\right) + \frac{1}{4}\left([1\ 1]\begin{bmatrix}\sigma^E_{11} & \sigma^E_{12}\\ \sigma^E_{21} & \sigma^E_{22}\end{bmatrix}\begin{bmatrix}1\\ 1\end{bmatrix}\right)\\ &= \frac{1}{12}\left(\sigma^E_{11} + \sigma^E_{22} - 2\sigma^E_{12}\right) + \frac{1}{4}\left(\sigma^E_{11} + \sigma^E_{22} + 2\sigma^E_{12}\right)\\ &= \frac{1}{3}\left(\sigma^E_{11} + \sigma^E_{22} + \sigma^E_{22}\right), \end{aligned}$$

which is identical with equation (4) derived in Section 2. Superscript E in σ^E_{ij} explicates that the covariances involved are in fact those of the measurement error component $\mathbf{e}$ of equation (18).

7.4.2 TIN interpolation. For random prediction locations within a TIN element, the resulting weights follow triangular distributions in [0, 1] with modes at 0. Consequently, their mean is 1/3, their variance is 1/18, and their pairwise covariance is $-1/36$. This entails

$$\mathbf{m}_{\mathrm{W}} = E\{\mathbf{w}\} = [1/3\ 1/3\ 1/3]'$$

and

$$\mathbf{V}_{\mathrm{W}} = \begin{bmatrix}\mathrm{Var}\{W_1\} & \mathrm{Cov}\{W_1, W_2\} & \mathrm{Cov}\{W_1, W_3\}\\ \mathrm{Cov}\{W_2, W_1\} & \mathrm{Var}\{W_2\} & \mathrm{Cov}\{W_2, W_3\}\\ \mathrm{Cov}\{W_3, W_1\} & \mathrm{Cov}\{W_3, W_2\} & \mathrm{Var}\{W_3\}\end{bmatrix} = \frac{1}{18}\begin{bmatrix}1 & -1/2 & -1/2\\ -1/2 & 1 & -1/2\\ -1/2 & -1/2 & 1\end{bmatrix},$$

as per equation (7).

Therefore, equation (41) becomes

$$\begin{aligned} E_s\{\widehat{\sigma}_{\mathrm{M}}(\mathbf{s})\} &= \frac{1}{18}\mathrm{tr}\left(\begin{bmatrix}\sigma^E_{11} & \sigma^E_{12} & \sigma^E_{13}\\ \sigma^E_{21} & \sigma^E_{22} & \sigma^E_{23}\\ \sigma^E_{31} & \sigma^E_{32} & \sigma^E_{33}\end{bmatrix}\begin{bmatrix}1 & -1/2 & -1/2\\ -1/2 & 1 & -1/2\\ -1/2 & -1/2 & 1\end{bmatrix}\right)\\ &\quad + \frac{1}{9}\left([1\ 1\ 1]\begin{bmatrix}\sigma^E_{11} & \sigma^E_{12} & \sigma^E_{13}\\ \sigma^E_{21} & \sigma^E_{22} & \sigma^E_{23}\\ \sigma^E_{31} & \sigma^E_{32} & \sigma^E_{33}\end{bmatrix}\begin{bmatrix}1\\ 1\\ 1\end{bmatrix}\right)\\ &= \frac{1}{18}\left(\sigma^E_{11} + \sigma^E_{22} + \sigma^E_{33} - \sigma^E_{12} - \sigma^E_{13} - \sigma^E_{23}\right)\\ &\quad + \frac{1}{9}\left(\sigma^E_{11} + \sigma^E_{22} + \sigma^E_{33} + 2\sigma^E_{12} + 2\sigma^E_{13} + 2\sigma^E_{23}\right)\\ &= \frac{1}{6}\left(\sigma^E_{11} + \sigma^E_{22} + \sigma^E_{33} + \sigma^E_{12} + \sigma^E_{13} + \sigma^E_{23}\right), \end{aligned}$$

which is identical with equation (10) derived in Section 3.

7.4.3 Bilinear interpolation in a rectangle. Going back to equation (41), the entries of vector $\mathbf{m}_W=[E\{W_1\}\ E\{W_2\}\ E\{W_3\}\ E\{W_4\}]'$ are

$$E\{W_1\}=E\{\Phi\Psi\}=E\{\Phi\}E\{\Psi\}=\frac{1}{4}$$

$$E\{W_2\}=E\{\Phi(1-\Psi)\}=E\{\Phi\}E\{(1-\Psi)\}=\frac{1}{4}$$

$$E\{W_3\}=E\{(1-\Phi)(1-\Psi)\}=E\{(1-\Phi)\}E\{(1-\Psi)\}=\frac{1}{4}$$

$$E\{W_4\}=E\{(1-\Phi)\Psi\}=E\{(1-\Phi)\}E\{\Psi\}=\frac{1}{4},$$

since RVs Φ and Ψ (as well as their complements) are uniformly distributed in [0, 1] and are uncorrelated.

We now derive the entries of the variance–covariance matrix $\mathbf{V}_W$, using the results obtained in Section 5:

$$\mathrm{Var}\{W_1\}=E\{\Phi^2\Psi^2\}-[E\{W_1\}]^2=\frac{1}{9}-\frac{1}{16}=\frac{7}{144}$$

$$\mathrm{Var}\{W_2\}=E\left\{\Phi^2(1-\Psi)^2\right\}-[E\{W_2\}]^2=\frac{1}{9}-\frac{1}{16}=\frac{7}{144}$$

$$\mathrm{Var}\{W_3\}=E\left\{(1-\Phi)^2(1-\Psi)^2\right\}-[E\{W_3\}]^2=\frac{1}{9}-\frac{1}{16}=\frac{7}{144}$$

$$\mathrm{Var}\{W_4\}=E\left\{(1-\Phi)^2\Psi^2\right\}-[E\{W_4\}]^2=\frac{1}{9}-\frac{1}{16}=\frac{7}{144}$$

$$\begin{aligned}\mathrm{Cov}\{W_1,\ W_2\}&=E\{\Phi\Psi\Phi(1-\Psi)\}-E\{W_1\}E\{W_2\}\\&=E\{\Phi^2\}E\{\Psi(1-\Psi)\}-\frac{1}{4}\frac{1}{4}=\frac{1}{3}\frac{1}{6}-\frac{1}{16}=-\frac{1}{144}\end{aligned}$$

$$\begin{aligned}\mathrm{Cov}\{W_1,\ W_3\}&=E\{\Phi\Psi(1-\Phi)(1-\Psi)\}-E\{W_1\}E\{W_3\}\\&=E\{\Phi(1-\Phi)\}E\{\Psi(1-\Psi)\}-\frac{1}{4}\frac{1}{4}=\frac{1}{6}\frac{1}{6}-\frac{1}{16}=-\frac{5}{144}\end{aligned}$$

$$\begin{aligned}\mathrm{Cov}\{W_1,\ W_4\}&=E\{\Phi\Psi(1-\Phi)\Psi\}-E\{W_1\}E\{W_4\}\\&E\{\Phi(1-\Phi)\}E\{\Psi^2\}-\frac{1}{4}\frac{1}{4}=\frac{1}{6}\frac{1}{3}-\frac{1}{16}=-\frac{1}{144}\end{aligned}$$

$$\begin{aligned}\mathrm{Cov}\{W_2, W_3\} &= E\{\Phi(1-\Psi)(1-\Phi)(1-\Psi)\} - E\{W_2\}E\{W_3\} \\ &= E\{\Phi(1-\Phi)\}E\left\{(1-\Psi)^2\right\} - \frac{1}{4}\frac{1}{4} = \frac{1}{6}\frac{1}{3} - \frac{1}{16} = -\frac{1}{144}\end{aligned}$$

$$\begin{aligned}\mathrm{Cov}\{W_2, W_4\} &= E\{\Phi(1-\Psi)(1-\Phi)\Psi\} - E\{W_2\}E\{W_4\} \\ &= E\{\Phi(1-\Phi)\}E\{(1-\Psi)\Psi\} - \frac{1}{4}\frac{1}{4} = \frac{1}{6}\frac{1}{6} - \frac{1}{16} = -\frac{5}{144}\end{aligned}$$

$$\begin{aligned}\mathrm{Cov}\{W_3, W_4\} &= E\{(1-\Phi)(1-\Psi)(1-\Phi)\Psi\} - E\{W_3\}E\{W_4\} \\ &= E\left\{(1-\Phi)^2\right\}E\{(1-\Psi)\Psi\} - \frac{1}{4}\frac{1}{4} = \frac{1}{3}\frac{1}{6} - \frac{1}{16} = -\frac{1}{144}.\end{aligned}$$

Therefore, equation (41) becomes

$$\begin{aligned}E_s\{\hat{\sigma}_M(s)\} &= \frac{1}{144}\mathrm{tr}\left(\begin{bmatrix}\sigma_{11}^E & \sigma_{12}^E & \sigma_{13}^E & \sigma_{14}^E \\ \sigma_{21}^E & \sigma_{22}^E & \sigma_{23}^E & \sigma_{24}^E \\ \sigma_{31}^E & \sigma_{32}^E & \sigma_{33}^E & \sigma_{34}^E \\ \sigma_{41}^E & \sigma_{42}^E & \sigma_{43}^E & \sigma_{44}^E\end{bmatrix}\begin{bmatrix}7 & -1 & -5 & -1 \\ -1 & 7 & -1 & -5 \\ -5 & -1 & 7 & -1 \\ -1 & -5 & -1 & 7\end{bmatrix}\right) \\ &\quad + \frac{1}{16}\left([1\ 1\ 1\ 1]\begin{bmatrix}\sigma_{11}^E & \sigma_{12}^E & \sigma_{13}^E & \sigma_{14}^E \\ \sigma_{21}^E & \sigma_{22}^E & \sigma_{23}^E & \sigma_{24}^E \\ \sigma_{31}^E & \sigma_{32}^E & \sigma_{33}^E & \sigma_{34}^E \\ \sigma_{41}^E & \sigma_{42}^E & \sigma_{43}^E & \sigma_{44}^E\end{bmatrix}\begin{bmatrix}1 \\ 1 \\ 1 \\ 1\end{bmatrix}\right) \\ &= \frac{1}{144}\left[7\left(\sigma_{11}^E + \sigma_{22}^E + \sigma_{33}^E + \sigma_{44}^E\right) - 2\left(\sigma_{12}^E + \sigma_{23}^E + \sigma_{34}^E\right) - 10\left(\sigma_{13}^E + \sigma_{24}^E\right)\right] \\ &\quad + \frac{1}{16}\left[\left(\sigma_{11}^E + \sigma_{22}^E + \sigma_{33}^E + \sigma_{44}^E\right) + 2\left(\sigma_{12}^E + \sigma_{23}^E + \sigma_{34}^E\right) + 2\left(\sigma_{13}^E + \sigma_{24}^E\right)\right] \\ &= \frac{1}{9}\left[\sigma_{11}^E + \sigma_{22}^E + \sigma_{33}^E + \sigma_{44}^E + \sigma_{12}^E + \sigma_{23}^E + \sigma_{34}^E + \frac{1}{2}\left(\sigma_{13}^E + \sigma_{24}^E\right)\right],\end{aligned}$$

which is identical with equation (16) derived in Section 5.

8. Discussion and conclusions

As Kubik and Botman (1976), Dutton (1992), Shi (1998) and others have pointed out, prediction error variances associated with the interpolation procedures considered in this paper (linear interpolation along a line segment, TIN interpolation and bilinear interpolation in a rectangle) appear paradoxical at first sight. In general, prediction error variance is minimized not at the vertices of a polygon, where observations are available, but at locations within that polygon. In the special case of equal variances, prediction error variance is minimized at the

polygon's centroid. In this paper we have shown that this apparent paradox is a natural consequence of the fact that the above interpolation procedures can be regarded as local, interpolating, trend surface models. The fact that such trend-surface models reproduce the observations at the polygon vertices is just a consequence of considering the same number of data as the number of parameters characterizing the surface models (local fitting). A line, for example, will always pass through two points used to fit it, hence piecewise linear interpolation will always reproduce the observations at the endpoints of line segments, since one just fits a line locally for each segment to the measurements of its two endpoints. This data reproduction characteristic is independent of: (i) the particular predictors adopted, as long as the number of predictor variables equals the number of data considered for fitting, (ii) the definition of the measurement error component, and (iii) its covariance.

In this paper, it is shown that the general geostatistical framework of Universal Kriging (UK) underlies 1D linear interpolation, as well as TIN and bilinear interpolation. Prediction error variances are minimal at inner points when the prediction objective is that of the drift (mean) component, not of the attribute itself. In UK of the mean component, such variances increase as the predictor values at a point (in this case, the coordinates of that point) deviate from the overall mean of each predictor variable. Since the coordinates of any polygon vertex are maximally different from the mean of the polygon coordinates, prediction error variance is naturally maximized at polygon vertices. The other two variants of Universal Kriging considered in this work, namely UK of the attribute itself and factorial UK, yield identical weights and associated predictions with 1D linear interpolation, as well as TIN and bilinear interpolation in 2D. The corresponding prediction error variances, however, are fundamentally different, owing to the different definitions of measurement error (and consequently different prediction objectives) adopted in these latter UK variants. From a pure data standpoint, these alternative definitions of measurement error cannot be distinguished one from another, and one needs additional prior information to identify the relative importance of measurement error versus stochastic deviations from an expected surface not associated with such error.

From a modelling standpoint, however, we believe that UKM should be used with caution if at all in real world applications. More precisely, in UKM one places extreme confidence in a linear expected attribute surface, and assumes that any other attribute variability stems from measurement error. In many practical cases, however, the main source of interpolation error might be precisely the assumption of a linear surface to begin with: real world attributes need not vary linearly between the endpoints of a line or between the vertices of a triangle or a rectangle. Consequently, it is expected that interpolation errors would generally be smaller rather than larger as one gets closer to measurement locations (polygon vertices). This entails that UKM does not yield a realistic assessment of interpolation error, and should only be used in those cases where the analyst is extremely confident that the attribute does indeed vary linearly in space. On the other hand, the alternative formulations of UKZ and UKF are more flexible, since they allow for deviations from a dogmatic linear surface. Of course other forms of expected surfaces (e.g. polynomial) could be adopted, and such surfaces are indeed commonly used with imposed constraints of continuity in finite element modelling (Carey 1995), and as a replacement for linear interpolation in some applications of TIN (Akima 1978). Our

results could be extended to polynomial interpolation over triangles and quadrilaterals, but such an extension is beyond the scope of this work.

The results given in this paper offer rigorous ways of quantifying the prediction error variance associated with an interpolated value in the case of linear interpolation on lines and within triangles, and bilinear interpolation within rectangles. In practice, it is quite common for users to have some basis for estimating measurement error variances, typically from known characteristics of measuring devices, such as GPS receivers, thermometers or soil sampling instruments. Root-mean-square errors, for example, are commonly provided for digital elevation models (Hunter and Goodchild 1995). Alternatively, the user might be able to estimate spatially varying measurement error variances based on a theoretical or empirical relationship between the observations and some characteristic, say, gradient, of the underlying signal. Knowledge of covariances is less common, and we suspect that our results will most often be implemented in cases where covariances are assumed to be zero, either because observations are believed to be statistically independent, or because no information on covariances is available. Our results are sufficiently simple to be readily implementable in GIS, and we believe that GIS designers should use them to provide prediction error estimates routinely for interpolated values in such functions as TIN interpolation and raster resampling.

In addition to deriving propagated measurement error variances for individual points, we have provided estimates of average prediction error variance over a set (ensemble) of randomly chosen points for the particular case of Universal Kriging of the mean component. We are struck by the simplicity of many of the results: the progression in the case of unit variances and zero covariances from 2/3 for the line, to 1/2 for the triangle, to 4/9 for the rectangle; and the equivalence between interpolation weights and areas in the polygons shown in figures 2 and 4. These estimates could also be evaluated routinely in GIS, and in cases where measurement error is constant for all observations in a data set, the results could be added to the data set's metadata description. When measurement errors vary, estimates of average prediction error variance might be pre-computed and stored with line segments, triangles or rectangles.

Acknowledgements

We are grateful to the following researchers for stimulating our interest in this problem: C. Q. Zhu and G. X. Wang of the Zhengzhou Institute of Surveying and Mapping and the Chinese Academy of Sciences; W. Z. Shi and Tracy C. K. Cheung of the Hong Kong Polytechnic University; Q. Q. Li of Wuhan University; and Erfu Dai of the Chinese Academy of Sciences. We would like to thank Klaus Tempfli of the International Institute for Geo-Information Science and Earth Observation (ITC) in the Netherlands for providing us with some early references on this problem. We extend our appreciation to three anonymous reviewers, whose constructive comments led to significant improvements in the original manuscript. We also gratefully acknowledge the funding that was provided by the National Geospatial-Intelligence Agency (NGA) for the project *Strategic Enhancement of NGA's Geographic Information Science Infrastructure*.

References

Akima, H., 1978, A method of bivariate interpolation and smooth surface fitting for irregularly distributed data. *ACSM Transactions on Mathematical Software*, **4**, pp. 148–159.

BALL, W.R. and COXETER, H.M., 1960, *Mathematical Recreations and Essays* (New York: Macmillan).

BOTMAN, A.G. and KUBIK, K., 1979, On the theoretical accuracy of the moving average method for surface estimation. *The ITC Journal*, **1979-1**, pp. 68–84.

Carey G.F. (Ed.) 1995, *Finite Element Modeling of Environmental Problems: Surface and Subsurface Flow and Transport* (New York: Wiley).

CHILÈS, J.P. and DELFINER, P., 1999, *Geostatistics: Modeling Spatial Uncertainty* (New York: Wiley).

CRESSIE, N.A., 1993, *Statistics for Spatial Data* (revised edn) (New York: Wiley).

DUTTON, G., 1992, Handling positional uncertainty in spatial databases. In *Proceedings of the 5th International Symposium on Spatial Data Handling*, Charleston, SC, August 1992, Vol. 2, pp. 460–469.

EVANS, M., HASTINGS, N. and PEACOCK, B., 2000, *Statistical Distributions* (3rd edn) (New York: Wiley).

HEUVELINK, G.B.M., 1998, *Error Propagation in Environmental Modelling with GIS* (London: Taylor & Francis).

HUNTER, G.J. and GOODCHILD, M.F., 1995, Dealing with error in spatial databases: A simple case study. *Photogrammetric Engineering and Remote Sensing*, **61**, pp. 529–537.

JOHNSON, R.A. and WICHERN, D.W., 2002, *Applied Multivariate Statistical Analysis* (Upper Saddle River, NJ: Prentice Hall).

KUBIK, K. and BOTMAN, A.G., 1976, Interpolation accuracy for topographic and geological surfaces. *The ITC Journal*, **1976-2**, pp. 236–274.

RIPLEY, B.D., 1981, *Spatial Statistics* (New York: Wiley).

SEARLE, S.R., 1971, *Linear Models* (New York: Wiley).

SEARLE, S.R., 1982, *Matrix Algebra Useful for Statistics* (New York: Wiley).

SHI, W.Z., 1998, A generic statistical approach for modelling error of geometric features in GIS. *International Journal of Geographical Information Science*, **12**, pp. 131–143.

ZHU, C.Q., SHI, W.Z., LI, Q.Q., WANG, G.X., CHEUNG, T.C.K., DAI, E. and SHEA, G.Y.K., 2005, Estimation of average DEM accuracy under linear interpolation considering random error at the nodes of TIN model. *International Journal of Remote Sensing*, **26**, pp. 5509–5523.