

# 1. Introduction

## 1.1. Background

This volume is complementary to Practical Geostatistics 2000. However, it can be used on its own if you feel you do not need the basic textbook.

In each of the first 10 Chapters of Practical Geostatistics 2000, there were exercises suggested as valuable to the reader or student. These exercises were based on the data sets supplied with the textbook and on smaller examples drawn from undergraduate and graduate level teaching over the last 28 years. This volume supplies possible answers to those questions.

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 textbook and keep the same conventions in these worked answers.

If you have the CD version of Practical Geostatistics 2000, the larger data sets are included along with software to reproduce the analyses (for the most part). If you do not, the data sets and software can be downloaded from the Web.

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. Data sets

The sort of applications presented within Practical Geostatistics 2000 are mainly geological with some hydrology and environmental case studies. The potential applications include any form of measurable spatial data and some which cannot be given a quantitative measure, such as rock type, land use etc. We have included data sets in the Appendix and on the Web as follows:

- **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.
- **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.

- Samples: this data set is based on a Wits type gold mine some decades into production. The samples are chipped from the face of the reef in a working section of the mine (stope). As the face advances, new chip samples are taken. Values within a stope are traditionally estimated using the sample values from the face. This data is totally fictitious except for the locations of the samples, which are taken from a real Wits type gold mine.
- Copper: a simulation based on a stockpile of mined material in the former Soviet Union. Boreholes have been drilled into the dump. The drill core is cut every 5 metres and assayed for copper and cobalt content in percentage by weight. This is the only three dimensional set of tutorial data. Co-ordinates are in metres.
- Geevor: this is sample data from a hydrothermal tin deposit in Cornwall, England. The mineralisation appears as a continuous vein which is sub-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 ( $\text{SnO}_2$ ) 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.*

- 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 1980's 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.*

- 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*

- **Dioxin:** A truck transporting dioxin contaminated residues dumped an unknown quantity of these wastes onto a farm Road in Missouri. In November, 1983, the U.S. EPA collected samples of the site. In order to reduce the number of samples required, samples were composited along transects. The transects run parallel to the highway, and this direction is designated as the X-direction. The direction perpendicular to the highway is designated as the Y-direction. Data are TCDD concentration (tetrachlorodibenzo-p-dioxin) in micro grams per kilogram ( $\mu\text{g}/\text{kg}$ ). Co-ordinates and transect length are given in feet.

*Reference: Zirschy, J.H., and Harris, D.J. 1986. "Geostatistical analysis of hazardous waste site data". Journal of Environmental Engineering, 112:770-784.*

- **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*

- **Velvetlf:** Subsample of the number of velvetleaf weeds counted in 7 meter<sup>2</sup> 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.*

Two data sets from Practical Geostatistics 1979 are used in the main 2000 textbook, both based on simulations of an actual iron ore deposit. One is from Chapter 2 and one is from later in the 'old' book and is generally referred to as "Page 95".

If you download the software from the Web or the CD of Practical Geostatistics 2000, you will also find some other data sets not directly used in the book. These include:

- **Agromet:** this is a data set which Isobel acquired whilst giving an in-house course to the Institute for Soil, Water and Climate in Pretoria, South Africa. It is simply a square area in the Natal Highlands, giving topographic height on a grid at 400 metre spacing.

- **Rainstat:** another data set courtesy of Soil, Water and Climate. This is some summary statistics on rainfall in South Africa and Namibia over the last 70 years. Lots of variables, lots of non-stationarity and *LOTS* of samples. This file comes with a companion boundary file, outlining the Republic of South Africa.
- **Coalmine:** this is the data set which ‘coal.dat’ is based on. These are the real locations of the samples taken (but not the real values).
- **Pennsylv:** a companion data set to the Wolfcamp. These wells are in the Pennsylvanian aquifer in the same study area as the Wolfcamp data.

### 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.