

UC Santa Barbara

Core Curriculum-Geographic Information Science (1997-2000)

Title

Unit 098 - Uncertainty Propagation in GIS

Permalink

<https://escholarship.org/uc/item/2110t9ns>

Authors

098, CC in GIScience
Heuvelink, Gerard B.M.

Publication Date

2000

Peer reviewed

Unit 098 - Uncertainty Propagation in GIS

Written by: Gerard B.M. Heuvelink, Faculty of Environmental Sciences,
University of Amsterdam, Nieuwe Prinsengracht 130, 1018 VZ
Amsterdam, The Netherlands

This section was edited by Gary Hunter, Department of Geomatics, University of Melbourne, Australia.

This unit is part of the *NCGIA Core Curriculum in Geographic Information Science*. These materials may be used for study, research, and education, but please credit the author and the project, *NCGIA Core Curriculum in GIScience*. All commercial rights reserved. Copyright 1997 by Gerard B.M. Heuvelink.

Advanced Organizer

Unit Topics

- this unit outlines
 - an introduction to the problem of uncertainty propagation in GIS
 - the definition and identification of a stochastic error model for quantitative spatial attributes
 - a description of common error propagation techniques
 - applications of the theory
 - how the results of an uncertainty analysis may be used to improve the accuracy of GIS products

Intended Learning Outcomes

- after reading this unit, you should be able to
 - present an overview of the main areas where error propagation within GIS is currently of concern
 - describe how errors in spatial attributes can be defined using statistical terminology
 - discuss the principles of common error propagation techniques and their pro's and con's
 - have an idea about how the theory of error propagation in GIS may be applied in

- practice
- have sufficient clues and references to dig into this problem more thoroughly if interested

[Instructors' Notes](#)

[Full Table of Contents](#)

[Metadata and Revision History](#)

Uncertainty Propagation in GIS

1. Introduction

- One of the most powerful capabilities of GIS, particularly for the earth and environmental sciences, is that it allows to derive new attributes from attributes already held in the GIS database. The many basic types of function used for derivations of this kind are often provided as standard functions or *operations* in many GISs, under the name of "*map algebra*" (link to other relevant unit(s)).
- No map stored in a GIS is truly error-free. Note that the words used here in its widest sense to include not only "mistakes" and "blunders", but also to include the statistical concept of error meaning "variation" (in this text, the words "error" and "uncertainty" are treated as synonymous).
- When maps that are stored in a GIS database are used as input to a GIS operation, then the errors in the input will *propagate* to the output of the operation. Moreover, the error propagation continues when the output from one operation is used as input to an ensuing operation. Consequently, when no record is kept of the accuracy of intermediate results, it becomes extremely difficult to evaluate the accuracy of the final result.
- Although users may be aware that errors propagate through their analyses, in practice they rarely pay attention to this problem. No professional GIS currently in use can present the user with information about the confidence limits that should be associated with the results of an analysis.
- The purpose of this unit is to describe a methodology for handling error and error propagation in spatial modelling with GIS. Note that this unit mainly deals with the propagation of *quantitative attribute* errors in GIS. However, the propagation of *positional* errors can be studied using a similar approach. The propagation of *categorical* errors is more difficult because it involves error probability distributions that cannot easily be reduced to a few parameters.

2. Definition of an error model for quantitative spatial attributes

- The "error" in a quantitative attribute can be conveniently defined as the difference between *reality* and our representation of *reality* (i.e. the map). For instance, if the nitrate concentration of the shallow groundwater at some location equals 68.6 g/m³, while according to the map it is 62.9 g/m³, then there will be no disagreement that in this case the error is 68.6-62.9=5.7⁰ g/m³. Generalising this example, let the true value of a spatial attribute at some location x be $a(x)$, and let the representation of it be $b(x)$. Then, according to the definition, the error $v(x)$ at x is simply the arithmetical difference $v(x)=a(x)-b(x)$.
- We should be well aware that the error $v(x)$ is never exactly known, because if it were, then it could simply be eliminated. Rather, knowledge about $v(x)$ is limited to specifying a range or distribution of possible values. This type of information can best be conveyed by representing the error as a random variable $V(x)$. For instance, since we do not know the true nitrate concentration of the shallow groundwater, we may think that it is a value drawn from a large set of values that surround the estimated value of 62.9 g/m³. Although we are aware that the attribute has only one fixed, deterministic value $a(x)$, our uncertainty about $a(x)$ allows us to treat it as the outcome of some random mechanism $A(x)$. We must now proceed by specifying the probability distribution of the error $V(x)$.
- First consider the error at a single location x only. Denote the mean of $V(x)$ by $u(x)$ and the variance by $q^2(x)$. The mean $u(x)$ represents the systematic error or bias, the standard deviation $q(x)$ characterises the on-systematic, random component of the error $V(x)$.
- Next consider the spatial and multivariate extension of the error model. Let x and x' be two locations. Apart from the means and variances of $V(x)$ and $V(x')$ we now also need to specify their spatial auto-correlation $p(x,x')$. For instance, the error in the nitrate concentration of the shallow groundwater might be spatially correlated as a negative exponential $p(x,x')=\exp[-0.004*|x-x'|]$, implying that the errors at two locations 500 meters apart equals 0.14. When there are multiple attributes $A_i(x)$ and errors $V_i(x)$, $i=1,\dots,m$, then for each of the attributes an error model $A_i(x)=b_i(x)+V_i(x)$ must be defined, where each error $V_i(x)$ follows some distribution with mean $u_i(x)$ and variance $q_i^2(x)$, and where the (cross-)correlation of $V_i(x)$ and $V_j(x')$ may be denoted by $p_{ij}(x,x')$.
- To illustrate that errors in spatial attributes are often correlated, consider the example of soil pollution by heavy metals, such as is the case in the river Geul valley, in the south of the Netherlands. Maps of the concentrations of lead and cadmium in the soil are obtained from interpolating point observations. In this case the interpolation errors $V_{\text{lead}}(x)$ and $V_{\text{cadmium}}(x)$ are likely to be positively correlated, because *unexpectedly*

high lead concentrations will often be accompanied by *unexpectedly* high cadmium concentrations. Unforeseen low concentrations will also often occur simultaneously.

- The observation that errors in spatial attributes are often correlated is important because in what

follows we will see that presence of non-zero correlation can have a marked influence on the outcome of an error propagation analysis.

3. Identification of the error model

- To estimate the parameters of the error random field V in practice, certain stationarity assumptions have to be made. This can be done in various ways. The most obvious way is to impose the assumptions directly on V . This is acceptable when inference on V is based solely on observed errors at test points. For instance, to assess the error standard deviation of an existing DEM it may be sensible to assume that the standard deviation of the error is spatially invariant, so that it can be estimated by the Root Mean Squared Error, computed from the differences between the DEM and the true elevation at the test points. In addition, it may be sensible to assume that the spatial auto-correlation $\rho(x, x')$ is a (decreasing) function of only the distance $|x - x'|$, such as the example of the negative exponential given before.
- In many cases it is advisable not to assess the error parameters after the map has been made, but to include the uncertainty assessment in the mapping procedure itself. This unit does not go into why this is advisable and how it should be done. Here we will only mention that it often involves the use of kriging (link to core curriculum^{02.13}).

4. Error propagation techniques

- The discussion hereafter will be confined to point operations, i.e. GIS operations that operate on each spatial location x separately. This is no principal restriction because non-point operations can be handled by minor modification. For notational convenience, the spatial index x will be dropped.
- The error propagation problem can be formulated mathematically as follows. Let U be the output of a GIS operation g on the m input attributes A_i :

$$U = g(A_1, A_2, \dots, A_m) \quad (1)$$

The objective is to determine the error in the output U , given the operation g and the errors in the input attributes A_i . Thus our main interest is in the uncertainty of U , as contained in its variance t^2 .

- It must first be observed that the error propagation problem is relatively easy when g is a linear function. In that case the mean and variance of U can be directly and analytically derived. However, for the general situation analytical methods are not very suitable. Two alternative methods will now be discussed.

4.1 Taylor series method

The idea of the Taylor series method is to approximate g by a linear function that is locally a good approximation of g . The linearization greatly simplifies the error analysis, but at the expense of introducing an approximation error. The resulting expression shows that the variance of U is the sum of various terms, which contain the correlations and standard deviations of the A_i and the first derivatives of g :

$$\tau^2 \approx \sum_{i=1}^m \sum_{j=1}^m \rho_{ij} \sigma_i \sigma_j \left(\frac{\partial g}{\partial a_i} \right) \left(\frac{\partial g}{\partial a_j} \right) \quad (2)$$

The derivatives reflect the sensitivity of U for changes in each of the A_i . Note that the correlations of the input errors can have a marked effect on the variance of U .

4.2 Monte Carlo method

- The Monte Carlo method uses an entirely different approach to analyse the propagation of error through the GIS operation. The idea of the method is as follows:
 - repeat N times:
 - generate a set of realisations $a_i, i=1, \dots, m$
 - for this set of realisations a_i , compute and store the output $u=g(a_1, \dots, a_m)$
 - compute and store sample statistics from the N outputs u
- The accuracy of the Monte Carlo method is inversely related to the square root of the number of runs N . This means that to double the accuracy, four times as many runs are needed. The accuracy thus slowly progresses as N increases.

4.3 Comparison of error propagation techniques

- The main problem with the Taylor method is that the results are approximate only. It will not always be easy to determine whether the approximations involved using this method are acceptable. The Monte Carlo method does not suffer from this problem, because it can reach an arbitrary level of accuracy.
- With the Monte Carlo method, high accuracies are reached only when the number of runs is sufficiently large, which may cause the method to become extremely time consuming. Another disadvantage of the Monte Carlo method is that the results do not come in an analytical form.
- As a general rule it seems that the Taylor method may be used to obtain crude

preliminary answers. These should provide sufficient detail to be able to obtain an indication of the quality of the output of the GIS operation. When exact values or percentiles are needed, the Monte Carlo method may be used. The Monte Carlo method will probably also be preferred when error propagation with complex operations is studied, because the method is easily implemented and generally applicable.

5. Examples

- Let us consider a few simple examples to get a feel of how error propagation may be applied in practice:
 - Let the estimated cadmium concentration at some location in the Geul river valley be $4.7 \text{ }^{\circ}\text{mg/kg}$ with estimation uncertainty $\sigma_{\text{Cd}}=1.2 \text{ mg/kg}$. Let the estimated lead concentration at the same location be 210°mg/kg with $\sigma_{\text{Pb}}=35 \text{ mg/kg}$. Let a risk factor R be defined as $R=\text{Pb}+13*\text{Cd}$ (cadmium is 13 times as harmful as lead). With equation (2) it is not difficult to verify that when the errors in cadmium and lead are uncorrelated, then the estimated risk factor will be 271.1 mg/kg with associated uncertainty 38.3°mg/kg . Observe also that if the errors would have been positively correlated with correlation coefficient 0.8, then the estimation error would increase to 43.6°mg/kg (why an increase?).
 - Assume that a soil sample consists of only the three fractions clay, silt and sand. These fractions must add up to one but there is uncertainty about the individual fractions: clay= 0.278 ± 0.045 , silt= 0.419 ± 0.073 , sand= 0.303 ± 0.052 . The correlations between the errors are $\rho_{\text{clay,silt}}=-0.591$, $\rho_{\text{clay,sand}}=-0.561$, $\rho_{\text{silt,sand}}=-0.467$ (why are they negative?). Applying equation (2) yields that the sum of the three fractions equals 1.000 ± 0.000 (verify this). In fact, this is as expected, because there cannot be any uncertainty about the sum of the three fractions, it will always equal one.
 - In his 1986 book, Burrough gives the example of the Universal Soil Loss Equation $A=R*K*L*S*C*P$, with A the annual soil loss in tonne/ha, R a measure of erosion caused by rainfall, K the erodibility of the soil, L the slope length in $^{\circ}\text{m}$, S the slope in per cent, C is the cultivation parameter and P represents protection measures. The values of the factors and their error standard deviations used by Burrough are: $R=297 \pm 72$, $K=0.10 \pm 0.05$, $L=2.130 \pm 0.045$, $S=1.169 \pm 0.122$, $C=0.50 \pm 0.15$ and $P=0.50 \pm 0.10$. Verify that when all errors are assumed uncorrelated and when the Taylor method is applied, this yields $A=18.5 \pm 12.4$. However, recall that in this case the Taylor method is approximate only because the USLE model is highly non-linear. The Monte Carlo method is more appropriate here and yields the solution $A=18.3 \pm 13.2$ (this can also be verified but it would require a computer exercise). Apparently,

in this case the Taylor method does not such a bad job after all.

- More elaborate examples that are merely described here are:
 - The uncertainty in a DEM can have a profound effect on derived attributes such as slope and aspect. For this class of neighbourhood operations the spatial autocorrelation of error becomes a crucial factor. Several studies have demonstrated that the error in the derived products decreases as the degree of spatial autocorrelation increases (explain this).
 - In soil science, expensive-to-measure soil attributes are often derived from cheaper ones using so-called pedo-transfer functions, which are often regression-type functions. These procedures involve the propagation of model error (the residual variance and the uncertainty about the regression coefficients) and input error (measurement and interpolation errors in the independent variables of the regression). The error analysis will not only produce the uncertainty in the output of the transfer function, but it can also show how much the individual error sources contribute to the final output error

6. Discussion and conclusions

- There is no perfect, easy method to analyse the propagation of errors in spatial modelling with GIS. Nonetheless, it can be done and the available methods are in a sense complementary.
- Error propagation can only be used once the input errors to the analysis are available. Unfortunately, in practice often there will only be crude and incomplete estimates of input error available. It is important that map makers become aware that they should routinely convey the accuracy of the maps they produce, even when accuracy is less than expected. It is also important that GIS manufacturers increase their efforts to add error propagation functionality to their products.
- The ability to determine how much each individual input contributes to the output error is extremely valuable. It allows users to explore how much the quality of the output improves, given a reduction of error in a particular input.
- When there are multiple error sources then in many cases it will be most rewarding to strive for a *balance of errors*. When the error in an attribute has a marginal effect on the output, then there is little to be gained from mapping it more accurately. In that case, extra sampling efforts can much better be directed to an input attribute that has a larger contribution to the output error. For instance, if a pesticide leaching model is sensitive to soil organic carbon and less so to soil bulk density, then it is more important to map the former more accurately.
- Error analysis may also be used to compare the contributions of input and model error. It is clearly unwise to spend much effort on collecting data if what is gained is immediately thrown away by using a poor model. On the other hand, a simple model

may be as good as a complex model if the latter needs lots of data that cannot be accurately obtained.

7. Reference Material

- Burrough, P.A. and McDonnell, R.A., 1998, *Principles of Geographical Information Systems*, Oxford: Oxford University Press.
- Forier, F. and Canters, F., 1996, A user-friendly tool for error modelling and error propagation in a GIS
- environment. In: Mowrer, H.T., Czaplewski, R.L. and Hamre, R.H. (Eds.) *Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*. Fort Collins, USDA Forest Service General Technical Report RM-GTR-277: 225-34.
- Goodchild, M.F., Sun, G. and Yang, S., 1992, Development and test of an error model for categorical data, *International Journal of GIS*, 6, 87-104.
- Hammersley, J.M. and Handscomb, D.C., 1979, *Monte Carlo Methods*, London: Chapman and Hall.
- Heuvelink, G.B.M., 1998, *Error Propagation in Environmental Modelling with GIS*. London: Taylor & Francis.
- Heuvelink, G.B.M., Burrough, P.A., Stein, A., 1989, Propagation of errors in spatial modelling with GIS. *International Journal of Geographical Information Systems*, 3: 303-322.
- Hunter, G.J., Goodchild, M.F., 1997, Modeling the uncertainty of slope and aspect estimates derived from spatial databases. *Geographical Analysis*, 29: 35-49.
- Kiiveri, H.T., 1997, Assessing, representing and transmitting positional uncertainty in maps, *International Journal of GIS*, 11, 33-52.
- Lanter, D.P. and Veregin, H., 1992, A research paradigm for propagating error in layer-based GIS, *Photogrammetric Engineering & Remote Sensing*, 58, 825-833.
- Pebesma, E.J. and Wesseling, C.G., 1997, Gstat, a program for geostatistical modelling, prediction and simulation, *Computers and Geosciences* (in press).
- Stanislawski, L.V., Dewitt, B.A. and Shrestha, R.L., 1996, Estimating positional accuracy of data layers within a GIS through error propagation. *Photogrammetric Engineering & Remote Sensing*, 62: 429-433.
- Taylor, J.R., 1982, *An Introduction to Error Analysis*. Mill Valley: University Science Books.

- Van Deursen, W.P.A. and Wesseling, C.G., 1995, PCRaster Software, Department of Physical Geography, University of Utrecht.
 - Wingle, W.L., Poeter, E.P. and McKenna, S.A., 1996, UNCERT: a geostatistical uncertainty analysis \package applied to groundwater and contaminant transport modeling.
-

Citation

To reference this material use the appropriate variation of the following format:

Heuvelink, Gerard B.M., (1998) Geographic Information Technologies in Society, *NCGIA Core Curriculum GIScience*,
<http://www.ncgia.ucsb.edu/giscc/units/u098/u098.html>, posted February 05, 1998.

Last revised: February 05, 1998.

Uncertainty Propagation in GIS (098)

Instructors' Notes

- Teaching this unit to students requires that they (and you!) have some background in mathematics and statistics.

[Back to the Unit](#)

Unit 098 - Uncertainty Propagation in GIS

Table of Contents

[Advanced Organizer](#)

[Unit Topics](#)

[Intended Learning Outcomes](#)

[Instructors' notes](#)

[Metadata and revision history](#)

[Body of unit](#)

1. [Introduction to Uncertainty Propagation in GIS](#)
2. [Definition of an error model for quantitative spatial attributes](#)
3. [Identification of the error model](#)
4. [Error propagation techniques](#)
 1. Taylor series method
 2. Monte Carlo method
 3. Comparison of error propagation techniques
5. [Examples](#)
6. [Discussion and conclusions](#)
7. [Reference Material](#)

[Citation](#)

[Back to the Unit](#)

Unit 098 - Uncertainty Propagation in GIS

Metadata and Revision History

1. About the main contributors

- Gerard BM Heuvelink

Faculty of Environmental Sciences
University of Amsterdam, Nieuwe Prinsengracht

2. Details about the file

- unit title
 - Uncertainty Propagation in GIS
- unit key number
 - 098

3. Key words

4. Index words

5. Prerequisite units

6. Subsequent units

7. Other contributors to this unit

8. Revision history

- Created on January 29, 1998

[Back to the Unit.](#)