Is there anything wrong with the ordinary kriging variance?

Article · January 2001

Source: OAI

CITATIONS
0

READS
984

2 authors:

Gerard Heuvelink
Wageningen University & Research
445 PUBLICATIONS **25,342** CITATIONS

SEE PROFILE

Edzer Pebesma
University of Münster
260 PUBLICATIONS **24,402** CITATIONS

SEE PROFILE

IS THE ORDINARY KRIGING VARIANCE A PROPER MEASURE OF INTERPOLATION ERROR?

Gerard B.M. Heuvelink
Institute for Biodiversity and Ecosystem Dynamics, Universiteit van Amsterdam,
Nieuwe Achtergracht 166, 1018 WV Amsterdam, the Netherlands
Email: g.b.m.heuvelink@science.uva.nl

Edzer J. Pebesma
Faculty of Geographical Sciences, University of Utrecht
P.O. Box 80.115, 3508 TC Utrecht, the Netherlands
Email: e.pebesma@geog.uu.nl

Abstract

Ordinary kriging (OK) is probably the most widely used kriging technique. This is because it is simple and robust. Recently, however, OK has received much criticism. The criticism concentrates on the OK variance. Many geostatisticians consider the OK variance an unreliable representation of the spatial interpolation error. Some even go as far as to consider it completely useless. In this paper we analyse under which circumstances the criticism on the OK variance is right and under which circumstances it is wrong. If we assume that the regionalised variable under study is a realisation of a stationary Gaussian random function, then the OK variance is entirely valid. Even when a sample taken from a realisation of a stationary Gaussian random function indicates that in some part of the area the variation is greater (or smaller) than elsewhere (and this is likely to happen by chance), then the OK variance still is a correct assessment of the local uncertainty. We verify this analytical result by means of a simulation study. From realisations generated using sequential Gaussian simulation, we analyse whether the observed interpolation error is larger (smaller) in areas where the local variation is greater (smaller). This turns out not to be the case. The numerical results thus confirm that in the Gaussian case the local prediction error is not affected by the degree of local variation. We conclude from this that one should be careful with applying alternative kriging methods that force the prediction error variance to be larger in areas with greater local variability. These findings do not extend to the non-Gaussian case, and so we also advocate that methodologies are developed that can judge the fairness of making the Gaussian assumption.

Keywords: geostatistics, kriging, interpolation error, local uncertainty, stochastic simulation

1. INTRODUCTION

It is now more than twenty-five years ago that the attention of the GIS community was drawn to the problem of spatial error propagation (Burrough 1986). Much progress has been made since, and recent publications addressing realistic case studies demonstrate the maturing of spatial uncertainty propagation analysis in GIS (e.g. Kros et al. 1999, Kyriakidis and Dungan 2001, Bachmann and Allgöwer 2002, Canters et al. 2002).

The errors or uncertainties that propagate through GIS analyses may stem from many sources. One important error source and the one that we will focus on in this paper is interpolation error. Interpolation error arises when the value of a spatial attribute at some location is unknown but must be predicted from neighbouring observations. This is a situation that frequently occurs in practice, because in many practical situations it is impossible to sample exhaustively. Instead, maps must be constructed from a limited number of observations taken at sampling locations in the field. Regardless of which interpolation method is used, there will always be interpolation error. An attractive property of geostatistical interpolation (kriging) is that it quantifies (and minimises) the interpolation error. The simplest kriging technique is simple kriging (SK), which, in mathematical terms, works as follows:

We consider a spatial attribute $z = \{z(x) \mid x \in D\}$ that is observed at n locations x_i, $i = 1,\ldots,n$, and for which a prediction is required at the unobserved location x_0. We assume that z is a realisation of a random function Z, which has a constant mean (the expectation $E[Z(x)] = m$ does not depend on x) and whose autocovariance function does not depend on the locations but only on the distance between them ($\mathrm{Cov}(Z(x), Z(x+h)) = R(h)$). These two assumptions entail that Z is (second-order) stationary. Now we can predict $Z(x_0)$ as follows:

$$Z^*(x_0) = m + \sum_{i=1}^{n} \lambda_i \left(Z(x_i) - m\right) \tag{1}$$

where the SK weights λ_i are chosen such that the SK variance is minimised. The SK variance is defined as:

$$\sigma^2_{SK}(x_0) = \mathrm{Var}\left(Z(x_0) - Z^*(x_0)\right) \tag{2}$$

Minimising $\sigma^2_{SK}(x_0)$ leads to a unique solution for the kriging weights λ_i:

$$\lambda = C^{-1}c \tag{3}$$

where $\lambda = [\lambda_1\, \lambda_2 \cdots \lambda_n]^T$ is the vector of SK weights, C is the n×n-covariance matrix of the $Z(x_i)$ and where c is the vector of covariances between the $Z(x_i)$ and $Z(x_0)$. By substituting the optimal kriging weights in Eq. (2) and maintaining matrix notation we get:

$$\sigma^2_{SK}(x_0) = R(0) - c^T C^{-1} c \tag{4}$$

The interesting property of the SK variance is that it does not depend on the data values $z(x_i)$ themselves. This property, which is the key focus of this paper, provides the basis for methodologies that optimise sampling density given a maximum tolerable interpolation error (McBratney et al. 1981, Van Groenigen et al. 1999).

Simple kriging is not that often used in practice because the assumption that the mean m of Z is known is not very realistic. The extension of SK to the situation of unknown m is known as ordinary kriging (OK). OK is the most frequently used kriging technique in practice. It has much resemblance with SK, although the equations are slightly more involved. This is why in this paper we have chosen to restrict the mathematical-statistical discussion to SK, although the arguments made hold equally well for OK.

The derivations above are all very basic and presented in any standard geostatistical text book (e.g. Goovaerts 1997). It was briefly repeated here because we will refer to it later on. One interesting observation can already be made: in the derivation of the kriging prediction and variance nowhere was the assumption made that Z should be normally (Gaussian) distributed.

2. WHY DID THE OK VARIANCE FALL IN DISGRACE?

Goovaerts (1997, p. 179) gives three key properties of the OK variance:

1) it is dependent on the covariance model,
2) it is dependent on the data configuration,
3) it is independent of data values.

Goovaerts (1997) qualifies the first two properties as excellent features. But he considers the third property as a bad feature. Intuitively, "the potential for prediction error is expected to be greater at a location surrounded by data that are very different from one another than at a location surrounded by similarly valued data" (Goovaerts, 1997, p. 180). The OK variance fails to recognise this. Some authors consider this to be such an unattractive property that they consider the OK variance as effectively useless. In their opinion, the OK variance is merely a ranking measure of data configuration (Journel et al. 2000). Alternative measures are suggested to overcome the fact that the OK variance presumably is not a full measure of uncertainty (Yamamoto 2000). In less than a decade the OK variance has transformed from a valuable parameter to the outcast of the geostatistical toolkit.

The question that we ask ourselves in this paper is whether it is right to put aside the OK variance that easily. Is it indeed a useless parameter? And if so, how can this be explained, given that it is the result of a derivation that is mathematically and statistically sound? Where do things get wrong, if they do? We investigate this problem first analytically and next verify the analytical results using a numerical experiment.

3. STATISTICAL PREDICTION THEORY

The objective of kriging is to predict the value of the random variable $X=Z(x_0)$ given the outcome of the random vector $Y=\left[Z(x_1)\ Z(x_2)\ Z(x_3)\ldots Z(x_n)\right]$. As is well known from statistical prediction theory, ideally for this we should work with the conditional probability of X given Y=y (Bayes' rule):

$$p_{X|Y}(x \mid y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} \tag{5}$$

where $p_{X,Y}(x,y)$ is the joint probability density of X and Y and where $p_Y(y)$ is the marginal probability density of Y. We can compute the mean and variance of the conditional probability density of X given Y=y. An important result from mathematical statistics is that the conditional mean is the minimum variance estimate of X given Y=y. In other words, it provides the best estimate (in minimum variance sense) for X from the observations y. This explains why it is so attractive to use the conditional mean as a predictor. However, the difficulty in practice is that is not always easy to compute the conditional density, and from it, its mean and variance. An exception is when X and Y are jointly normally distributed, in

which case the conditional distribution is also normally distributed with conditional mean and variance given by:

$$E[X \mid Y = y] = m_X + c^T C^{-1}(y - m_Y) \tag{6}$$

$$\mathrm{Var}(X \mid Y = y) = R(0) - c^T C^{-1} c \tag{7}$$

where m_X is the mean of X, m_Y is the mean of Y, R(0) is the variance of X, C is the covariance matrix of Y, and c is the vector of covariances between X and Y.

Some important observations can now be made. First, we note that Eqns. (6) and (7) are identical to the simple kriging equations. Second, we see that the conditional variance is independent of y. Thus, under the assumptions made, the simple kriging predictor is the best predictor of $z(x_0)$ from the observations $z(x_1),\ldots,z(x_n)$, and moreover the accuracy of the prediction is not affected by the degree of local variation in the observations. In other words, there is nothing wrong with the much disputed third property of kriging. We made optimal use of our observations to predict $z(x_0)$, and the prediction error we make is independent of the observations.

Of course the above is only valid under the assumptions made. The crucial assumption we made is that we assumed that X and Y are jointly Gaussian. If this is not the case then the conditional variance computed from the conditional density in Eq. (5) may well depend on y. The conditional variance being independent of y is a special feature of Gaussian densities which is not true in general.

To establish Eqns. (6) and (7) we assumed that X and Y were jointly Gaussian. But simple kriging yielded the same results without requiring normality. How can this be explained? This is because SK restricts itself to predictions that are linear combinations of the observations. In the non-Gaussian case this is a real restriction because the conditional mean is non-linear in the observations. In that case SK does not yield the conditional mean and variance and hence it is sub-optimal. It does not exploit all information contained in the observations. If it would, then the prediction error variance would equal the conditional variance and thus would almost surely depend on the observations. In other words, the fact that the SK and OK variance do not depend on the data values is indeed an unattractive property in the non-Gaussian case. Apparently, the validity of the Gaussian assumption is a critical element in the analysis. We will come back to this in the Discussion.

In the next section we first verify the obtained theoretical results with a numerical experiment.

4. SIMULATION EXPERIMENT

We used the following procedure:

1. We generated 50 realisations of a Gaussian random function on a 30×30 grid, using unconditional sequential Gaussian simulation. The variogram used is of the spherical type, with 20 per cent nugget variance and range equal to 15 grid cells. Figure 1 gives three example realisations.
2. For each realisation, we took eight observations in the upper right corner and eight observations in the lower left corner, by overlaying the realisation with a mask. The eight observations are taken along the sides of a 7×7 square window. The results for the three example realisations are given in Figure 2.

3. For each realisation we then computed the sample variance for each corner, which we used as a measure for the degree of local variation.
4. Next for each realisation and for each corner we kriged to the centre of the window, using ordinary kriging. We only used the eight neighbouring observations. Note that for symmetry reasons, the OK variance is the same for both corners (and for all realisations, for that matter).
5. We then compared the kriged value in the centre of the window with the 'true' (simulated) value and computed the absolute difference between the two.
6. We finally compared for each realisation the two corners: we determined which corner had the largest local variation, and we determined which corner had the largest prediction error. We looked how often these were the same corners.

Table 1 gives the results of this experiment. For each realisation it gives the ratio of the local variation between the upper right and lower left corner, the ratio of the observed absolute prediction errors between the upper right and lower left corner, and a result on whether the two ratios mutually agree: is the observed prediction error indeed larger at the corner with the largest local variation?

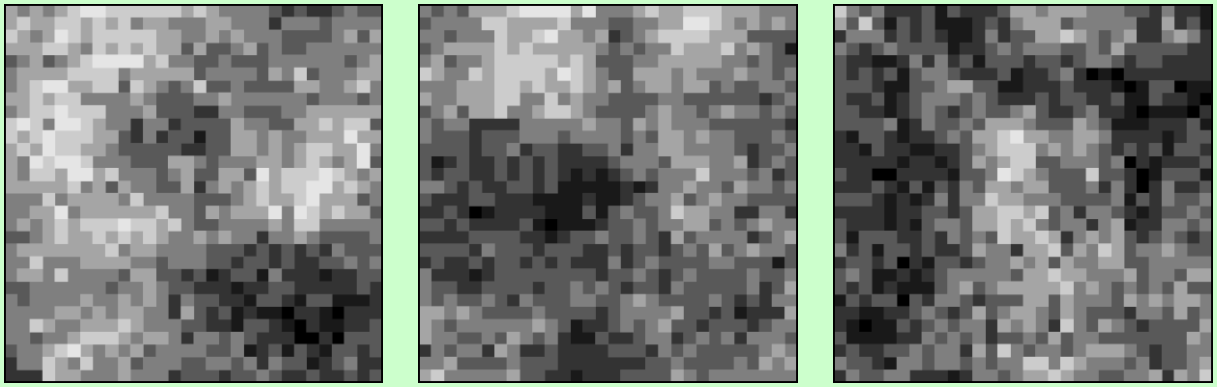

Figure 1. Three example realisations (from the total of 50) of a stationary Gaussian random function simulated on a 30×30 grid, using sequential Gaussian simulation and a spherical variogram with 20 per cent nugget variance and range equal to 15 cell widths. Left: simulation number 6. Middle: simulation number 11. Right: simulation number 16. Grey tones represent attribute value (light=low, dark=high).

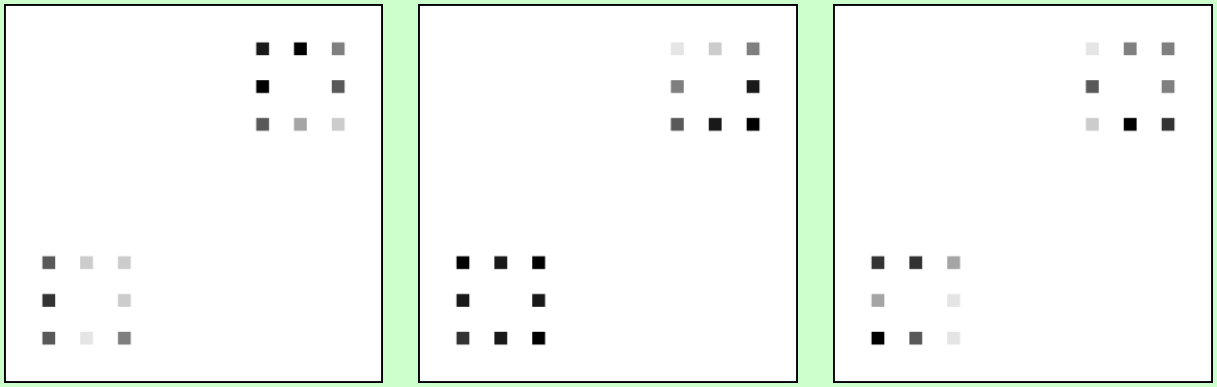

Figure 2. Eight observations taken from the upper right and eight from the lower left corner of the three simulated maps given in Figure 1. Ratios of sample variances (upper right corner over lower left corner) are: 1.34 (Left), 13.39 (Middle) and 1.03 (Right).

sim. nr	ratio local variation	ratio pred. errors	agree-ment?
1	0.54	2.50	0
2	1.38	0.10	0
3	0.80	2.46	0
4	2.18	0.25	0
5	1.34	1.95	1
6	1.34	7.61	1
7	2.64	0.14	0
8	1.07	5.87	1
9	1.10	0.82	0
10	0.51	0.65	1
11	13.99	4.63	1
12	0.45	168.00	0
13	7.70	2.61	1
14	0.71	1.89	0
15	6.18	1.27	1
16	1.03	24.53	1
17	0.37	0.39	1
18	0.51	2.81	0
19	12.00	0.79	0
20	0.33	1.36	0
21	4.68	0.10	0
22	1.56	0.39	0
23	0.43	1.30	0
24	2.52	3.11	1
25	1.25	0.36	0
26	23.30	3.74	1
27	2.51	0.88	0
28	0.47	0.96	1
29	2.83	1.12	1
30	1.70	1.09	1
31	0.47	9.96	0
32	1.83	1.35	1
33	0.48	12.23	0
34	0.20	0.80	1
35	1.17	1.79	1
36	1.75	1.65	1
37	0.12	0.84	1
38	0.38	1.01	0
39	0.92	3.56	0
40	1.18	0.10	0
41	1.04	0.17	0
42	17.46	1.07	1
43	0.66	4.20	0
44	0.73	0.45	1
45	0.12	1.04	0
46	1.36	8.42	1
47	8.60	0.47	0
48	0.42	0.02	1
49	2.65	1.57	1
50	4.92	1.72	1

Table 1. Results for the 50 Gaussian realisations. For each realisation the ratios of the local variation and the observed absolute prediction errors between the upper right and lower left corner are given; 'agreement?' refers to whether the observed prediction error is indeed larger for the corner with the largest local variation (agreement=1) or not (agreement=0).

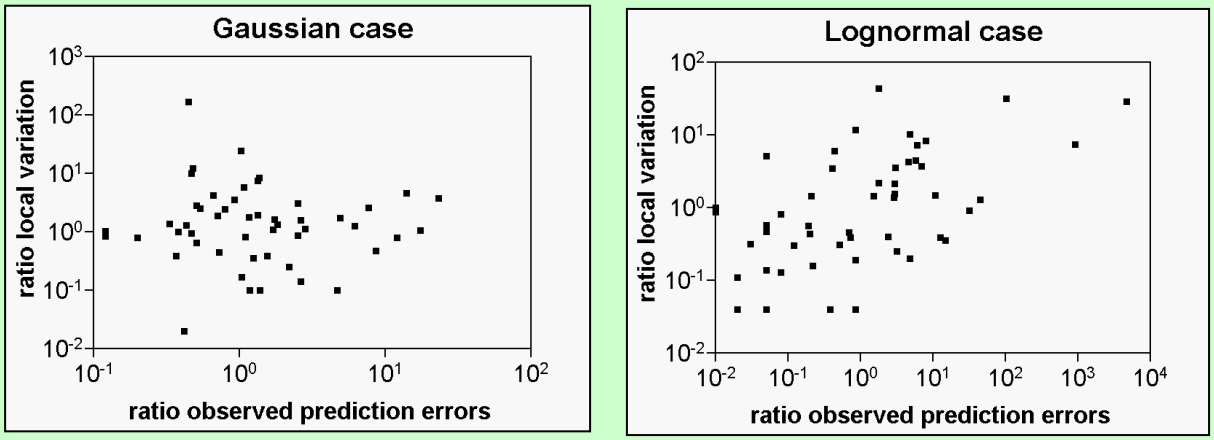

Figure 3. Scatter plots of ratio of observed absolute prediction errors against ratio of local variation. Left: Gaussian case. Right: Lognormal case.

It turns out that there is no relationship between the two ratios: there is agreement for exactly 25 out of 50 realisations. This is confirmed by Figure 3 (Left), where the scatter plot shows that the two ratios are independent. What this means is that there is no relationship between the degree of local variation and the magnitude of interpolation error. The numerical experiment confirms the analytical result, as it should. It also means that human intuition can be quite deceiving, because the argument that "it is clear that the potential for prediction error ought to be greater at a location surrounded by data that are very different from one another than at a location surrounded by similarly valued data" (Goovaerts, 1997, p. 180) simply does not hold.

Of course we should realise that this is true only in the Gaussian situation. We therefore repeated the simulation experiment but now with a non-Gaussian random function. We extended step 1 of the procedure above by taking the antilog of the realisation before moving on to the next steps. This resulted in realisations of a lognormal random function, which were

very skew as evidenced by an example realisation and its frequency distribution given in Figure 4. The results of the analysis are given in Figure 3 (Right). It now turns out that there is a relationship between the ratios, although there still is considerable scatter. Large prediction errors mainly occur there where local variation is large. For 38 of the 50 realisations the largest prediction error occurs there where the local variation is largest, which is significantly (α=0.001) different from what would be expected under the null hypothesis of no dependence.

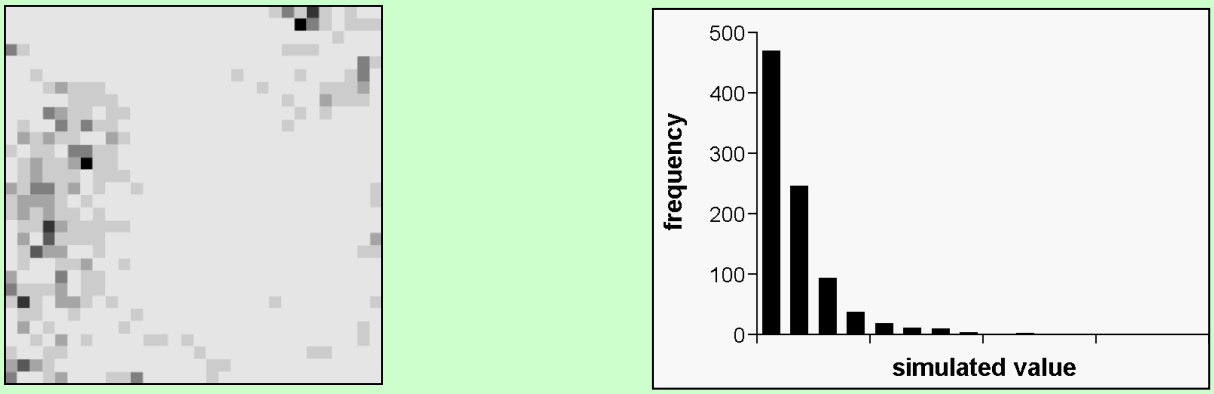

Figure 4. One of the 50 realisations generated for the lognormal case (Left) and frequency distribution of its 900 cell values (Right).

5. DISCUSSION AND CONCLUSIONS

Both the mathematical-statistical and numerical analysis show that it is not a matter of course that interpolation error should be larger in areas where local variation is larger. When it is reasonable to assume that the attribute under study is a realisation of a stationary Gaussian random function, then the prediction error variance should be independent of the data values.

This is an important and remarkable finding.

It is important, because it has important consequences for the way in which we react to observed differences in local spatial variation. For stationary Gaussian random functions, the local variation can markedly vary between locations. In the numerical experiment carried out in this paper it was not uncommon for the local variation in one corner to be five times as large as the local variation in the other corner. Apparently local variations in realisations of stationary Gaussian random fields can vary much more than one would normally expect. The differences may be less pronounced when a different variogram or more observations had been taken, but there is no doubt in our mind that many a practitioner would decide from the observations given in the middle map in Figure 2 that he or she is dealing with a situation of nonstationarity and/or non-normality. He or she would then probably but unrightfully force the interpolation error to be smaller in the bottom left corner.

It is also a remarkable finding, because it runs counter to what one would intuitively expect. Although the independence of the conditional variance to the data values in the Gaussian situation may be well known to statisticians and informed geostatisticians, many practitioners still find it hard to accept, because the result is so counter-intuitive. It is therefore important to emphasise that indeed there is an important class of situations where the prediction error variance is independent of the data values. Alternative kriging techniques, notably indicator kriging, by construction will always yield larger prediction error variances there where local variation is larger. Indicator kriging will also do so in the stationary Gaussian situation, which

is incorrect. Apparently, one can also rely too much on the data in building a model of spatial uncertainty.

One other conclusion that we can draw is that it is crucially important to be able to tell whether the stationarity and Gaussian assumptions (better to say: decisions) are realistic or not. The choice that is made here determines whether the SK and OK variances are proper measures of interpolation error or not. The simulation experiment showed that even in the stationary Gaussian case there can be large differences in local variation. What is needed is an unambiguous method that allows to determine how likely it is that a given dataset derived from a multiGaussian distribution.

It would also be interesting to analyse how robust the SK and OK variances are to certain deviations from the Gaussian assumption. The results of the numerical experiment showed that the dependence between absolute prediction error and local spatial variation still was rather small for the lognormal case, even though the distribution was very skew (see Figure 4). In fact, the distribution was that skew that no skilled geostatistician would apply a direct kriging interpolation to it without first applying some kind of transformation of the data.

References

Bachmann, A. and Allgöwer, B., 2002, "Uncertainty propagation in wildland fire behaviour modelling". *International Journal of GIS*, 16, pp. 115-127.

Burrough, P.A., 1986, *Principles of Geographical Information Systems for Land Resources Assessment*, Clarendon Press, Oxford.

Canters, F., De Genst, W. and Dufourmont, H., 2002, "Assessing effects of input uncertainty in structural landscape classification". *International Journal of GIS*, 16, pp. 129-149.

Goovaerts, P., 1997, *Geostatistics for Natural Resources Evaluation*, Oxford University Press.

Journel, A.G., Kyriakidis, P.C. and Mao, S., 2000, "Correcting the smoothing effect of estimators: a spectral postprocessor". *Mathematical Geology*, 32, pp. 787-813.

Kros, J., Pebesma, E.J., Reinds, G.J. and Finke, P.F., 1999, "Uncertainty assessment in modelling soil acidification at the European scale: a case study". *Journal of Environmental Quality*, 28, pp. 366-377.

Kyriakidis, P.C. and Dungan, J.L., 2001, "A geostatistical approach for mapping thematic classification accuracy and evaluating the impact of inaccurate spatial data on ecological model predictions". *Environmental and Ecological Statistics*, 8, pp. 311-330.

McBratney, A.B., Webster, R. and Burgess, T.M., 1981, "The design of optimal sampling schemes for local estimation and mapping of regionalized variables: 1. Theory and method", *Computers and Geosciences*, 7, pp. 331-334.

Van Groenigen, J.W., Siderius, W. and Stein, A., 1999, "Constrained optimisation of soil sampling for minimisation of the kriging variance", *Geoderma*, 87, pp. 239-259.

Yamamoto, J.K., 2000, "An alternative measure of the reliability of ordinary kriging estimates", *Mathematical Geology*, 32, pp. 489-509.