Review of Best Practices Guide

Clayton V. Deutsch

A PDF and hardcopy of a first draft of the Best Practices Guide have been delivered as part of the meeting. I did not want to release the guide yet. Much remains to be done; however, an early release within the CCG family seemed reasonable since I have promised it for two years now. This short note summarizes much of the introduction and plan for the Guide. The Guide is for CCG researchers and staff from member companies. Readers are expected to have legitimate access to the accompanying software and data files.

1. Introduction

The CCG has evolved over the last 10 years, yet the basic idea remains the same: member organizations provide unrestricted research funding for preferential access to research results. The CCG delivers many results including reports like this one, guidebooks and monographs, source code, compiled programs, generated data, and training images. The Guide to Best Practice in Geostatistics (Guide for short) clarifies the workflow and software tools that we would recommend for practical application. Our goal is to help the members of CCG make sense of the vast number of techniques and tools that are available for their use.

Best practice is the most efficient (least amount of effort) and effective (best results) way of accomplishing a task, based on repeatable procedures that have proven themselves over time for large numbers of people (http://en.wikipedia.org/wiki/Best_practice (March 15, 2010)). Documenting best practice seems like a good idea. Disclosing recommended workflows, procedures implementation decisions and documentation standards would help newcomers find their way through the many various options of geostatistics. Weak areas would be identified and additional research would be triggered.

The Guide is intended to help relative newcomers to geostatistics scope out a geostatistical study, choose specific workflows and tools, make implementation decisions and support the numerous interdependent decisions that must be made during a geostatistical study. There is certainly no unique view of best practice, but the views expressed in this Guide may be useful.

Some people will hide behind the rules and use them as an excuse not to consider the specific problem at hand. Best practices are not general because of many problem and site specific considerations. The rich complexity of geology cannot be captured in a brief guidebook. Moreover, some of best practice is personal preference based on background and experience. Everything else equal, we would choose to use a technique we are familiar with and tools that we have used before. Each practitioner is influenced by past successes and failures. It would be a shame if practitioners follow a workflow without careful consideration of the site specific factors.

Best practice is often to choose methods/techniques that are simple, robust and within your level of expertise. The methods, techniques, and analysis should be appropriate to the goals of the study. We should be able to explain the background and critical assumptions behind every technique and provide documentation for critical input parameters.

2. Plan

The Best Practices Guide is not a book that explains geostatistical principles. There are many books on geostatistics. There are other resources in geostatistics including guidebooks and many papers including more than 500 papers archived in the CCG reports. These resources can be reviewed for more details on specific subjects. The aim of the Guide is to explain how to formulate a solution methodology and how to take different implementation decisions. There are some recurrent themes that will be emphasized repeatedly. Checking should be an integral part of all geostatistical calculations. Checking takes place at all levels: of each specific calculation, of the final model and with new information as it becomes available.

The aim of the Guide is toward the construction of geostatistical models to meet specific project objectives; it is not aimed at the students of geostatistics who construct little toy models to test out some research ideas. The student or researcher can work in isolation with ill defined objectives, few data and limited access to the geological setting of the data. Research is very important, but the Guide is aimed at more realistic modeling where the geostatistician has access to commercial software, IT support, geological and geophysical support, engineering objectives and senior technical staff that would review the project. Working in isolation is encouraged in academic research, but has little place in the best practice of geostatistics.

Real geostatistical modeling takes place in a variety of settings such as petroleum exploration and production, mining and environmental. The objectives of the geostatistical studies and the geological settings can be very different. The Guide is primarily aimed at mining and petroleum applications because of the author's experience and the focus of the CCG. Many of the best practice suggestions are independent of the area of application and this Guide could prove useful in other areas, but that is not the goal.

The core chapters of this Guide and the basic structure:

2. Prerequisites reviews the starting point of a geostatistical study: defining the study objectives and assembling site specific and conceptual data.
3. Unit Operations discusses some of the common analysis and decisions that are taken in geostatistical studies. Stationarity – the choice of how to subset the data and specify location-dependence of statistical parameters – is an important subjects. Issues of scale and grid systems are discussed. Best practice is described for data transformation, declustering and debiasing, calculating uncertainty in statistical parameters and assembling a reliable variogram.
4. Estimation consists of creating a single model for some purpose such as visualizing data or assessing resources. Best practice for this is also discussed for the estimation of local distributions with a consistent multivariate distribution or empirically.
5. Simulation of categorical and continuous variables is central to a modern geostatistical study. Best practice for simulation of one or many variables is presented together with checking the results.
6. Common Challenges include deciding whether or not to model a trend, modeling a trend, integrating large scale data, and the integration of dynamic data.
7. Post Processing of geostatistical models can be as important as how the models are constructed in the first place. Details of best practice related to model operations, displaying and managing uncertainty, scaling the models and ranking/selecting a few models (if necessary) are presented.
8. Some Specific Applications are described in this Chapter. Often, high resolution models are not required and the mapping of trends and uncertainty is adequate for the study objectives. Modeling for heavy oil, permeability, grade control and recoverable reserves are classic problems in geostatistics.
9. Documentation, Disclosure and Classification are important subjects from a regulatory and best practices perspective. This document supplements the regulations and the judgment of the competent person.
10. Appendices describe mostly techniques that are not in best practices and explains why. This is not to say these techniques should not be used, but their application is so narrow that they are not of widespread use for effective and efficient problem solving.

There is little geostatistics possible without software. Data must be explored and summarized, displays must be created, inferences made and large volumes modeled at a reasonably high resolution. There are commercial software alternatives and practitioners should have access to full-featured modeling software. There are many best practices described in this Guide that may not be available in all commercial software, then the practitioner can limit their work or they can search for different software. The tools described herein provide focused tools to perform specific geostatistical tasks. The programs are neither well documented nor easy to use.

3. Workflow

1. **Specify** goals and data
2. **Divide** volume into subsets
3. **Choose** how mean varies
4. **Infer** statistical parameters
5. **Estimate** each variable
6. **Simulate** multiple realizations
7. **Post process** models

Geostatistics is applied in many disciplines with various data types at different scales. The fundamental approach and many of the tools are the same (Chiles and Delfiner, 1999; Deutsch, 2002). A generic workflow for geostatistics could be summarized by seven steps, see summary list to the right. (1) Specify the goals of the study and take inventory of the available measurements and conceptual data. (2) Divide the area/volume of interest into subsets that are relevant for the specific situation. (3) Choose how the mean of each variable depends on location within each chosen subset. (4) Infer all required statistical parameters for creating spatial models of each variable within each subset. (5) Estimate the value of each variable at each unsampled location. (6) Simulate multiple realizations to assess joint uncertainty at different scales. Finally, (7) Post process the statistics, estimated models and simulated realizations to provide decision support information. The detailed implementation of these steps will depend on the purpose of the study.

First: the goals of the study must be specified to determine the work effort required for the study, the variables to be predicted, the scale relevant for evaluation and the specific estimation, simulation and post processing steps. A data inventory must be taken to review all available measured data from wells, seismic and production data. The numerical models should reproduce all of these measured data within the scale and accuracy of the data. Conceptual data must also be assembled including a geological understanding of the spatial distribution and analogue data. The conceptual model expressed in this first step may include schematic pictures and illustrations of the features that should be contained in the final model.

Second: the entire volume being modeled is rarely modeled in one step. There are logical subsets based on geological zones, rock types and facies. If possible, rock that genetically belongs together is kept together. The subdivisions must be large enough to contain sufficient data for reliable statistics, yet small enough to isolate geological features for local accuracy. A hierarchical system is chosen where the site is divided into different zones and, perhaps, into areas where the deposition was controlled by different processes. The surfaces and geometric limits that separate these zones may be modeled by geostatistical tools. Then, the rock types and facies within each zone are modeled at the chosen grid resolution. Finally, continuous rock properties are assigned within each category. The choice of the hierarchical subdivision for modeling has a significant influence on the final numerical geological model.

Third: the mean value of each variable may depend on location within the chosen subset. There are often significant trends in the distribution of rock type and facies proportions within a geological zone. These trends are understood even with few data. Categorical variables almost always have a locally varying mean model. Continuous variables within categories are more likely to have a constant mean model. The results of the second step (subsets of the volume for geostatistical analysis) and third step (modeling the location dependence of the mean) are collectively known as the decision of stationarity.

Fourth: infer all required statistical parameters. The required statistical parameters will depend on the chosen technique that, in turn, depends on the conceptual model chosen for each stationary subset of the domain. Almost always, there will be a need to infer the univariate proportions of each category and the histograms of each continuous variable within each category. These univariate distributions are computed from the data and calculated to be representative of the entire subset. Some measures of spatial variability must also be inferred. In traditional Matheronian geostatistics (Matheron, 1971), variograms are the measures that quantify the spatial variability of each category and rock property. There are other techniques that require size distributions (object based modeling) or training images (multiple point statistics). In presence of sparse data, these statistical parameters are considered uncertain and a number of scenarios are documented. Depending on the purpose of the study, the uncertainty in the parameters is quantified and accounted for in subsequent geostatistical modeling.

Fifth: calculate an estimate of each variable at each unsampled location. These estimates are based on the data and do not involve a random number generator. The estimation is commonly a form of kriging considering indicators, data transformation, cokriging, and/or a locally varying means as required. Whenever possible, the uncertainty is estimated directly with indicators for categorical variables and normal scores in a multivariate Gaussian context for continuous variables. This provides a single best estimate at each unsampled location together with a measure of uncertainty. This is based entirely on the data and decisions taken in the first four steps. The results are useful for resource assessment and checking. Some geostatistical techniques such as object based modeling and multiple point statistics techniques are only used in simulation mode. In these cases, the practitioner would run multiple realizations and summarize the results as facies proportions.

Sixth: multiple realizations of all surfaces, facies and reservoir property variables are simulated to quantify joint uncertainty and to provide a model of heterogeneity suitable for flow simulation. The simulation techniques are often closely linked to the estimation techniques. The estimation results are used for checking the realizations and for a first estimate of the resource/reserve. Uncertainty over a large volume depends on the simultaneous uncertainty at many locations; simulating multiple realizations is the only practical approach to quantify large-scale uncertainty. Also, the details of the geological heterogeneity may have a large influence on recovery and reserve calculations.

Seventh: post-process all of the model results. Sometimes, the statistical parameters from Steps 3 and 4 are useful in themselves; variogram ranges may be used to understand data spacing and expected length scales of geological features. The estimated model provides expected results at unsampled locations and measures of local uncertainty that are useful for data collection and management. Models of different variables must be combined and important response variables calculated. The simulated models provide large scale uncertainty and input to subsequent engineering design. Often, multiple realizations must be ranked to permit a few to be selected for more detailed post processing.

These seven steps provide an overview of the workflow of geostatistics to address specific study objectives. Some of the details will be explained in the Guide; other details are available in textbooks, papers and software user's guides. A geostatistical study is undertaken with data. There are site specific data of different variables at different scale and conceptual data related to the geological formations being modeled.

7. Conclusions

This Guide will be updated annually for a period of time. This early first edition is a little terse and additional explanation and justification is needed in many places. Also, there are valid arguments for alternative best practices.