

Practical Geostatistics 2000 series Case Studies 2009

Isobel Clark
Geostokos (Ecosse) Limited,
Scotland

William V Harper
Otterbein College, Westerville,
Ohio, USA

PRACTICAL Geostatistics 2000

Case Studies

Isobel Clark
Geostokos (Ecosse) Limited, Scotland

William V Harper
Otterbein College, Westerville OH, USA



Ecosse North America LLC, Columbus, Ohio, USA.



First US printing, November 2008

Copyright © 2008 by Geostokos (Ecosse) Limited, Scotland.

All rights reserved.

Reproduction or translation of any part of this work beyond that permitted by Section 107 or 108 of the 1976 United States Copyright Act without the permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to Permissions Department, Geostokos (Ecosse) Limited, Alloa Business Centre, Whins Road, Alloa FK10 3SA, Central Scotland.

Tel: +44 773 663 1632 Fax: +44 1259 215274

e-mail: PG2000_CS@kriging.com

<http://www.kriging.com> <http://ecossenorthamerica.com>

Library of Congress Cataloging-in-Publication Data:

Clark, Isobel

Practical geostatistics 2000, case studies / Isobel Clark, William V. Harper.
382p. 2.5cm.

ISBN: 0-9703317-4-6

ISBN: 978-0-9703317-4-8

1. Geology—Statistical methods. 2. Mining engineering. 3. Environmental engineering I. Harper, William V. II. Title.

QE33.2.S82 C431 2000

519.5'02'4553—dc21

Ecosse North America LLC, Columbus, Ohio, USA, publishers.

Typeset using Scientific Word 5.5, MacKichan Software Inc,

<http://www.mackichan.com>

**This one is for Eleanor
who has already learned how to say "I don't know".**

and for Paula and Rachel who bring joy to life.

For those who seek definitive answers:

*“Let us all be quiet in unison
and hear what the oracle has to tell us.*

*The oracle says, when winter comes
the autumn is over. Are we comforted?”*

Edwin Morgan

{blank page}

CONTENTS

Preface

Notation

1. Introduction **Page 1**

1.1 What we are trying to do here

1.2 Geostatistics in the real world

1.3 Notes for teachers

1.4 Cautionary note

1.5 Data sets

2. Coal **Page 13**

2.1 Relationship between variables

2.1.1 Correlation coefficient

2.1.2 Predicting one variable from another

2.1.3 Calorific Value versus Ash Content

2.1.4 Calorific Value versus Sulphur Content

2.1.5 Other combinations of variables

2.1.6 Relationship involving more than two variables

2.1.7 Predicting Sulphur from Calorific Value and Ash Content

2.2 Calorific values

2.2.1 Statistical analysis

2.2.2 Inverse distance interpolation

2.2.3 Semi-variogram calculation and modelling

2.2.4 Estimation

2.2.5 In the real world

2.3 Sulphur content

2.3.1 Semi-variogram calculation and modelling

2.3.2 Cross validation

2.3.3 Kriging of Sulphur values

3. GASA

Page 57

3.1 Statistical analysis

3.1.1 reef widths

3.1.2 Sichel's t estimation

3.1.3 Estimating the mean of a lognormal population

3.1.4 gold values

3.1.5 The three parameter lognormal distribution

3.1.6 More complex distribution and mixtures

3.2 Two variables or One variable?

3.3 Spatial estimation

3.4 Geostatistical methods

3.4.1 Using similar information from another source

3.4.2 Realistic grade tonnage curves for production

4. Geevor

Page 91

4.1 Summary of the Geevor study

4.2 Statistical analysis

4.2.1 thickness of vein

4.2.2 tin grades

4.2.3 One variable or Two?

4.3 Geostatistics

4.3.1 Semi-variograms

4.3.2 Nugget effect or Sampling error?

4.3.3 Summary of nugget effect study

4.3.4 Cross validation

4.3.5 Kriging a map – point estimation

4.3.6 The lognormal kriging transformation

4.3.7 The lognormal shortcut

5. Organics

Page 113

5.1 Statistical study

5.1.1 Possible outlier

5.2 Inverse distance interpolation

5.3 Semi-variogram calculation and modelling

5.3.1 trend surface analysis

5.3.2 studying the residuals from the trend surface

5.3.3 Summary of semi-variogram calculation and modelling

5.4 Cross validation

5.4.1 other things to do with cross validation statistics

5.5 Kriging

6. Scallops

Page 131

6.1 Summary statistics

6.1.1 Fitting a statistical distribution: the Normal distribution

6.1.2 Fitting a distribution: the lognormal model

6.1.3 Estimation of the mean of the study area

6.1.4 Confidence levels on the mean

6.2 Inverse distance interpolation

6.3 Lognormal kriging

6.3.1 kriging with logarithms

6.3.2 Worked example for lognormal kriging

6.3.3 cross validation

6.3.4 mapping with lognormal kriging

6.4 Rank uniform kriging

7. Sunshine Mine

Page 161

7.1 Silver grade

7.1.1 Estimation of the mean of the study area

7.1.2 Confidence limits on the mean

7.1.3 Grade tonnage curves

7.2 Semi-variogram calculation and modelling

7.2.1 Cross validation of semi-variogram model

7.3 Mapping the silver values

7.3.1 kriging values over stoping panels

8. Velvetleaf **Page 175**

8.1 Finding the appropriate distance function – the semi-variogram

8.1.1 Cross validate the semi-variogram model

8.2 A more realistic(?) analysis

8.2.1 Indicator transform at 1 (well 0.5!)

8.2.2 Indicator transform at 40

8.2.3 Indicator at 40 given that weeds are present

9. Wolfcamp **Page 197**

9.1 The data

9.2 Summary of chapter

9.3 Descriptive statistics

9.4 Spatial data

9.5 Calculating the semi-variogram

9.6 Polynomial trend surface analysis

9.6.1 Recalculating the semi-variogram

9.6.2 Universal kriging – the process and the impact of the nugget effect

9.7 Cross Validation

9.8 Kriging

9.9 kriging over an area

10. Brooms Barn **Page 227**

10.1 Potassium (K) values

10.1.1 Semi-variogram calculation and interpretation for K values

10.1.2 Fitting and testing a semi-variogram model

10.1.3 judging how well the model fits the data

10.1.4 cross validation

10.1.5 when is an outlier not an outlier?

10.1.6 maybe these outliers are actually outliers

10.2 Kriging maps for K values

10.3 One more time: K without the outliers

10.3.1 cross validation for $K < 50$

10.3.2 kriging the swathes

10.4 A different variable – pH

10.4.1 pause to review the data again

10.4.2 mapping pH without knowing the field boundaries

10.4.3 Combining the components for a map of pH

10.4.4 What if we just kriged it normally?

11. Old Brenda

Page 265

11.1 Statistical analysis of the copper values

11.2 Semi-variogram calculation and modelling

11.3 Cross validation

11.4 Mapping copper values level by level

11.5 Block kriging for mine planning

11.6 Trend Surface Analysis

12. Bibliography

Page 279

Tables

Data Sets

{blank page}

Preface

This is a companion volume to Practical Geostatistics 2000. Here we present data sets worked as complete case studies. We have chosen mostly to present larger data sets in this volume. All of these can be found either in the Appendix at the end of the book and/or on our web site at <http://www.kriging.com>.

If you have questions please contact one or other of us or visit <http://www.kriging.com> or <http://ecossenorthamerica.com/> to check whether your question has already been asked and answered.

Given the nature of the universe, we can give no guarantee that this volume is typo free. We have taken as much care as we can but none of us are perfect. We apologise for any stress caused by any remaining errors in the text.



Isobel Clark
Geostokos (Ecosse) Limited
Alloa Business Centre
Whins Road, Alloa, FK10 3SA
Central Scotland
Tel: 01259 726600
Fax: 01259 215274
e-mail: PG2000_CS@kriging.com
<http://uk.geocities.com/drisobelclark>



William V. Harper
Otterbein College, Mathematical Sciences
Towers Hall 139, 1 Grove Street
Westerville, OH 43081-2006 U.S.A
Phone: 614-823-1417 Fax: 614-823-3201
WHarper@otterbein.edu
<http://ecossenorthamerica.com/>

{blank page}

Notation

<i>a</i>	the range of influence or distance scaling parameter in a model semi-variogram
<i>b</i>	breadth of a panel or block under study
<i>b_k</i>	coefficients estimated in a regression or trend surface analysis
<i>c</i>	cutoff or ‘pay’ value used in a grade/tonnage calculation
<i>d</i>	distance between a sample and a location at which an estimate is required
<i>e</i>	2.718281828..... the natural anti-logarithm function
<i>f</i>	a secondary or ancillary variable measured on a samples
<i>g</i>	the primary value measured at a sample location
<i>h</i>	the distance between any two specified locations within the study area
<i>i</i>	a subscript identifier
<i>j</i>	a subscript identifier
<i>k</i>	a subscript or integer power
<i>l</i>	length of a panel or block under study
<i>m</i>	number of samples used in a ‘local’ estimation, such as inverse distance or kriging
<i>n</i>	number of samples available in the current study
<i>p</i>	slope of the line in a linear semi-variogram model
<i>r</i>	estimate for the correlation coefficient between two variables
<i>s</i>	unbiased estimate for the standard deviation of a population
<i>t</i>	Student’s distribution for the arithmetic mean of Normal samples Sichel’s estimator for the average of a lognormal population
<i>v</i>	a volume
<i>w</i>	weighting factor
<i>x</i>	standardised Normal variate, mean 0 standard deviation 1
<i>y</i>	logarithm of <i>g</i> value
<i>z</i>	residual from a trend surface
A	an area or volume whose value is to be estimated
C	nugget effect C_0 or sill C_i on a semi-variogram model
F	auxiliary function for the variance within a panel or block
H	auxiliary function for a corner point with a 2d panel
M	a specified point within an area or volume
N	number of pairs of samples found in a semi-variogram calculation
P	proportion of the population above a specified cutoff value or pay limit may also be used to represent percentage rather than proportion
Q	an auxiliary function between two ‘lines’ at right angles
S	sum of squared residuals from a regression exercise
T	the unknown value at an unsampled location the distribution of Sichel’s <i>t</i> estimate for the lognormal mean
V	the variance of samples around their arithmetic mean
X	co-ordinate for a location, usually east/west
Y	co-ordinate for a location, usually north/south
Z	co-ordinate for a location, usually elevation or depth

Notation continued

lower case Greek letters

α	alpha	power to which distance is raised in a semi-variogram model the significance level of a hypothesis test
β	beta	additive constant used in a three parameter lognormal distribution
γ	gamma	Sichel's correction factor for the lognormal mean estimation a model semi-variogram or function calculated from such
δ	delta	a small change
ϵ	epsilon	the error of estimation, actual value—estimated value
θ	theta	an angle usually expressed in degrees
λ	lambda	lagrangian multiplier
μ	mu	arithmetic mean of values in a Normal population
ν	nu	degrees of freedom
ξ	eta	damping or decay parameter in a hole effect semi-variogram model
π	pi	3.141592654.... the ratio of the perimeter of a circle to its diameter
ρ	rho	the population correlation coefficient between two variables
σ	sigma	the standard deviation of values in a Normal population
τ	tau	the arithmetic mean of values in a lognormal population
ϕ	phi	the standard Normal density function
χ	chi	the distribution of sample variances from Normal samples
ω	omega	the standard deviation of values in a lognormal population

Upper case Greek letters

Σ	Sigma	a summation sign, e.g. $\sum_{i=1}^n$ is “sum for series $i=1,2,3,\dots,n-1, n$ ”
Φ	Phi	cumulative proportion under the standard Normal distribution
Ψ	Psi	Sichel's confidence factors for the lognormal distribution

1. Introduction

1.1. What we are trying to do here

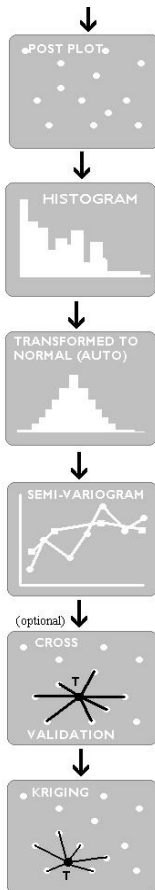
This volume is complementary to Practical Geostatistics 2000. It was produced as a further companion to *Practical Geostatistics 2000, Answers to the Exercises* on the suggestion of some of our teaching colleagues. They thought it would be a good idea to present the analysis of a data set as a “case study”, following the analysis through from first statistical summaries to final kriging maps. Each chapter here is a study of a single data set.

To read this book and benefit from it you need to be fairly comfortable with basic algebra. That is, with the notion of using symbols as shorthand for longer statements. We have worked hard to bring you a consistent notation throughout the textbooks and keep the same conventions in these case studies.

The data sets and software used in these analyses can be downloaded from the Web at http://www.kriging.com/teaching_software.

As in the main textbook, we do advise you to keep your sense of humour and common sense to the fore at all times while reading this book. These sort of analyses work much better if you expect the answers to make sense.

1.2. Geostatistics in the real world



In many textbooks, documentation and professional reports, geostatistics is presented as a straightforward step-by-step process. Examples which are used for teaching processes are often chosen *because* they are straightforward and uncomplicated.

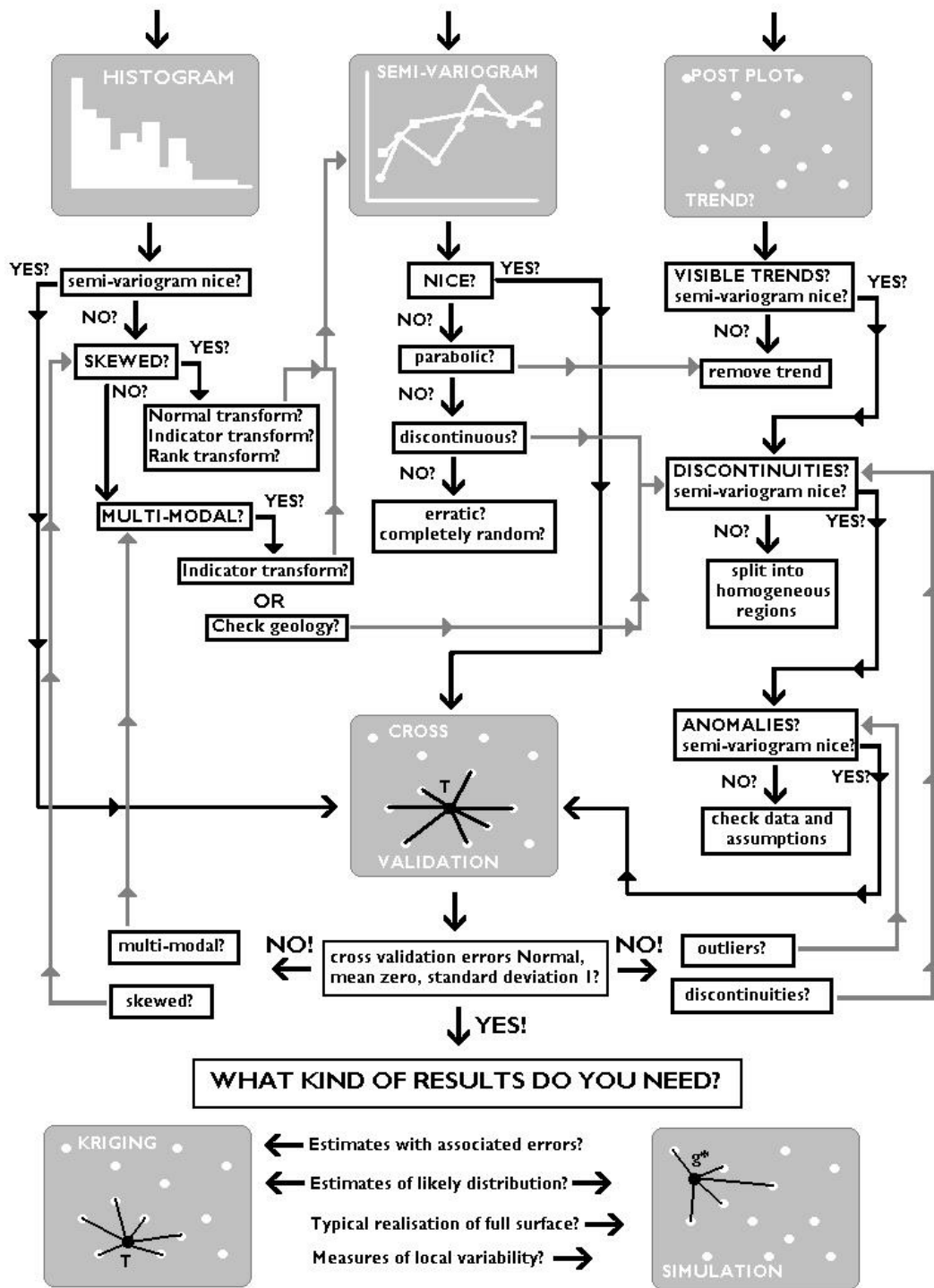
In this volume, we have tried to present examples which are a little more interesting than the usual classroom exercises. Most of these data sets are of reasonable size – from around 100 samples to over 5,000 samples.

Some analyses presented here follow the simple flow illustrated in the figure to the left. For example, the coal data is pretty linear.

Others are more realistic and require a more iterative approach – refining the interpretation and models to find an optimal technique for a particular data set. For example, the Velvet Leaf data set repays close attention to its statistical nature.

Some data sets, such as the scallops data, can be tackled in a variety of ways which may or may not produce similar results.

Finally, it is worth re-iterating that the case studies provided in this volume are intended as examples only and form no definitive guide as to how to analyse other data sets or similar problem.



Still better than the real thing

Comments and suggestions are always welcome from our readers – although, it has to be said, that positive suggestions are more gracefully received!

1.3. Notes for teachers

If you need an example and cannot find an appropriate one here, remember that:

- the behaviour of Normal data is not affected by adding a constant and/or multiplying by a factor;
- the behaviour of two parameter lognormal data is not affected by a multiplicative constant. Three parameter lognormals are more tricky.
- spatial relationships are not affected by a multiplying factor. If the data is Normal, with or without a trend, an additive constant will not affect trend analysis or other spatial relationships.

So, do what we do, use your imagination and experience to adapt these exercises for your own teaching purposes.

Precision: It is also important to remember that all of the answers given here have been computed using proprietary software, spreadsheets and/or hand calculators. All of these have different levels of precision in their makeup. Do not worry if your answer is out by anything up to a couple of percent compared to the one in this book. This is particularly so when you have to square, cube or raise numbers to a larger power.

If you make a real mistake, your answer will be quite different from ours. In most cases, mistakes during calculation lead to huge differences in the answers.

1.4. Cautionary note

Please note that the worked answers in this textbook are – especially in the later chapters – *possible* answers. It is a well established fact that different geostatisticians will produce different conclusions from exactly the same data set.

Both of us have experience of tackling applications where another (or many other) worker(s) have drawn their own conclusions. One of our biggest problems in the field of inference is that there are no *right* answers. There are *wrong* ones. There are *inappropriate* ones. There are *suboptimal* approaches. There are *silly* answers and there are *stupid* answers. There are even *funny* answers.

This volume contains *our* answers. Disagreement is your privilege.

1.5. Data sets

The applications presented within the book are geological (with a bias towards mineability) with some hydrology, soil science and environmental case studies. We have included applications of varied geostatistical techniques in order for the reader to see as many approaches in practice as possible in a volume of this size.

The *potential applications* of geostatistics include any form of measurable spatial data and some which cannot be given a quantitative measure, such as rock type, land use etc.

1.5.1. Coal:

a simulated set of data based on a real coal seam in Southern Africa. Boreholes drilled into the coal seam are measured for: thickness of coal (metres), energy content or ‘calorific value’ of coal (Megajoules per tonne); ash content (%) and sulphur content (%). Three co-ordinates in metres are available for the top of the coal seam where intersected by the drillhole.

In this example we study the following:

- relationships among the economic variables, including: covariances, correlations, predictive relationships between two variables and multi-variable relationships.
- a full study of the calorific value of the coal, including: summary statistics; histograms and probability plots; fitting a Normal distribution model; grade/tonnage curves from Normal model; inverse distance interpolation; calculation and modelling of semi-variograms; cross validation of the semi-variogram model; ordinary kriging.
- a similar study for the sulphur content of the coal samples.

1.5.2. GASA:

this data set is named for the Geostatistical Association of South Africa and was used in an illustration of geostatistical techniques at a meeting in April 1987 in Johannesburg. The sample data are taken from deep boreholes drilled into a typical Witwatersrand type gold reef. The measurements of interest are the grade of the gold in grams per tonne of rock (parts per million) and the thickness of the reef intersection in the borehole (centimetres). The 27 boreholes lie approximately 1 kilometre apart and constitute a typical data set for the planning and design of a new Wits gold mine. The values have been disguised by a factor but are otherwise unaltered. Co-ordinates are in metres.

This case study is all about dealing with highly skewed data using the lognormal distribution. We investigate simple variations on the lognormal, namely the “three parameter” lognormal model. In this model, a constant is added to each sample value to provide data which will fit the standard lognormal.

We discuss:

- fitting the distribution models;
- estimation of the mean and standard deviation;
- confidence intervals for both logarithmic and backtransformed parameters;
- grade/tonnage curves using these models.

For this case study, we have been able to obtain some underground development data at a spacing of around 1.4 metres. This enables us to get a generic semi-variogram which we can use for theoretical calculations of grade/tonnage curves for this project.

We also discuss briefly the possibility of using a single economic variable – the accumulation in centimetre grams per tonne.

1.5.3. Geevor:

Mining in Cornwall dates back to between 1000 and 2000 B.C. when Cornwall is thought to have been visited by metal traders from the eastern Mediterranean. They even named Britain as the “Cassiterides” - “Tin Islands”. Cornwall along with the far west of Devon provided the vast majority of the United Kingdom’s tin and arsenic and most of its copper. Initially the tin was found as alluvial deposits in the gravels of stream beds, but before long some sort of underground working took place. In fact, where the tin lodes outcropped on the cliffs underground mines sprung up as early as the 16th century.

This study concerns sample data from a hydrothermal tin vein deposit in Cornwall, England. From the Online version of the Encyclopedia Britannica:

Main article: vein in geology, ore body that is disseminated within definite boundaries in unwanted rock or minerals (gangue). The term, as used by geologists, is nearly synonymous with the term lode, as used by miners. There are two distinct types: fissure veins and ladder veins.

hydrothermal ore deposit a vein, which forms when a hydrothermal solution flows through an open fissure and deposits its dissolved load.

The mineralisation appears as a continuous vein which is almost vertical. Samples of around 1kg are chipped across the vein, which averages about 24 inches wide. Measurements are grade of tin in pounds of black tin (SnO_2) per ton of rock. The thickness of the vein or ‘lode’ is measured to the nearest inch. Co-ordinates are in feet along section and elevation above an arbitrary base level.

Clark, I., 1979, “Does geostatistics work?”, Proc. 16th APCOM, Thomas J O’Neil, Ed., Society of Mining Engineers of AIME Inc, New York, 213-225.

Clark, I., “Geostatistical Reserve Estimation in Cornish Tin Mining”, Ph.D. Thesis, University of London, 1979

In this case study, we discuss:

- the statistical behaviour of the vein width and the tin value and the complexities introduced by the deposition of the mineral(s).
- whether or not to combine these two variables into a single accumulation value for economic analysis.
- semi-variogram calculation and modelling for both variables, using development samples only.
- whether the nugget effect can be interpreted as sampling error or inherent geological variability (or both!). The assaying technique which is used to produce the measured grades is examined in detail with a special sampling scheme.
- cross validation of the semi-variogram for grades, using stope samples to check back on a model calculated on development samples.

- comparing lognormal kriging with the lognormal shortcut – which kriges on untransformed values.

The interesting conclusion from this work is that lognormal kriging can be used in this case, even though the underlying statistical distribution is not lognormal. A side issue is also resolved when we find that stope samples can be cross validated from a model produced from development drives.

1.5.4. Organics:

Data are Soil Organic Matter values (in grams per kilogram) derived from soil samples taken in a research field at the University of Nebraska West Central Research and Extension Center near North Platte, Nebraska, USA. Data were taken as part of experiments on variable-rate fertilizer technology. Co-ordinates are in metres.

- *Reference: Gotway, C.A. and Hergert, G.W. (1997). "Incorporating Spatial Trends and Anisotropy in Geostatistical Mapping of Soil Properties". Soil Science of America Journal, 61:298-309*

The abstract for the cited paper reads:

The spatial variation in soil parameters often differs with direction. These differences may occur naturally or may be due to management practices. Regardless of their origin, they present a challenge in geostatistical mapping of soil parameters. Recommendations pertaining to the selection of an appropriate geostatistical method based on the current literature are often incomplete or contradictory. The purpose of this investigation was to provide a unified description, comparison, and discussion of different geostatistical methods for handling trend and anisotropy that may be present in measured soil properties. Soil organic matter content of the 0- to 20-cm depth from a field in continuous ridge-tilled corn (*Zea mays* L.) was used to compare five geostatistical methods: ordinary kriging with an isotropic semi-variogram (OKI); ordinary kriging with an anisotropic semivariogram (OKA); ordinary kriging within local neighborhoods (OKN); universal kriging (UK); and median polish kriging (MPK). Organic matter maps produced from the five methods showed similar large-scale features but marked differences in the finer features. A comparison of percentage of total area in each organic matter range among mapping methods also showed strong similarities; however, the proportion of the field assigned to each range differed by as much as 7%. Larger differences would be expected at large sample spacing. Although the five methods produced similar maps, selection of the "best" technique should be based on selection of an associated model that best accounts for and describes the nature of the cause of the variation.

Here we present a slightly different approach to the data.

- Statistical analysis reveals a possible outlier with a very low value compared to all other sampling. We discuss the problem briefly and continue to analyse the data with this in mind.
- Inverse distance is used to sketch maps and to illustrate the question of the appropriate distance function to use for estimation purposes. We see that there appears to be a strong anisotropy in the maps.
- The semi-variogram is calculated. Our interpretation of the semi-variogram is that there seems to be a strong trend in the values.
- Trend surface analysis is carried out and the significance of the trend tested in the usual ways (cf. Chapter 7 in Practical Geostatistics 2000).
- The semi-variogram is recalculated using residuals from the fitted cubic trend surface. This semi-variogram can be modelled.
- Cross validation is undertaken and we find that there are now two significant outliers in the data: one is the extreme low value found previously; the other is an apparently acceptable sample value which just does not mesh with the neighbouring sample values around it.
- Universal kriging is carried out for the full data set and for the data set less the two apparent outliers. Comparisons are illustrated and discussed.

It is interesting to note that, while Universal Kriging was one of the 5 methods used in the original paper there was no discussion of any outliers in the data or of what to do with them.

1.5.5. Scallops:

Scallop data were collected during a 1990 survey cruise off the east coast of North America. Scallop counts were obtained using a dredge. Any scallop smaller than 70 mm was termed a prerecruit. Total catch is the sum of prerecruits and recruits. Measurements included in the data file are:

- National Marine Fisheries Service (NMFS) 4 digit strata designator in which the sample was taken;
- sample number per year ranging from 1 to approximately 450;
- location in terms of latitude and longitude of each sample in the Atlantic Ocean;
- total number of scallops caught at the sample location;
- number of scallops whose shell length is smaller than 70 millimeters;
- number of scallops whose shell length is 70 millimeters or larger.

Reference: Ecker, M.D., and Heltshe, J.F. 1994. "Geostatistical estimates of Scallop Abundance", In, Case Studies in Biometry, Lange et al., editors. Wiley, New York

This is a marvellous data set, provided by colleague Carol Gotway Crawford, now of CDC in Atlanta. This data set can be analysed in more ways than "you can shake a stick at". In this section, you will see:

- summary statistics for "total catch" with discussion of precision problems introduced by large sample values;
- discussion of Normal distribution assumption with calculation of confidence intervals and selectivity abundance/catch limit calculations;
- discussion of a possible lognormal model, fitting of an additive constant and comparison of abundance/catch limit from theoretical model and actual data;
- inverse distance interpolation, a worked example and a full map with discussion of possible anisotropy;
- lognormal geostatistics: semi-variogram calculation, cross validation, backtransforms and lognormal kriging;
- rank uniform transforms: a distribution-free approach to geostatistics to find patterns of abundance.

You could also try – for your own benefit – lognormal geostatistics with trend: with full trend surface analysis on logarithmic values and universal lognormal kriging; Alternatively you could try a two stage process, with indicator kriging for presence/absence of scallops and lognormal kriging for when scallops are present (similar to Velvet Leaf example).

1.5.6. Sunshine Mine

This data set was provided by Pierre Mousset-Jones of Mackay School of Mines in Reno, Nevada. The data is a historical sampling set from development drives in an almost vertical vein. Silver and gold assay are available, as well as the width of the vein.

This study illustrates:

- fitting of lognormal distributions with additive constant;
- estimation of mean value and confidence levels, using large sample theory;
- grade tonnage curves for three parameter lognormal distribution;
- calculation and modelling of logarithmic semi-variograms;
- kriging maps compared with inverse distance squared;
- estimation of the average value over mining panels.

1.5.7. Velvetleaf:

Subsample of the number of velvetleaf weeds counted in 7 meter² area in a field in Nebraska. Data were collected by Gregg Johnson (see 2nd reference), as part of a research program in weed management at the University of Nebraska.

References: Data set taken from: *Gotway, C.A., and Stroup, W.W. 1997. "A generalized linear model approach to spatial data analysis and prediction". Journal of Agricultural, Biological, and Environmental Statistics, 2:157-178.*

Data collected by: *Johnsen, G.A., Mortensen, D.A., and Gotway, C.A. 1996. "Spatial and temporal analysis of weed seedling populations using geostatistics". Weed Science, 44:704-710.*

This case study illustrates the need for a statistical analysis of the data before venturing into a geostatistical analysis. This chapter includes sections where:

- We illustrate the sensitivity of inverse distance interpolation to choice of distance function and search radius;
- calculation and modelling of semi-variogram graphs;
- cross validation of the semi-variogram model.

We find that the cross validation exercise is extremely unsatisfactory – even though the average and standard deviation of the error statistics are close to the ideal [0,1]. We start our analysis again by first considering the statistical behaviour of the values. We find that the data is highly positively skewed but not lognormal (or any variation thereof). Studying a logarithmic probability plot leads us to the conclusion that there are actually three “populations” within the study area:

- quadrats with no weeds (absence of weeds);
- quadrats with up to 40 weeds (colonisation?);
- quadrats with more than 40 weeds (climax vegetation?).

We illustrate the usefulness of indicator transforms for the characterisation and location of the various stages of weed growth, including indicator semi-variogram and kriging.

It becomes apparent from this map that the field under study is largely non-homogeneous and that further investigation into the previous history of the area might provide more realistic estimation techniques.

1.5.8. Wolfcamp:

Measurements of water pressure (potentiometric level) in 85 water wells in the Texas panhandle. This data set was part of a study carried out by the Office for Nuclear Waste Isolation in the mid 1980s on a potential site for a high level nuclear waste repository. The Wolfcamp aquifer underlies the planned repository. One aspect of repository planning is to quantify the risks inherent in a breach of the storage facility. Should radionuclides leak into the local aquifers, the scope and speed of potential contamination has to be assessed. The pressure of fluid within the aquifer was one of several variables used to determine the travel path and speed of travel for escaped radionuclides.

Reference: Harper, W. V., and Furr, J.M., 1986. "Geostatistical analysis of potentiometric data in the Wolfcamp Aquifer of the Palo Duro Basin, Texas", BMI/ONWI-587, April, Office of Nuclear Waste Isolation, Battelle Memorial institute, Columbus, Ohio.

This study of the Wolfcamp data illustrates:

1. descriptive statistics, the Normal distribution and its applicability to this data;
2. nearest neighbour analysis and sketch maps produced by inverse distance methods;
3. semi-variogram calculation, identification of extremely strong trend in the values;
4. trend surface analysis, recalculation of the semi-variogram on residuals from the trend;
5. modelling of the residual semi-variogram and cross validation;
6. kriging 'point' maps and polygonal averages.

1.5.9. Brooms Barn

This is another soil science application, data supplied by Dick Webster – co-author of Webster & Oliver's excellent "Geostatistics for Environmental Scientists". Brooms Barn is an agricultural experimental station in East Anglia (UK) which hosts several fields within its area. The data set includes Potassium (K mg/l), Phosphorus (P mg/l) and pH levels in the soil. Over 400 samples were collected on a regular grid at 40 metres spacing. The data file consists of 434 samples and the following fields for each sample:

- East and North position on the sampling grid – this is not in metres but in grid spacing, i.e. 1 unit of distance equals 40 metres;
- K – potassium value in the soil, mg/l;
- log₁₀ K – logarithms to the base 10 of K values;
- pH – universal measurements for acidity (or lack of) in the soil;
- P – phosphorus levels in the soil, mg/l;
- log₁₀ P – logarithm to the base 10 for P values.

It is obvious from the data file that Dr Webster had identified positive skewness on the K and P values. In our analysis of this data, we use the K values for illustrative purposes. We show that the K values are moderately skewed but that untransformed semi-variograms are adequate once a certain number of apparent outliers have been identified. Thanks to Google EarthTM we can see where these outliers are located and posit reasons for eliminating them from the data set.

We also present a brief analysis of the pH (acidity) values for contrast with the K analysis. This variable shows inhomogeneities which are interpreted as two “populations” within the study area. We show how an indicator transform leads us to the conclusion that three fields have probably been treated with lime or some other ‘basic’ preparation. We show how pH could be mapped if the field boundaries are not known, using a combination of indicator and ordinary kriging. We compare this compound analysis with a map kriged using the indicator semi-variogram to approximate the actual pH semi-variogram.

1.5.10. Brenda Mine

Our final example in this volume was provided by our esteemed colleague Pierre Mousset-Jones. The major feature of this data set is that it is a full three dimensional example. Brenda Mine is a porphyry copper/molybdenum deposit in the southern interior of British Columbia, approximately 22 kilometers west of Peachland in the Central Okanagan, and was closed for economic reasons around 1990. In its time Brenda Mines processed 182 million tonnes of rock resulting in:

Metals Produced	
Copper	278,000 tonnes
Molybdenum	66,000 tonnes
Silver	125 tonnes
Gold	2 tonnes

The files provided by Professor Mousset-Jones were in standard mine planning format, with separate files for borehole collar co-ordinates, survey information and mineral values. These were combined into a single text data file for teaching purposes. The data file `Old_Brenda.dat` contains 1,856 samples. Core sections are between 2 and 56 in length (we assume this to be feet). The longer cores are generally very low grade. The data file supplied at <http://www.kriging.com> comprises the following information:

- X_co-ordinate
- Y_co-ordinate
- Z_co-ordinate
- Cu%
- Mo%
- length_of_core

- From (top of core section sample relative to collar)
- To (bottom of sample relative to borehole collar)

The analysis described here is a fairly straightforward one with comments on traditional geological sampling along section lines.

The Copper (Cu) values are used for our illustration. A very few extreme outliers are identified. Three dimensional semi-variogram graphs are constructed and discussed.

Cross validation is used to determine the necessary order of trend to be used for kriging purposes. Possible outlier values are discussed.

Kriging is applied to produce example maps and accompanying standard errors.

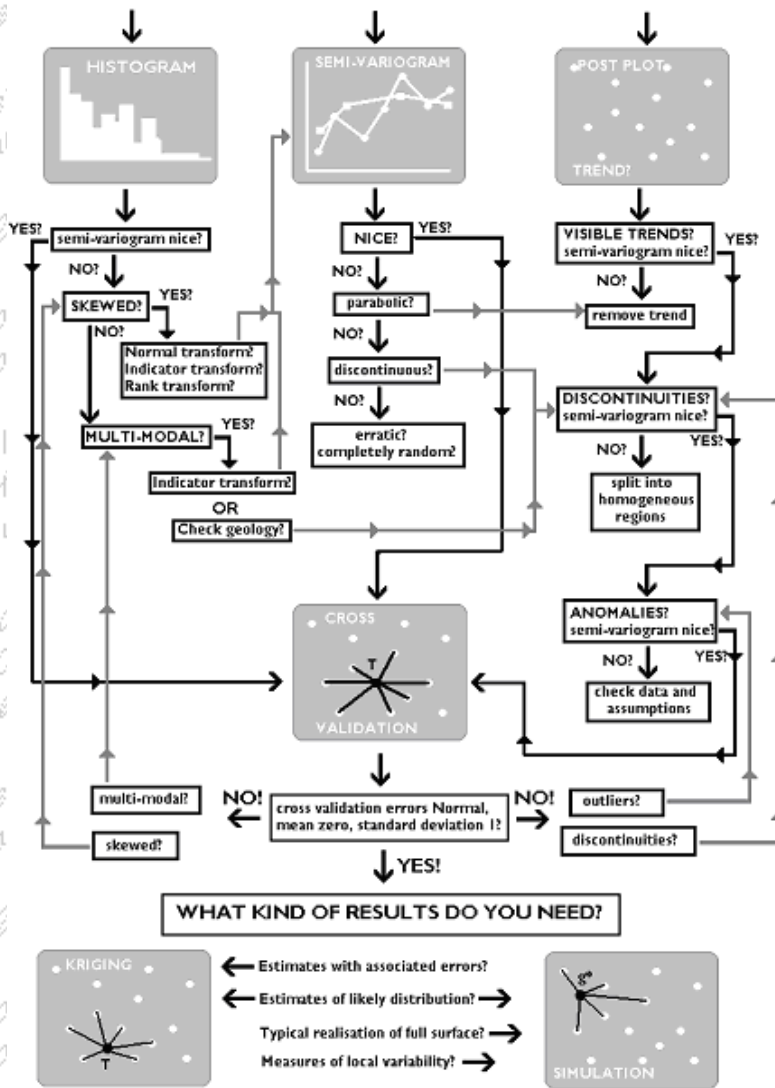
Block kriging is discussed briefly and a small example shown.

Finally, we mention why we did not use trend surface analysis.

Practical Geostatistics 2000 was written as a traditional style text-book, with a bit of theory, full worked examples and exercises at the end of each chapter.

The companion volume **Answers to the Exercises** provide possible answers to the exercises in each chapter of Practical Geostatistics 2000.

On the suggestion of some of our academic colleagues we have produced a third volume which we have entitled **Case Studies**. This volume takes each data set and works through the analyses in the way we might tackle them in the *real world* rather than a text-book.



Ecosse North America, LLC
 Attn: William V. Harper, wharper@otterbein.edu
 Mathematical Sciences, Towers Hall 139
 Otterbein College
 One Grove Street
 Westerville, OH 43081-2006 USA
<http://ecossenorthamerica.com/>

ISBN 0-9703317-4-6



9780970331748

