ENVIRONMETRICS
Environmetrics 2010; **21**: 541–548
Published online 18 August 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/env.1015

Conditional versus unconditional mean-squared prediction errors for Gaussian processes with constant but unknown mean

Andreas Dominik Cullmann[1*,†] and Joachim Saborowski[2]

[1] *FVA Baden-Württemberg, Freiburg, Germany*
[2] *Department of Ecoinformatics, Biometrics and Forest Growth, University of Göttingen, Germany*

SUMMARY

For prediction in a Gaussian random field, we give an explicit formulation of the conditional mean-squared prediction error (cmspe). If the prediction method is ordinary kriging, we find that this error in most applications is likely to be very close to the ordinary kriging variance. This is additionally demonstrated based on a case study. Finally, we discuss the difference between these two errors compared to the error introduced by using estimated instead of true covariance parameters. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: conditional mean squared prediction error; ordinary kriging; Gaussian random field

1. INTRODUCTION

Whenever we deal with the general problem of predicting the value of an unobserved random variable Z_0 from the realisation z_n of an observed random vector $\mathbf{Z}_n = [Z_1, \ldots, Z_n]'$, whether the special case of the problem be a mixed linear model or spatial prediction (Harville and Jeske, 1992, p. 724), the mean squared prediction error is the means to evaluate the goodness of the prediction. Spatial prediction is widely used in environmental sciences, but whether the prediction itself is that of soil characteristics (Webster and Oliver, 1985), of tree growth increment (Biondi *et al.*, 1994) or of the percentage of damaged leaves (Köhl and Gertner, 1992): its error can either depend on the realisation of the observed random vector (the data, in practice) or not. For an arbitrary predictor $\hat{Z}_0$, the mean squared prediction error conditioned by the data is given by

$$E[\{\hat{Z}_0 - Z_0\}^2 \mid \mathbf{Z}_n] = \mathrm{var}[Z_0 \mid \mathbf{Z}_n] + (\hat{Z}_0 - E[Z_0 \mid \mathbf{Z}_n])^2 \tag{1}$$

This conditional mean-squared prediction error (cmspe) is unavailable in most applications because it requires the knowledge of the expectation $E[Z_0 \mid \mathbf{Z}_n]$ and variance $\mathrm{var}[Z_0 \mid \mathbf{Z}_n]$ of the conditional distribution of Z_0 given $\mathbf{Z}_n = z_n$. Therefore usually the unconditional mean squared prediction error (umspe) is computed. Its independence of the data has often been criticised (for example Armstrong,

*Correspondence to: Andreas Dominik Cullmann, FVA Baden-Württemberg, Freiburg, Germany.
†E-mail: acullma@gwdg.de

Received 14 May 2008
Accepted 7 June 2009

1994; Chilès and Delfiner, 1999; Goovaerts, 1997, chap. 5.8.3; Lloyd and Atkinson, 1999; Deutsch and Journel, 1998 and Yamamoto, 2000).

In the following, we will derive an explicit formulation of the cmspe of the best unbiased linear predictor (*blup*) for a Gaussian joint distribution of Z_0 and $\boldsymbol{Z}_n$ with constant but unknown mean (i.e. $E[Z_i] = \mu \,\forall\, Z_i; i = 0, \ldots, n$). Since it depends on the unknown mean, we examine the distribution of that difference between umspe and cmspe of the *blup* and—in case the *blup* is the ordinary kriging predictor—the absolute value of this difference.

2. GAUSSIAN PROCESSES WITH CONSTANT BUT UNKNOWN MEAN AND SPATIAL PREDICTION

2.1. Conditional mean-squared prediction error

As is well known, $E[Z_0 \mid \boldsymbol{Z}_n]$ is linear in $\boldsymbol{Z}_n$ for a Gaussian joint distribution of Z_0 and $\boldsymbol{Z}_n$ (Cressie, 1991, p. 109) and if additionally $E[Z_i] = \mu \,\forall\, Z_i; i = 0, \ldots, n$ then

$$E[Z_0 \mid \boldsymbol{Z}_n] = \mu + \boldsymbol{c}'\Sigma^{-1}(\boldsymbol{Z}_n - \mathbf{1}_n\mu) \tag{2}$$

(Stein, 1999, p. 3), where $\boldsymbol{c} = \mathrm{cov}[Z_0, \boldsymbol{Z}_n]$ and $\Sigma = \mathrm{cov}[\boldsymbol{Z}_n, \boldsymbol{Z}_n]$. Furthermore, if the joint distribution of Z_0 and $\boldsymbol{Z}_n$ is Gaussian then $\mathrm{var}[Z_0 \mid \boldsymbol{Z}_n]$ does not depend on $\boldsymbol{Z}_n$ (Johnson and Kotz, 1972, p. 41) and if additionally $\mathrm{var}[Z_i] = \sigma^2 \,\forall\, Z_i; i = 0, \ldots, n$ then

$$\mathrm{var}[Z_0 \mid \boldsymbol{Z}_n] = \sigma^2 - \boldsymbol{c}'\Sigma^{-1}\boldsymbol{c} \tag{3}$$

holds (Stein, 1999, p. 3). Since the mean μ is unknown, we estimate it using $\hat{\mu} = (\mathbf{1}_n'\Sigma^{-1}\mathbf{1}_n)^{-1}\mathbf{1}_n'\Sigma^{-1}\boldsymbol{Z}_n$. Now we have the *blup* given by

$$\hat{Z}_0 = \hat{\mu} + \boldsymbol{c}'\Sigma^{-1}(\boldsymbol{Z}_n - \mathbf{1}_n\hat{\mu}) \tag{4}$$

which has umspe

$$\begin{aligned} E(\hat{Z}_0 - Z_0)^2 &= \sigma^2 - \boldsymbol{c}'\Sigma^{-1}\boldsymbol{c} + (1 - \boldsymbol{c}'\Sigma^{-1}\mathbf{1}_n)^2(\mathbf{1}_n'\Sigma^{-1}\mathbf{1}_n)^{-1} \\ &= \sigma^2 - \boldsymbol{c}'\Sigma^{-1}\boldsymbol{c} + (1 - \boldsymbol{c}'\Sigma^{-1}\mathbf{1}_n)^2 E(\hat{\mu} - \mu)^2 \end{aligned} \tag{5}$$

By plugging Equations (2–4) into Equation (1), we derive the cmspe of the *blup* as

$$E[\{\hat{Z}_0 - Z_0\}^2 \mid \boldsymbol{Z}_n] = \sigma^2 - \boldsymbol{c}'\Sigma^{-1}\boldsymbol{c} + (1 - \boldsymbol{c}'\Sigma^{-1}\mathbf{1}_n)^2(\hat{\mu} - \mu)^2 \tag{6}$$

The difference between the umspe and the cmspe of the *blup* equates to

$$\text{Equation (5)} - \text{Equation (6)} = (1 - \boldsymbol{c}'\Sigma^{-1}\mathbf{1}_n)^2[E(\hat{\mu} - \mu)^2 - (\hat{\mu} - \mu)^2] \tag{7}$$

DOI: 10.1002/env

This difference becomes negative if and only if $(\hat{\mu} - \mu)^2 > E(\hat{\mu} - \mu)^2$, and since in the special case considered here $\hat{\mu}$ is a weighted sum of normally distributed $\boldsymbol{Z}_n$, then following Anderson (1958, p. 19)

$$\hat{\mu} \sim N\left(\mu, (\mathbf{1}_n' \Sigma^{-1} \mathbf{1}_n)^{-1}\right) \quad \text{and} \quad \left(\frac{\hat{\mu} - \mu}{\sqrt{E(\hat{\mu} - \mu)^2}}\right)^2 \sim \chi_1^2$$

Thus, $P[(\hat{\mu} - \mu)^2 > E(\hat{\mu} - \mu)^2] = P[(\hat{\mu} - \mu)^2 / E(\hat{\mu} - \mu)^2 > 1] = P[\chi_1^2 > 1] \approx 0.32$, that is in about 32% of all cases, the umspe of the *blup* under the special case considered here is less than the corresponding cmspe.

2.2. *Spatial prediction*

Let us suppose that $Z_0, \ldots, Z_n$ are marginal distributions of a spatial random process and $\boldsymbol{c}$ and Σ are matrices consisting of values of a spatial autocovariance which is a function, $\text{cov}[\boldsymbol{d}]$, of the spatial distance vectors between $Z_0, \ldots, Z_n$ only. Then Equation (4) is known as the ordinary kriging predictor refer to Cressie (1991, p. 123). From Equation (4)

$$\hat{Z}_0 = \hat{\mu} + \boldsymbol{c}'\Sigma^{-1}(\boldsymbol{Z}_n - \mathbf{1}_n\hat{\mu}) = \boldsymbol{c}'\Sigma^{-1}\boldsymbol{Z}_n + \hat{\mu}(1 - \boldsymbol{c}'\Sigma^{-1}\mathbf{1}_n) \tag{8}$$

we see that the first factor in Equation (7), $(1 - \boldsymbol{c}'\Sigma^{-1}\mathbf{1}_n)^2$, is the squared weight of the (estimated) mean for ordinary kriging.

Rivoirard (1984, p. 67), shows that the weight of the mean is close to zero if some of the spatial locations of $Z_1, \ldots, Z_n$ are close to the spatial location of Z_0, close with respect to the range of the spatial autocovariance function.

We can illustrate this by calculating $(1 - \boldsymbol{c}'\Sigma^{-1}\mathbf{1}_n)^2$ for an exponential spatial autocovariance function

$$\text{cov}_e(\boldsymbol{d}) = c_p \exp\left(-\frac{\boldsymbol{d}}{a_0}\right) \tag{9}$$

where $\boldsymbol{d}$, c_p and a_0 denote the vector of spatial distance, the sill and the range, respectively. Note that for $3\,a_0$ (which Deutsch and Journel (1998, p. 25), call effective range), $\text{cov}_e(\boldsymbol{d})$ reaches approximately $0.95\,c_p$.

Figure 1 shows the squared weight of the mean for $\hat{Z}_0$ when n observed random variables, $Z_1, \ldots, Z_n$, are arranged equally spaced on a circle around Z_0. With the range increasing from 0 to the radius of the circle (where at $a_0/\text{radius} = 1/3$ the observed random variables 'come into reach' of the spatial autocovariance function), the squared weight of the mean diminishes rapidly, whereas n, the number of observed random variables, does not have much influence when there is more than just a couple of observed random variables.

Let us now consider the more realistic assumption of systematic sampling. If the observed random variables are located on a square grid and Z_0 lies in the middle of the central square, then Figure 2(a) shows the squared weights of the mean using Equation (9) again. In Figure 2(a), the squared weights of the mean are shown over the range relative to the size of a quadratic grid with 100 grid points. Not surprisingly, the weights decrease with increasing relative range, which is due to more observed random variables coming into reach of the spatial autocovariance function. From

DOI: 10.1002/env

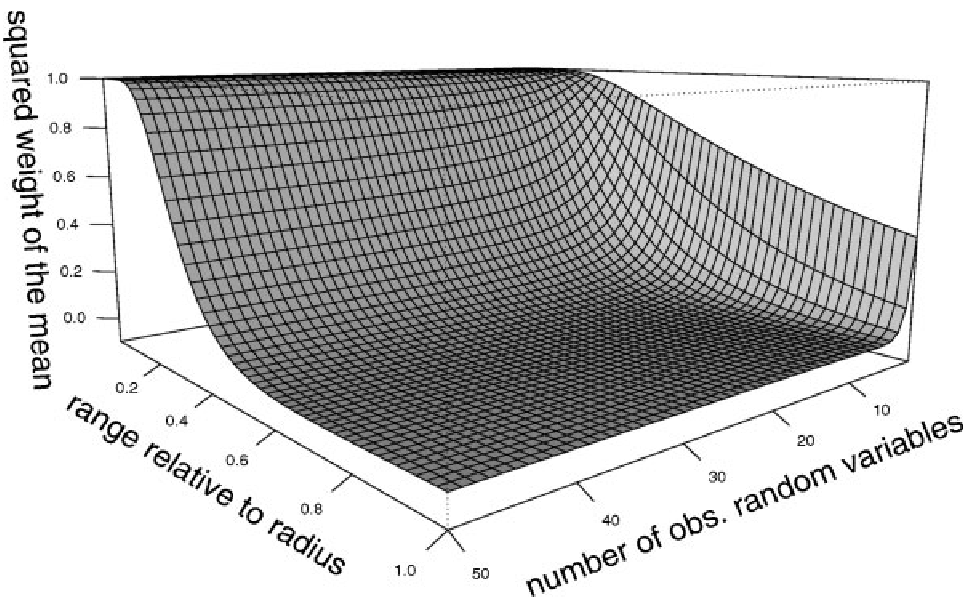

Figure 1. $(1 - \boldsymbol{c}'\Sigma^{-1}\mathbf{1}_n)^2$ for radial data using Equation (9). Graphics were produced using the code available from http://www.uni-forst.gwdg.de/˜acullma/cmspe/cmspe.r under R version 2.6.2 (2008-02-08), ISBN 3-900051-07-0, with the additional packages 'fields' version 4.1 and 'spatial' version 7.2-41

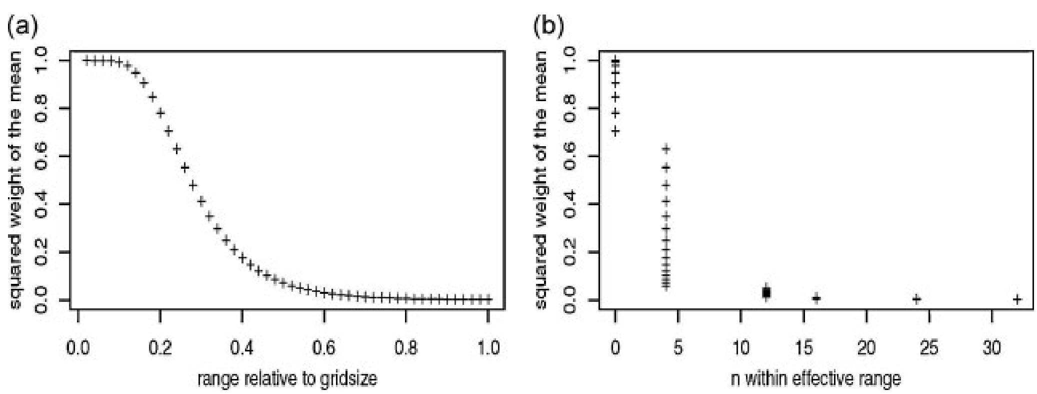

Figure 2. Squared weights of the mean using observed random variables on a quadratic 10 × 10 grid, over range relative to gridsize (a) and number of observed random variables within a spatial distance of 3 a_0 from the location of Z_0 (b). Graphics were produced using the code available from http://www.uni-forst.gwdg.de/˜acullma/cmspe/cmspe.r under R version 2.6.2 (2008-02-08), ISBN 3-900051-07-0, with the additional packages 'fields' version 4.1 and 'spatial' version 7.2-41

Figure 2(b) we see that the number of observed random variables, not more than the effective range apart from Z_0 in space, is a good indicator for the amount of $(1 - \boldsymbol{c}'\Sigma^{-1}\mathbf{1}_n)^2$. Both figures show that the squared weight of the mean for ordinary kriging ought to be very close to zero in most applications.

DOI: 10.1002/env

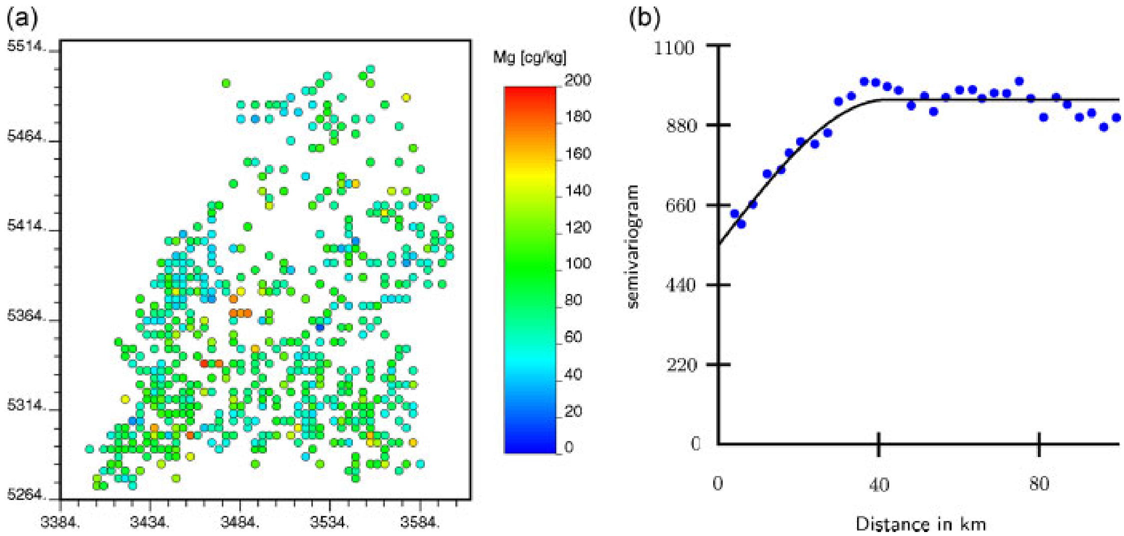

Figure 3. Magnesium content of spruce needles, Baden-Württemberg 1994, Gauß–Krüger coordinates in kilometers (a); Empirical (dots) and fitted semivariogram (line) (b). Graphics were produced using the code available from http://www.uni-forst.gwdg.de/~acullma/cmspe/cmspe.r under R version 2.6.2 (2008-02-08), ISBN 3-900051-07-0, with the additional packages 'fields' version 4.1 and 'spatial' version 7.2-41. This figure is available in colour online at www.interscience.wiley.com/journal/env

3. A CASE STUDY

Magnesium deficiency causes disadvantageous nitrogen–magnesium ratios and can lead to increment loss in conifer stands (refer to Evers and Hüttl, 1992; Liu and Hüttl, 1991), Danneberg (2001) gives magnesium deficiency thresholds of about 70–80 cg/kg for spruce needles.

Figure 3(a) shows the magnesium content of spruce needles collected for the Immissionsökologische Waldzustandserfassung in the state of Baden-Württemberg, South West Germany, in 1994, Figure 3(b) the empirical and fitted semivariograms. The latter is the isotropic spherical autocovariance function

$$\mathrm{cov}_s|\boldsymbol{d}| = \begin{cases} 400 \cdot \left\{1 - \left[\frac{3}{2} \cdot \frac{|\boldsymbol{d}|}{42} - \frac{1}{2} \cdot \frac{|\boldsymbol{d}|^3}{42^3}\right]\right\} & \text{for } 0 < |\boldsymbol{d}| \leq 42 \\ 550 + 400 & \text{for } |\boldsymbol{d}| = 0 \\ 0 & \text{for } |\boldsymbol{d}| \geq 42 \end{cases} \tag{10}$$

Let us assume that the data are a realisation of $n = 578$ jointly distributed Gaussian random variables $\boldsymbol{Z}_n = [Z_1, \ldots, Z_{587}]'$ with constant first moments and second moments given by Equation (10), then the weight of the mean in Equation (8), $1 - \boldsymbol{c}'\Sigma^{-1}\mathbf{1}_n$, for predicting the magnesium content in spruce needles over the area of the state of Baden-Württemberg (assuming that for each prediction location the unobserved Z_0 joins the distribution of $\boldsymbol{Z}_n$) is shown in Figure 4(a).

We clearly see that this weight is close to zero whenever there are some data within a reasonable range of the prediction location. With this weight squared, the difference between umspe and cmspe, Equation (7), is likely to be negligible for most parts of the area of interest.

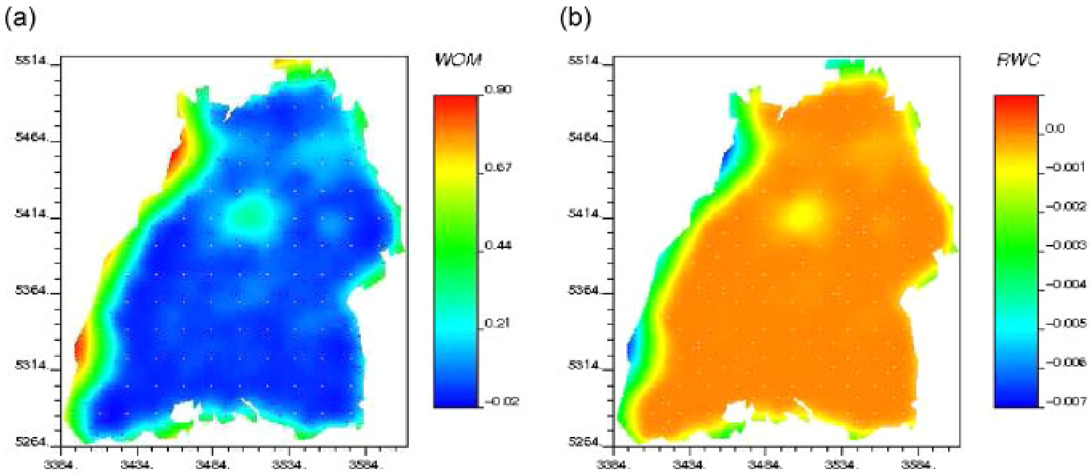

Figure 4. Weight of the mean (WOM) for ordinary kriging (a); relative width change (RWC) of confidence intervals (b) for $(\hat{\mu} - \mu)^2 = 2 \cdot E(\hat{\mu} - \mu)^2$. Gauß–Krüger coordinates in kilometres. This figure is available in colour online at www.interscience.wiley.com/journal/env

If we wanted to estimate the relative change in width

$$\text{RCW} = \frac{\text{width}\left(p \pm \sqrt{\text{umspe}} \cdot u_{1-(\alpha/2)}\right) - \text{width}\left(p \pm \sqrt{\text{cmspe}} \cdot u_{1-(\alpha/2)}\right)}{\text{width}\left(p \pm \sqrt{\text{cmspe}} \cdot u_{1-(\alpha/2)}\right)}$$

$$= \frac{\sqrt{\text{umspe}} - \sqrt{\text{cmspe}}}{\sqrt{\text{cmspe}}}$$

of Gaussian confidence intervals of type $[p \pm \sqrt{\text{mspe}} \cdot u_{1-(\alpha/2)}]$ that origins from substituting cmspe by umspe, we need not only to know the weight of the mean, but $E(\hat{\mu} - \mu)^2 - (\hat{\mu} - \mu)^2$, the second factor in Equation (7), too. Since the mean is unknown, we shall make the following assumptions which are ordered by the goodness of the estimation of the constant mean:

(a) $E(\hat{\mu} - \mu)^2 - (\hat{\mu} - \mu)^2 = E(\hat{\mu} - \mu)^2$
the estimation of the mean is perfect, there is no error ($\hat{\mu} - \mu = 0$), thus $(\hat{\mu} - \mu)^2/E(\hat{\mu} - \mu)^2 = 0$ and the probability of observing a realisation with $(\hat{\mu} - \mu)^2/E(\hat{\mu} - \mu)^2 \geq 0$ is $P[\chi_1^2 > 0] = 1$.

(b) $E(\hat{\mu} - \mu)^2 - (\hat{\mu} - \mu)^2 = 0$
the squared error of the estimation of the mean is as big as expected to be ($(\hat{\mu} - \mu)^2 = E(\hat{\mu} - \mu)^2$), this implies $(\hat{\mu} - \mu)^2/E(\hat{\mu} - \mu)^2 = 1$, the probability of observing a realisation with $(\hat{\mu} - \mu)^2/E(\hat{\mu} - \mu)^2 \geq 1$ is $P[\chi_1^2 > 1] \approx 0.3173105$.

(c) $E(\hat{\mu} - \mu)^2 - (\hat{\mu} - \mu)^2 = -E(\hat{\mu} - \mu)^2$
the squared error of the estimation of the mean is twice as big as expected ($(\hat{\mu} - \mu)^2 = 2 \cdot E(\hat{\mu} - \mu)^2$). $(\hat{\mu} - \mu)^2/E(\hat{\mu} - \mu)^2 = 2$, the probability of observing a realisation with $(\hat{\mu} - \mu)^2/E(\hat{\mu} - \mu)^2 \geq 2$ is $P[\chi_1^2 > 2] \approx 0.1572992$.

The relative change in width under assumption (c) is depicted in Figure 4(b): it is distinguishable from zero only close to the western border of Baden-Württemberg, where the weight of the mean $1 - \boldsymbol{c}'\Sigma^{-1}\mathbf{1}_n$ is close to 1 (Figure 4(a)). Table 1 sums up results for the three assumptions. We see

DOI: 10.1002/env

Table 1. Extrema and median for some statitics, RCW for assumptions (a) through (c)

Statistic	Assumption	Minimum	Maximum	Median
$E(\hat{Z}_0 - Z_0)^2$		0	950.0342	562.5342
$E(\hat{\mu} - \mu)^2$		16.1207	16.1207	16.1207
$1 - \boldsymbol{c}'\Sigma^{-1}\mathbf{1}_n$		−0.0105	0.8936	0.0638
RCW	(a)	0	0.0067	0.000059
RCW	(b)	0	0	0
RCW	(c)	−0.0068	0	−0.000059

that the width of the confidence intervals does not change remarkably even for relatively extreme realisations (having $(\hat{\mu} - \mu)^2/E(\hat{\mu} - \mu)^2 = 0$ and $(\hat{\mu} - \mu)^2/E(\hat{\mu} - \mu)^2 = 2$). This is probably partly due to the fact that $E(\hat{\mu} - \mu)^2$ is small compared to the median of $E(\hat{Z}_0 - Z_0)^2$, so that for most locations the value of $E(\hat{\mu} - \mu)^2 - (\hat{\mu} - \mu)^2$ in $E(\hat{Z}_0 - Z_0)^2 - E[\{\hat{Z}_0 - Z_0\}^2 \mid \boldsymbol{Z}_n]$ will be small compared to $E[\{\hat{Z}_0 - Z_0\}^2 \mid \boldsymbol{Z}_n]$, especially when $-E(\hat{\mu} - \mu)^2 \le E(\hat{\mu} - \mu)^2 - (\hat{\mu} - \mu)^2 \le E(\hat{\mu} - \mu)^2]$—which is the case with $P[0 \le \chi_1^2 \le 2] \approx 0.843$.

The same will be true for the RCW which is the relative difference of the square roots of $E(\hat{Z}_0 - Z_0)^2$ and $E[\{p - Z_0\}^2 \mid \boldsymbol{Z}_n]$.

4. DISCUSSION

As we have seen in Section 2, the unconditional mspe of the *blup*—under the assumption of a Gaussian joint distribution of Z_0 and $\boldsymbol{Z}_n$ with constant mean—is likely (with probability $1 - P[\chi_1^2 > 1] \approx 0.68$) to be higher than the conditional mean squared prediction error, resulting in approximate confidence intervals of type $\hat{Z}_0 \pm 1.96E(\hat{Z}_0 - Z_0)^2$ that are wider, i.e. more conservative, than confidence intervals of type $\hat{Z}_0 \pm 1.96E[\{\hat{Z}_0 - Z_0\}^2 \mid \boldsymbol{Z}_n]$. For ordinary kriging, from Section 2.2, the difference between conditional and umspe, $E(\hat{Z}_0 - Z_0)^2 - E[\{\hat{Z}_0 - Z_0\}^2 \mid \boldsymbol{Z}_n]$, is likely to be close to zero for most applications, because $(\hat{\mu} - \mu)^2$ will also be small compared to its expectation for moderate and larger sample sizes (see Equation (7)). Even if this difference, due to an increasing weight of the mean, increases, so does $E[\{\hat{Z}_0 - Z_0\}^2 \mid \boldsymbol{Z}_n]$ (according to Equation (6)), and the relative error $E(\hat{Z}_0 - Z_0)^2 - E[\{\hat{Z}_0 - Z_0\}^2 \mid \boldsymbol{Z}_n]/E[\{\hat{Z}_0 - Z_0\}^2 \mid \boldsymbol{Z}_n]$ ought to stay reasonably small. This states our opinion that in the special case considered here $E(\hat{Z}_0 - Z_0)^2$ is an appropriate approximation of $E[\{\hat{Z}_0 - Z_0\}^2 \mid \boldsymbol{Z}_n]$.

Furthermore, in practice the parameters $\boldsymbol{\theta}$ of the spatial autocovariance function giving the second moments of the joint distribution of Z_0 and $\boldsymbol{Z}_n$ are usually estimated from the realisation of $\boldsymbol{Z}_n$ (refer to Stein, 1999, p. vii); this estimation is denoted by $\hat{\boldsymbol{\theta}}$ in the following. It results in using the 'empirical' or 'estimated' best linear unbiased estimator, $p_2(\boldsymbol{Z}_n, \hat{\boldsymbol{\theta}})$, with actual umspe $m_2(\boldsymbol{\theta}) = E(Z_0 - p_2(\boldsymbol{Z}_n, \hat{\boldsymbol{\theta}}))^2$ instead of the *blup* $p_1(\boldsymbol{Z}_n, \boldsymbol{\theta})$ with umspe $m_1(\boldsymbol{\theta}) = E(Z_0 - p_1(\boldsymbol{Z}_n, \boldsymbol{\theta}))^2$ (refer to Stein, 1999, p. 200).

$m_2(\boldsymbol{\theta})$ is, '*aside from relatively simple special cases*' (Harville and Jeske, 1992, p. 725), unavailable, hence in practice it is estimated by the presumed error of $p_2(\boldsymbol{Z}_n, \hat{\boldsymbol{\theta}}), m_1(\hat{\boldsymbol{\theta}}) = E(Z_0 - p_1(\boldsymbol{Z}_n, \hat{\boldsymbol{\theta}}))^2$, which is obtained by substituting $\boldsymbol{\theta}$ with $\hat{\boldsymbol{\theta}}$ in $m_1(\boldsymbol{\theta})$.

Zimmerman (2006) shows that regularly spaced sampling designs for $\boldsymbol{Z}_n$ enhance prediction but clustered designs result in improved estimation of $\boldsymbol{\theta}$, so there seems to be a trade-off between estimating $m_2(\boldsymbol{\theta})$ through $m_1(\hat{\boldsymbol{\theta}})$ and keeping $m_2(\boldsymbol{\theta})$ small at the same time.

DOI: 10.1002/env

Zimmerman and Cressie (1992), example 2, show that for a Gaussian joint distribution of Z_0 and $\mathbf{Z}_n$ with a spherical spatial autocovariance the bias

$$\frac{E\left[m_1(\hat{\boldsymbol{\theta}})\right] - m_2(\boldsymbol{\theta})}{m_2(\boldsymbol{\theta})}$$

can mount up to over 0.33 if the spatial correlation is weak, i.e. if the weight of the mean in kriging is high. Similar problems can be expected to occur with the cmspe of the estimated *blup* for Z_0.

ACKNOWLEDGEMENTS

This research was partially supported by the DFG under grants SA 415/3-1 and SA 415/3-2. The data for the case study were kindly provided by Dr Klaus-Hermann von Wilpert, Forstliche Versuchs- und Forschungsanstalt Baden-Württemberg. The authors are also grateful for the helpful comments of an anonymous reviewer.

REFERENCES

Anderson TW. 1958. *An Introduction to Multivariate Statistical Analysis*. Wiley: New York.

Armstrong M. 1994. Is research in mining geostats as dead as a dodo? In *Geostatistics for the Next Century*, Dimitrakopoulos R (ed.). Kluwer: Dordrecht; 303–312.

Biondi F, Myers DE, Avery CC. 1994. Geostatistically modeling stem size and increment in an old-growth forest. *Canadian Journal of Forest Research* **24**: 1354–1368.

Chilès J-P, Delfiner P. 1999. *Geostatistics—Modelling Spatial Uncertainty*. Wiley: New York.

Cressie NAC. 1991. *Statistics for Spatial Data*. Wiley: New York.

Danneberg O, Jasser Ch., Katzensteiner K, Luckel W, Mutsch F, Reh M, Schuster K, Starlinger F. 2001. Wald(boden)sanierung. Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft. Fachbeirat für Bodenfruchtbarkeit und Bodenschutz, Wien.

Deutsch CV, Journel AG. 1998. *GSLIB: Geostatistcal Software Library and User'S Guide*. Oxford University Press: New York.

Evers FH, Hüttl RF. 1992. Magnesium-, Calcium- und Kaliummangel bei Waldbäumen – Ursachen, Symptome, Behebung. Merkblätter der Forstlichen Versuchs- und Forschungsanstalt Baden-Württemberg **42**, Forstliche Versuchs- und Forschungsanstalt Baden-Württemberg, Freiburg.

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

Harville DA, Jeske DR. 1992. Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association* **87**(419): 724–731.

Johnson NL, Kotz S. 1972. *Distributions in Statistics: Continuous Multivariate Distributions*. Wiley: New York.

Journel AG. 1986. Geostatistics: models and tools for the earth sciences. *Mathematical Geology* **18**: 119–140.

Köhl M, Gertner G. 1992. Geostatistische Auswertungsmöglichkeiten für Waldschadeninventuren: Methodische Überlegungen zur Beschreibung räumlicher Verteilungen. *Forstwissenschaftliches Centralblatt* **111**: 320–331.

Liu JC, Hüttl RF. 1991. Relations between damage symptoms and nutritional status of Norway spruce stands (Picea abies Karst.) in southwestern Germany. *Nutrient Cycling in Agroecosystems* **27**(1): 9–22.

Lloyd CD, Atkinson PM. 1999. Designing optimal sampling configurations with ordinary and indicator kriging. In *Proceedings of the 4th International Conference on GeoComputation*, Mary Washington College, Fredericksberg, Virginia, USA, 25–28 July. Greenwich GeoComputation CD-ROM. URL: http://www.geocomputation.org/1999/065/gc_065.htm

Rivoirard J. 1984. Le Comportement des Poids de Krigeage. Thèse de Docteur-Ingénieur en Sciences et Techniques Minières, Option Géostatistique, Ecole des Mines de Paris, Fontainebleu.

Stein ML. 1999. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer: New York.

Webster R, Oliver MA. 1985. Utilisation exploratoire de la geostatistique pour la cartographie du sol dans la forêt de Wyre (G.B.). *Sciences de la Terre, Informatique Géologique* **24**: 171–173.

Yamamoto JK. 2000. An alternative measure of the reliability of ordinary kriging estimates. *Mathematical Geology* **32**: 489–509.

Zimmerman DL. 2006. Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. *Environmetrics* **17**: 635–652.

Zimmerman DL, Cressie NAC. 1992. Mean squared prediction error in the spatial linear model with estimated covariance parameters. *Annals of the Institute of Statistical Mathematics* **44**(1): 27–43.