A Faster, More Accurate Gaussian Simulation

Colin Daly
Schlumberger, Abingdon, UK
Cdaly@slb.com

and

Sandra Quental
Schlumberger, Clamart, France
Squental@slb.com

and

Darcy Novak
Schlumberger, Calgary, Canada
Dnovak@slb.com

Introduction to Gaussian Random Function Simulation

The so-called Gaussian Random Function simulation (GRFS) differs substantially from the Sequential Gaussian simulation (SGS) from GSLIB. GRFS more accurately reproduces distributions. It is typically faster than SGS, with additional efficiencies due to its parallel architecture. GRFS also has an option to run a fast collocated cosimulation – fast enough that it can be used with an interactive correlation coefficient slide bar.

Background

The GRFS works using the well known decomposition which states:

conditional simulation = kriging + unconditional simulation

For the kriging part of the equation, GRFS uses a parallel kriging algorithm introduced in 2008. This kriging algorithm is substantially faster than the old GSLIB-based kriging algorithm, particularly in the case of a lot of well data, and so makes use of the above decomposition practical and beneficial. For example, on a test case with 3 million cells and 500 wells, the new algorithm runs in about 10 seconds compared to about 36 minutes for the old GSLIB based algorithm for identical results. The unconditional simulation term uses a Fast Fourier Transform based algorithm which gives good variogram reproduction for a wide class of variograms (Pardo-Iguzquiza and Chica-Olmo, 1993).

If using the collocated cosimulation option with GRFS, the user will notice that there isn't any systematic bias in the degree of variability of the simulated variable or in the correlation between the simulated variable and the secondary variable. For SGS, it is often found that the variance of the simulated primary variable is systematically different to the desired input variance. In the typical case where the secondary variable is smoother than the primary (often the secondary variable is a smooth seismic data volume), then SGS simulations will generally have higher variability than expected. Furthermore, the calculated correlation between the simulated primary variable and the secondary variable is not equal to the input correlation. This is a problem associated to the sequential nature of SGS. A 'variance reduction factor' was often used to partly remove the bias in an SGS result. This is no longer necessary with GRFS.

Another nice feature implemented for the GRFS is the fast collocated cosimulation. This is based on a new algorithm that extends a well known decomposition from the literature (e.g. Chiles and Delfiner, 1999). This decomposition states that collocated cokriging can be split into a kriging that is done once and a simple Bayesian cokriging update. We have further developed this by coupling it with a correctly chosen unconditional cosimulation of primary and secondary variables. It can be shown that this gives an exact collocated cosimulation. Updating to try a new correlation between primary and secondary variables is quite fast, so this has been implemented on a slider bar. The modeler can interactively see the results when changing the correlation (Figure 1). For the 3 million cell model mentioned above, the updating takes about 0.2 seconds.

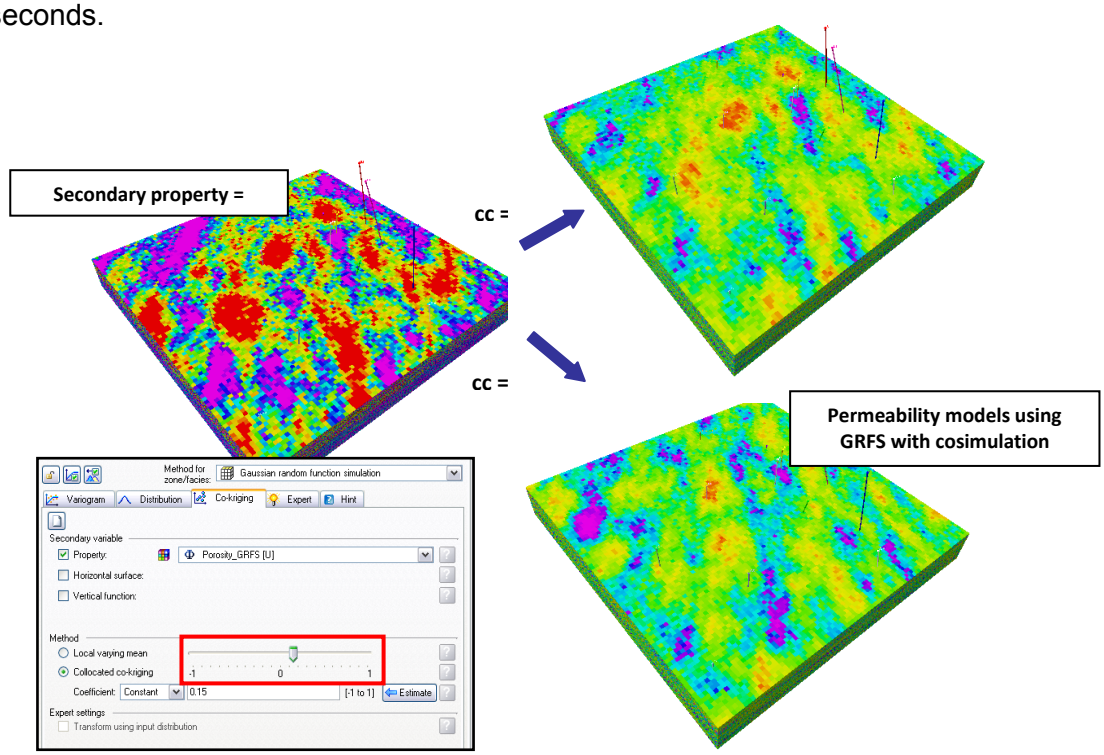

Figure 1: Using a slider bar available for cosimulation with GRFS, it is possible to change the correlation coefficient (cc) and see the results 'on the fly' in visualization windows.

There are several advanced options available in GRFS. The first is a 'layer declustered search' which is used in the kriging component of the Gaussian simulation (thus also present for the kriging method). When active, it ensures that when the kriging algorithm is searching for nearest neighbours of a cell to be kriged, it preferentially searches for neighbours in the current layer and then progressively for neighbours in nearby layers. This overrides the default mechanism which searches for neighbours according to variogram weighted distance. The primary application of this is when the variogram exhibits a long correlation in the vertical direction. In this case, the standard search would tend to find many highly correlated neighbours along vertical, or near vertical wells. This situation, when many of the data used for kriging are highly correlated between themselves, is similar to a situation in standard regression theory called collinearity. A typical solution used for kriging is to perform a declustering of the data and to choose neighbours which are less well-correlated between themselves. The layer declustered search is a simple but often effective method to perform such a declustering in the case of near vertical wells with a long vertical range. The effect is that a better spread of neighbouring data, with less correlation between one another are used for the kriging. This eliminates common artifacts generated by the traditional search methods (Figure 2).

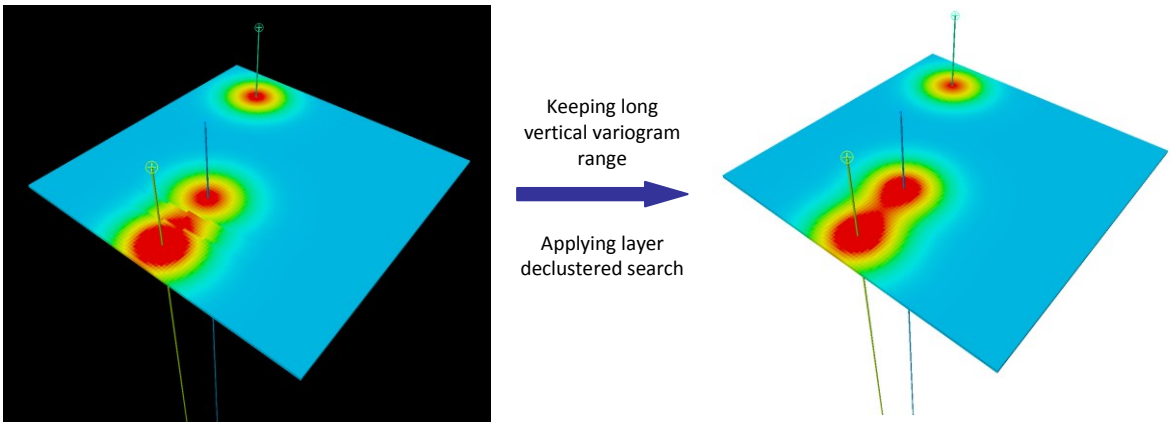

Figure 2: A very simple example showing a kriged porosity model using three wells.
Left: common kriging artifacts are due to a long vertical variogram range.
Right: layer declustered search has been applied, eliminating the artifacts.

Another expert option for kriging and GRFS is the 'approximate search', which provides a search algorithm that can often be substantially faster than the standard search algorithm. It is an approximate nearest neighbor search, thus not as accurate, and it must be used with some care. A typical application is when generating many Gaussian simulations. The user should first test that the fast search is giving results of acceptable quality on a trial realization with a given set of kriging parameters. If it works well for one realization then it will work just as well for all realizations using those same parameters. This option can then be switched on for the time consuming activity of generating many realizations.

Finally, the 'factor of simulation extent' is a variable associated with the unconditional simulation. When the range of the variogram becomes longer than the sides of the model, the Fast Fourier Transform based model will not give a good reproduction of the variogram, due to aliasing. This can be improved by simulating on a larger volume by padding the data, then 'cutting out' the region of interest. The factor of simulation extent can be used in such a case. Typically the value of 1 will be good enough, but if the correlation length becomes long, then it may need to be increased to a length of 2 or 3. Very high values can potentially cause memory problems for the machine. It should be noted that there is little reason for using correlation lengths much longer than the extent of the field as this type of low frequency variability is usually better treated as a trend.

Practical Example – GRFS vs SGS on Pore Volume Distributions

We now look at an example which shows that GRFS does a better job at modeling the uncertainty in the total pore volume of the reservoir than SGS.

In a test project with a regular grid of 90x90x200=1.62 million cells of 100x100x1 meters size, 200 realizations were run on the porosity model, 100 of them using GRFS and the other 100 using SGS. The only parameter varied was the seed - all other parameters were kept the same. The variogram model was spherical with ranges 2000, 2000, 5 in the X, Y and Z directions, respectively. The mean porosity was 0.15 with standard deviation of 0.05.

Our objective is to calculate the pore volume for each realization and then look at the total distributions of such volumes for both GRFS and SGS. The resulting distributions are shown in Figure 3. There is clearly a difference between the GRFS case and SGS case. Which gives the better result? Well, if we knew the expected standard deviation of the distribution we could just check and see.

To help us with this, we remember from basic statistics that if we have n independent points following the same distribution then the variance of the mean of those values is just variance of a single point divided by the number of points ($\sigma^2_v = \sigma^2/N$ where σ^2_v is the variance of the mean reservoir porosity, the variability being from one realization to the next). We have 1.62 million points and we know that the mean of the porosity distribution is 0.15 with standard deviation of 0.05. However, we cannot just choose N=1.62M because not all the data are independent (the resulting simulation is not just a pure nugget effect or white noise simulation so the values are correlated to one another). Roughly speaking we can consider points to be independent of one another when they are separated by a distance equal to the range of the variogram. More accurately, there is a known method in geostatistics for calculating the approximate number of truly independent points. It is called the method of integral range. We won't go into the details here, but the integral range for a spherical variogram is $A = \pi/6\ a^3$ where a is the range of the variogram and the number of equivalent data is then N = V/A where V is the bulk rock volume of the reservoir (Lantuejoul, 1991). In this case we find that N=1600 approximately. We can then use the formula $\sigma^2_v=\sigma^2/N$ to calculate the variability we might expect over the reservoir volume. This gives the standard deviation $\sigma_v=0.00125$. In a normal distribution, the size of the 95% confidence interval is twice the size of the standard deviation. Combining these results we should get 95% of our realizations having a mean porosity for the total reservoir of between 0.15 – 2*(0.00125) and 0.15 + 2*(0.00125), that is in the interval [0.1475, 0.1525]. Since the total rock volume of the reservoir is $(1.62\text{x}10^6)*100*100*1 = 16.2\text{x}10^9\ m^3$ (number of cells multiplied by volume of cell), then the expected range of pore volumes is approximately [$2.389\text{x}10^9$, $2.471\text{x}10^9$]. Looking at the results of Figure 3 we can see that the 100 realizations of the GRFS are consistent with this estimate while the results from SGS show considerably more variability than one would expect from the theory. The following table summarizes the results:

	Lower 95% conf interval	Upper 95% conf interval
Theoretically correct result	2.389	2.471
GRFS – observed result	2.386	2.482
SGS – observed result	2.350	2.502

Table 1: Confidence Intervals for the Total Pore volume variation for the reservoir.
Results are in units of $10^9\ m^3$.

Overall, this shows that the GRFS gives results that are more consistent, in terms of total pore volume modeled, with the information used to develop the model (in this case, the variogram, mean and standard deviations of any well data).

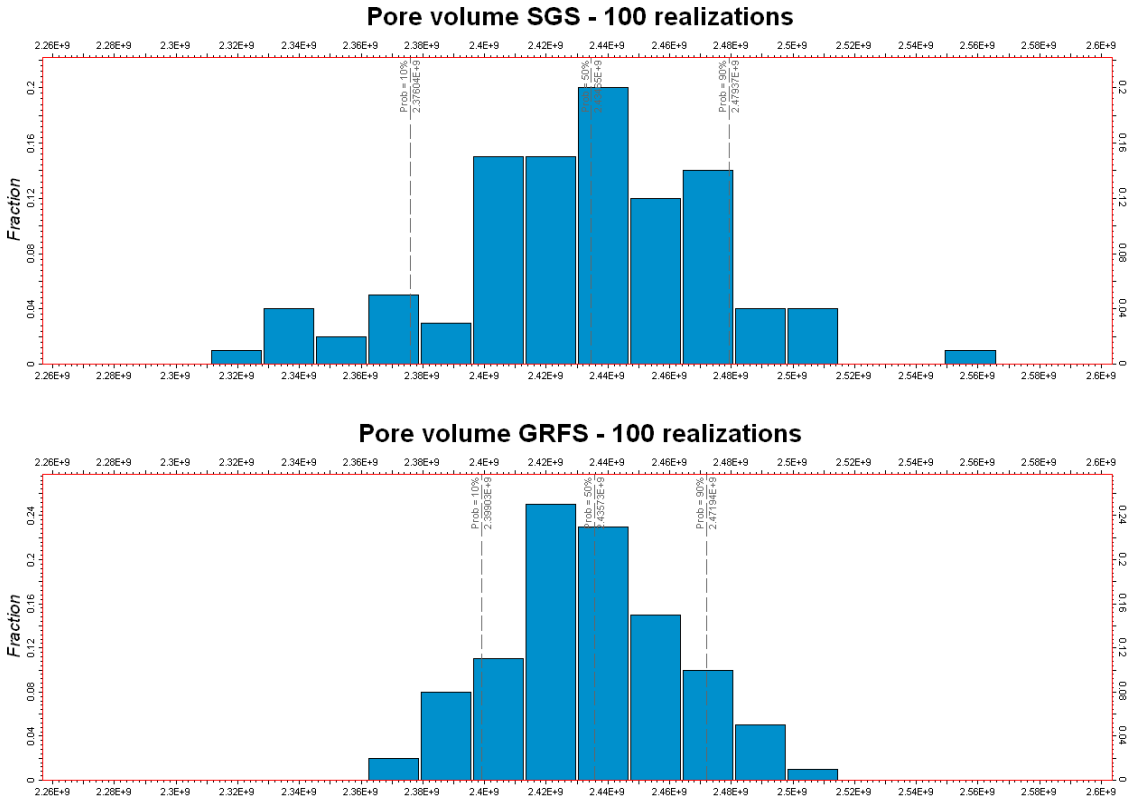

Figure 3: Above: pore volume distribution after running 100 porosity realizations using GRFS. Below: pore volume distribution after running 100 porosity realizations using SGS. The histogram for GRFS is less spread than for SGS, because SGS tends to give higher variance results than expected, based on the input distribution (in this case higher porosity variance, hence higher pore volume variance).

Conclusions

Gaussian Random Function simulation does a better job of modeling the expected variability in distributions. The speed gains of GRFS can be impressive, in part due to its parallel methodology. In addition, the effects of varying the correlation coefficient when cosimulating properties can be seen practically real-time.

Geomodelers are encouraged to use the new GRFS in workflows where SGS is usually applied. In large grids and/or in an uncertainty study context a great time-savings will be gained, as well as a visible improvement in achieving the desired distribution statistics.

References

Pardo-Iguzquiza, E. and Chica-Olmo, M., 1993, The Fourier Integral Method: An Efficient Spectral Method for Simulation of Random Fields, *Mathematical Geology*, Vol.25, No.2

Chiles, J. P. and Delfiner, P., 1999, Modeing Spatial Uncertainty, *Geostatistics*

Lantuejoul, C., 1991, Ergodicity and Integral Range, *Journal of Microscopy*, Vol.161 (3), p. 387-403