Some Developments towards Best Practice

C. V. Deutsch

The CCG was born 10 years ago with the mission of teaching and equipping professionals to build high quality geostatistical models that realistically reflect natural heterogeneity and accurately measure our uncertainty. *We are aggressively pursuing this mission. This paper summarizes CCG developments of this year and previous years. Attention is paid to actionable results that should affect current best practice.*

Introduction

Geostatistics is a fascinating field of study. My longstanding enthusiasm for geostatistics stems from the variety of challenges we face. In general, less than one trillionth of a geological site is sampled, and then we must predict multiple variables at many unsampled locations. The physical, chemical and biological processes that led to the present-day spatial distribution are complex, varied and result in interesting spatial features. Although we directly sample a tiny fraction of a site, we may have a variety of secondary data at different scales with varying levels of precision. We are asked for predictions and a quantification of how heterogeneity and uncertainty in geological variables affects engineering performance predictions. This field of study is sufficiently large and challenging to warrant the ongoing attention of many international research groups; it is certainly challenging and interesting enough for the CCG research group.

The focus of CCG is teaching and research. There were ten Master and Doctorate degree defenses since the last report. There are some standard questions in these defenses. A favorite question is "*If you had two minutes in the elevator with the vice president of [type favorite company name here], how would you describe your research?*" Another question in geostatistical cases is "*What changes should be made to [type favorite software here] as a result of your research?*" The responses are varied. Often, students are too enamored with the details of what they have done to see how their research fits in the big picture. The purpose of this paper is to stand back and answer these questions for some of the CCG research.

Much of our research deserves no action on the part of member companies. Some results support current practice. Some ideas do not work out, but are documented anyway. Some results are too immature for full implementation. Some results depend on other ideas to be developed and available. Some of the best research sets the stage for future developments, but deserves no action. This short note focuses on the CCG research that deserves action in modern best practice.

Report Nine: 2006/2007

Some modeling decisions are required to perform geostatistical calculations. The most important decision is the geological domains within which to perform calculations. Some aspects of this decision are straightforward. Major geological zones/rock types/facies are kept separate. Modeling proceeds hierarchically to ensure that "lower-level" modeling is being performed within reasonable geological subdivisions. There are two aspects to stationarity: (1) the choice of a population or group of data to perform statistical calculations, and (2) the location-dependence of statistical parameters. Much CCG research has contributed to this problem.

Locally Varying Anisotropy The locally-varying anisotropy and distance calculations presented in this report are unique (papers 102 and 103). The distance calculation actually *tracks through* the anisotropy field; we do not merely realign variogram parameters locally. For petroleum, this could be implemented automatically with directions obtained from facies and other geological modeling. For mining, this could be implemented with an expert interpretation of directions or distances automatically calculated. The automatic anisotropy detection (paper 112) or non-stationary variogram calculation would supplement this tool (papers 117 and 118).

Gibbs Sampler for MPS Those of you looking to model discrete variables with MPS and training images will benefit from the Gibbs Sampler algorithm (papers 106, 113, 114 and 115). The approach could be implemented in practice and as a tool in software.

Intrinsic Collocated Cokriging Anyone using collocated cokriging with a variant of the Markov model should stop. The biases resulting from variance inflation and lack of histogram reproduction are unacceptable. The intrinsic model of coregionalization (paper 105 and 116) requires exactly the same input parameters, that is, a variogram of the primary variable and a correlation coefficient with the secondary. The data are used differently, but the engine can be adapted with no changes to the user interface. Multiple secondary data could be merged with the super secondary variable concept (paper 109).

Wide Array Declustering A wide array of declustering tools should be used when there are sufficient data (paper 110), which is mostly in mining cases. Simple application of cell declustering, volume of influence or accumulated kriging weights is naïve and not best practice.

Successive or Distance Constrained Kriging Estimation in tabular bounded domains with significant vertical trends or edge effects, that is, unusually high or low values right at the limits should consider a modification to kriging to mitigate the influence of kriging weights to the edge samples. Successive kriging (paper 119) or distance constrained kriging weights (paper 120) calculate optimized weights for kriging-like estimates, but with no artifact large weights to edge samples.

Improved Variogram Calculation Choosing the correct tolerance parameters has always been a problem: until now (papers 121 and 122). The variogram ranges should be adjusted to account for the fact that some tolerance was used and the magnitude of continuity under/over estimated (paper 124). Variable-length compositing is a neat approach to establish a reliable nugget effect and a good first estimate of the variogram range (paper 123).

Uncertainty in Areal Limits Perhaps the first aspect of stationarity is defining the limits within which to assign soil or rock properties. A single deterministic limit provides no assessment of uncertainty for subsequent calculations. The distance function approach (papers 127 and 128) provides an approach to quantify uncertainty in areal limits.

Data Integration In petroleum applications, one is often faced with multiple secondary data at large scale imprecisely related to the facies and petrophysical properties that are being modeled. A variety of machine-learning-based algorithms to integrate the secondary data should be employed (papers 105 and 205). The integrated or calibrated secondary data should be used for high resolution modeling (papers 206 and 207).

Event-Based Facies Modeling In petroleum applications, object-based, surface-based and event-based modeling have become increasingly sophisticated. The notion is to mimic the genetic processes that led to the present day spatial distribution. The tools were launched in early years and have been refined and developed (papers 208, 209, 210).

Recoverable Reserves In mining applications, accounting for lost ore and dilution with relatively sparse exploration data is a challenge. The FGC approach of scaling up high resolution realizations (papers 301 and 302) should be considered for this purpose. The discrete Gaussian model (DGM) (AKA HERCO) has proven is use in global change of support with such sparse exploration data. The generalization of DGM to a wide variety of mineralization styles (paper 303) should be used in many cases.

Block LU Simulation In mining applications, one should use a local LU simulation implemented like a kriging program, but with output more like a simulation program (scaled up values, probabilities and so on). This is implemented for multiple variables, which is important in many cases (papers 304 and 305).

There are a number of other interesting results and deliverables in Report Nine. The robust solver (paper 401) is essential in many geostatistical applications. The tips for coding algorithms as plugins for commercial software will prove useful. The details of factorial kriging (particularly the newer simple kriging implementation) may be just what you need.

Previous Reports

Following are important developments that may already be best practice as a result of older CCG reports.

Bayesian Updating The challenge of uncertainty prediction in presence of sparse primary data and multiple secondary data is well handled by Bayesian Updating. This approach works for Gaussian and non-parametric distributions. We have applied with over 100 primary and secondary variables. The dissociation of the information contained within primary and secondary data, then their merging leads to very useful estimates of local uncertainty. The methodology and many implementation details have been addressed and resolved.

Stepwise Conditional Transformation Multivariate data often show non-Gaussian features such as non-linearity, heteroscedasticity and constraint features. The Stepwise transformation is a multivariate transformation technique expanded upon and developed at CCG. This approach is a candidate in all complex multivariate settings.

Training Image Library Multiple point statistics (MPS) have gained some popularity. A fundamental requirement is training images that reflect the features of the site being characterized. Training images have been constructed and catalogued by CCG for many petroleum and mineral deposit settings. The TI library should be reviewed whenever MPS are to be used.

McMurray Formation Oil Sands Many CCG tools have been developed for the peculiar problems faced by in-situ and mining of oil sands from the McMurray formation in Northern Alberta. Surface modeling, facies modeling, permeability modeling, ranking realizations for SAGD performance, and uncertainty prediction have all been addressed by CCG and are considered best practice in this setting.

Stope Boundaries and Sequencing and ***Dig Limits*** and ***Well Placement*** Optimization tools can be used for decision making in presence of uncertainty. Tools have been developed to establish the precise stope limits and order of mining for underground mining, the optimal dig limits that are achievable whilst minimizing dilution and lost ore, and for the positioning of wells to achieve maximum reservoir performance.

Declustering and Debiasing Virtually all geostatistical modeling requires input univariate statistics. Many tools have been developed at CCG for this purpose. Sometimes the original data can be weighted in clever ways (declustering). Sometimes secondary data must be considered to derive representative proportions or a representative distribution (debiasing). Tools are available for these purposes.

Validation and Checking Checking global uncertainty is very difficult, however, local uncertainty can and should be checked in cross validation mode and with new data. A number of tools and case studies have been developed. Criteria and tools have also been developed for checking simulated realizations.

Some Other Good Ideas

Some research themes have been under development, but have never quite reached prime time. Direct sequential simulation is an example. The challenges of histogram reproduction and the proportional effect have never been comfortably overcome. Documenting the problems with and then fixing P-field simulation has never quite worked out. The problems are clear, but the fix is complicated. A number of CCG tools to account for data of different scale have proven useful in specific settings, but a widespread general approach to multiscale modeling remains elusive. The challenge of non-stationary soft geological boundaries has been addressed; however, full practical implementation remains a challenge because of inference and software.

Consideration of unstructured grids is an active research area. The software to generate, manipulate and flow simulate unstructured grids is not mature enough to permit reasonable standards. Integrating production data is another example. A number of tools have been created, but they are cumbersome to use and require very precise input data.

Developing improved experimental design tools for sensitivity and uncertainty assessment is a research theme. These tools have never fully met the promise of replacing the reasonable practice of Monte Carlo Simulation, vary one at a time sensitivity, and ranking realizations.

The CCG is close to presenting a unified solution to locally varying parameters. Work has been undertaken toward locally varying histograms, correlation matrices, variograms, anisotropy, and other parameters. These have the promise of improving local prediction and resolution in models.

Software

In the beginning, CCG was aimed at using and developing the GSLIB code. That code is still used and developed when it makes sense. Increasingly, code is being developed in C++, C#, Matlab and other platforms. The CCG software catalogue will prove useful for those searching for a specific geostatistical tool. We will continue providing DLLs and plugins. Our focus is to prototype and develop algorithms and solutions and not develop pseudo-commercial code.

There are some very useful programs available from CCG. The spatial bootstrap, debiasing, variogram modeling, indicator and surface based facies modeling, image cleaning, ranking, volumetric uncertainty, volume variance, uncertainty matrix, the UltimateSGSIM and so on.

This paper is not intended to be an overview of the nearly 5000 pages of writing in the other 323 CCG papers; only important and reasonably general developments are mentioned. These warrant serious consideration by all practitioners and software developers.