Correcting Variogram Reproduction of P-Field Simulation

Julián M. Ortiz (jmo1@ualberta.ca) Department of Civil & Environmental Engineering University of Alberta

Abstract

Probability field simulation is a fast algorithm that is based on separating the task of calculating the conditional distribution from the task of drawing a simulated value.

Its main advantage is the speed with which multiple realizations can be generated. Unlike in sequential methods where each realization requires that the conditional distribution is calculated anew at every location, p-field requires that the conditional distributions are calculated only once. A fast algorithm can then be used to generate the unconditional realizations used as probability fields.

However, two known problems of p-field simulation are the generation of local extrema at conditioning points and the higher correlation that the final realization has near for short distances, that is, near the origin of the variogram. In this paper a solution to the second problem is proposed and several examples are presented to illustrate the correction.

Introduction

Probability field (p-field) simulation dissociates the task of generating the conditional distributions from the actual drawing of simulated values [2, 5]. The basic idea is to generate multiple simulated fields of the probabilities that are used to draw from the cumulative conditional distribution functions (ccdf). These probabilities cannot be random, since this would not preserve the spatial continuity of the variable. Conditional distributions are built considering only the sample data and, unlike in sequential methods, no previously simulated nodes are used to modify these conditional distributions. The probabilities used to draw from nodes that are close to each other should show some spatial correlation.

The algorithm is applied assuming that the correlation used to generate the p-fields corresponds to the correlation of the uniform transform of the conditioning data. However, in reality, the correlation of the probability fields is non-stationary and depends on the spatial configuration of the conditioning data. This incorrect assumption is the cause for the bias in the spatial correlation of the final simulated models, as we will discuss in the next section.

In practice, Gaussian methods are used to generate the unconditional realizations or p-fields. These Gaussian realizations should be transformed into uniform scores, that is, values uniformly distributed in [0,1].

Two approaches are typically used to calculate the conditional distributions from which the values are drawn using the p-fields: indicator kriging or multi-Gaussian kriging.

Indicator kriging consists of estimating the conditional probability at a number of cutoffs, by kriging the indicator transform of the original data at different thresholds. This approach permits integration of secondary and soft information and allows for the different characterization of the continuity for every threshold, thus avoiding the randomness at extreme thresholds implicit with Gaussian methods.

Notwithstanding their drawbacks, Gaussian methods are the most commonly used. Multi-Gaussian kriging consists simply of kriging the normal scores of the original values. The conditional distributions are built by back-transforming the non-standard Gaussian distributions with means and variances given by the kriging estimates and their corresponding estimation variances, respectively.

Covariance Bias in P-Field Simulation

This paper considers the case in which multi-Gaussian kriging is used to define the conditional distributions and a Gaussian method is used to generate the probability fields. Similar results are expected in other cases, although these are not explored in this note.

A simulated probability value at location \mathbf{u} is denoted by $p^{(l)}(\mathbf{u})$ for the l^{th} realization, with l = 1, ..., L. The corresponding simulated value is denoted $x^{(l)}(\mathbf{u})$. In this application, these simulated values can be generated with a matrix method, sequential simulation, or a spectral method. We can write:

$$p^{(l)}(\mathbf{u}) = G(x^{(l)}(\mathbf{u})) \tag{1}$$

The simulated values obtained by p-field simulation, $y^{(l)}(\mathbf{u})$, are calculated using the probability value drawn from the conditional distribution, which in general is not standard Gaussian:

$$y^{(l)}(\mathbf{u}) = G^{-1}(p^{(l)}(\mathbf{u})) \cdot \sigma_{SK}^2(\mathbf{u}) + m_{SK}(\mathbf{u})$$
(2)

where $m_{SK}(\mathbf{u})$ is the multi-Gaussian kriging estimate at \mathbf{u} , given the neighboring information (n), and $\sigma_{SK}^2(\mathbf{u})$ is the corresponding kriging variance.

Notice that we can replace the value of $p^{(l)}(\mathbf{u})$ as given in Equation 1, and rewrite Equation 2 as:

$$y^{(l)}(\mathbf{u}) = G^{-1}(G(x^{(l)}(\mathbf{u}))) \cdot \sigma_{SK}^2(\mathbf{u}) + m_{SK}(\mathbf{u}) = x^{(l)}(\mathbf{u}) \cdot \sigma_{SK}^2(\mathbf{u}) + m_{SK}(\mathbf{u})$$
(3)

The special case of unconditional simulation is interesting. When no data are available to modify the local conditional distributions, these result in the marginal distribution of uncertainty. Thus:

$$\begin{array}{c} m_{SK}(\mathbf{u}) = 0\\ \sigma_{SK}^2(\mathbf{u}) = 1 \end{array} \right\} \forall \mathbf{u} \in A$$

$$(4)$$

where A is the domain simulated.

A p-field is generated with a Gaussian method and the normal score variogram. The covariance between any two simulated Gaussian deviates $x^{(l)}(\mathbf{u})$ and $x^{(l)}(\mathbf{u} + \mathbf{h})$ is $C_Y(\mathbf{h})$, exactly the covariance of the normal scores. In this case, the simulated y values are:

$$y^{(l)}(\mathbf{u}) = x^{(l)}(\mathbf{u})$$

thus, the simulated values $y^{(l)}(\mathbf{u})$ reproduce exactly the target covariance model $C_Y(\mathbf{h})$. In this case, there is no bias in covariance reproduction [3, 4].

Now, considering the case where conditioning data have modified the conditional distributions, then the covariance between two simulated values is:

$$C\{y^{(l)}(\mathbf{u}), y^{(l)}(\mathbf{u}+\mathbf{h})\} = C\{x^{(l)}(\mathbf{u}) \cdot \sigma_{SK}(\mathbf{u}) + m_{SK}(\mathbf{u}), x^{(l)}(\mathbf{u}+\mathbf{h}) \cdot \sigma_{SK}(\mathbf{u}+\mathbf{h}) + m_{SK}(\mathbf{u}+\mathbf{h})\}$$

which can be expressed in terms of expected values as:

$$C\{y^{(l)}(\mathbf{u}), y^{(l)}(\mathbf{u} + \mathbf{h})\} = E\{[x^{(l)}(\mathbf{u}) \cdot \sigma_{SK}(\mathbf{u}) + m_{SK}(\mathbf{u})] \cdot [x^{(l)}(\mathbf{u} + \mathbf{h}) \cdot \sigma_{SK}(\mathbf{u} + \mathbf{h}) + m_{SK}(\mathbf{u} + \mathbf{h})]\} - E\{x^{(l)}(\mathbf{u}) \cdot \sigma_{SK}(\mathbf{u}) + m_{SK}(\mathbf{u})\} \cdot E\{x^{(l)}(\mathbf{u} + \mathbf{h}) \cdot \sigma_{SK}(\mathbf{u} + \mathbf{h}) + m_{SK}(\mathbf{u} + \mathbf{h})\}$$

Rearranging the terms and recalling that $E\{x^{(l)}(\mathbf{u})\} = 0$ and $E\{x^{(l)}(\mathbf{u} + \mathbf{h})\} = 0$, we can write:

$$C\{y^{(l)}(\mathbf{u}), y^{(l)}(\mathbf{u} + \mathbf{h})\} = C\{x^{(l)}(\mathbf{u}), x^{(l)}(\mathbf{u} + \mathbf{h})\} \cdot \sigma_{SK}(\mathbf{u}) \cdot \sigma_{SK}(\mathbf{u} + \mathbf{h}) + m_{SK}(\mathbf{u})) \cdot m_{SK}(\mathbf{u} + \mathbf{h})$$

If a covariance model $C_Y(\mathbf{h})$ is used to generate the p-field (x values), then the resulting covariance is different by a multiplicative factor $\sigma_{SK}(\mathbf{u}) \cdot \sigma_{SK}(\mathbf{u} + \mathbf{h})$, and an additive factor $m_{SK}(\mathbf{u})) \cdot m_{SK}(\mathbf{u} + \mathbf{h})$ [4]. Furthermore, this difference is location-dependent. The probability fields are non-stationary, that is, a different covariance function would be required at every location to yield the correct target covariance in the y simulated values.

Next, we propose a correction that gives the stationary covariance model that yields the correct target covariance in the y values.

Determining the Corrected Variogram Model for P-Field Simulation

Clearly, the p-fields must be generated with a covariance different than $C_Y(\mathbf{h})$. If we consider that a stationary covariance model $C_X(\mathbf{h})$ is going to be used to generate the Gaussian values x used as probability fields, the resultant covariance model for Y, denoted $C_Y^*(\mathbf{h})$, is:

$$C_Y^*(\mathbf{h}) = C_X(\mathbf{h}) \cdot E\{\sigma_{SK}(\mathbf{u}) \cdot \sigma_{SK}(\mathbf{u} + \mathbf{h})\} + E\{m_{SK}(\mathbf{u}) \cdot m_{SK}(\mathbf{u} + \mathbf{h})\}$$
(5)

That is, the covariance model used to generate the p-fields is modified by two factors: the first one corresponding to the mean product of kriging standard deviations for points \mathbf{h} apart, and the second one corresponding to the mean of the products of kriging estimates for points \mathbf{h} apart.

Since we are interested in identifying the term $C_Y^*(\mathbf{h})$ with the target covariance model $C_Y(\mathbf{h})$, we can infer the model $C_X(\mathbf{h})$ that yields this identity:

$$C_X(\mathbf{h}) = \frac{C_Y(\mathbf{h}) - E\{m_{SK}(\mathbf{u}) \cdot m_{SK}(\mathbf{u} + \mathbf{h})\}}{E\{\sigma_{SK}(\mathbf{u}) \cdot \sigma_{SK}(\mathbf{u} + \mathbf{h})\}}$$
(6)

Notice that the correction we propose is global and attempts to correct the reproduction of the variogram when considered over the entire domain A. Due to the non-stationary character of the covariance of the probabilities required for p-field simulation, locally, there will be bias in the covariance. Further research is required to devise an algorithm to locally condition the covariance model. However, this appears as an extremely difficult task, since the covariance model depends *simultaneously* on all conditional data, but in a locationdependent fashion. Imposing non-negative definiteness conditions to the locally modified covariances is not straightforward.

The covariance values $C_X(\mathbf{h})$ must be subtracted from the sill of the corrected covariance to calculate the corrected variogram, that is:

$$\gamma_X(\mathbf{h}) = C_X(\mathbf{0}) - C_X(\mathbf{h}) = \frac{C_Y(\mathbf{0}) - E\{m_{SK}^2(\mathbf{u})\}}{E\{\sigma_{SK}^2(\mathbf{u})\}} - \frac{C_Y(\mathbf{h}) - E\{m_{SK}(\mathbf{u}) \cdot m_{SK}(\mathbf{u} + \mathbf{h})\}}{E\{\sigma_{SK}(\mathbf{u}) \cdot \sigma_{SK}(\mathbf{u} + \mathbf{h})\}}$$
(7)

In conclusion, by generating the p-fields with a variogram model $\gamma_X(\mathbf{h})$, the simulated values obtained from drawing from the conditional distributions with these p-fields have a variogram $\gamma_Y(\mathbf{h})$, which is the target one.

Applications

The following examples have been prepared using a Fortran program that corrects the variogram required for obtaining realizations that reproduce the target variogram model. GAMPFIELD outputs the corrected variogram. This output must be fitted with licit variogram models prior to re-running the generation of the p-fields. The program is documented in the **Appendix**.

All the p-fields are generated using the program LUSIM from GSLIB [1]. Drawing from the conditional distributions given the p-fields has been done with PFSIM, also from GSLIB.

First Example in One Dimension

A first example is presented where a string of 1000 nodes is simulated. Conditioning data are available every 10 nodes in Gaussian units and kriging is performed to obtain the mean and variance of the normal scores at every location.

LUSIM is used to generate 1000 realizations of the 1000 nodes of the string of interest. A variogram model with 0.2 of nugget effect and a spherical model with sill contribution of 0.8 and a range of 10 nodes is used to impose spatial continuity.

PFSIM is used to draw from the conditional distributions using the p-fields generated unconditionally with LUSIM.

The artifact of p-field simulation is shown in **Figure 1**. PFSIM imparts a higher spatial correlation for short distances. Notice that the unconditional realizations to be used as p-fields (from LUSIM) reproduce very well the target variogram, but after drawing from the conditional distributions, the final realizations (from PFSIM) show the bias due to the conditioning information.

A new variogram is calculated to unbias the p-fields. The experimental variogram obtained has to be modelled prior to repeating the simulation with LUSIM. It is shown in **Figure 2**. Notice that the sill is higher than the original target variogram.



Figure 1: Performance of p-field simulation prior to correcting the variogram model to generate the p-fields. Top left: variograms of 100 LUSIM realizations used to generate the p-fields. Top right: average variogram from 1000 realizations with LUSIM. The match is perfect. Bottom left: variograms of 100 PFSIM realizations calculated with the LUSIM realizations as p-fields. Bottom right: average variogram from 1000 realizations with PFSIM. Notice the bias in variogram reproduction.



Figure 2: The calculated variogram to correct the bias generated by p-field simulation (dotted line with bullets). The original model is shown as a solid line (the sill is 1.0). The modelled variogram required to obtain unbiased realizations with PFSIM is shown as a dashed line (the sill is greater than 1.0).

The corrected model consists of a nugget effect of 0.21 plus a spherical model with a sill contribution of 0.90 and a range slightly shorter than the target: 9 nodes.

Finally, LUSIM is re-run with the corrected model and the p-fields are used to draw from the conditional distributions using PFSIM. The p-fields generated with LUSIM reproduce accurately the corrected model and the final models after running PFSIM on these corrected p-fields match the target variogram extremely well (**Figure 3**).

Second Example in One Dimension

An interesting challenge is to see the performance when there is only a change in the local variances. This was done with the same setting used in the previous example, but assigning a value of 0 to all the sample data. Since the work is done in Gaussian units, this means that the entire field is estimated as 0, but the kriging variances from multi-Gaussian kriging change depending on the configuration of the local data.

The average kriging variance is around 0.66, which, because there is no change in the local means, is reflected as a change in the sill of the output variograms from p-field simulation (**Figure 4**).

The corrected variogram is calculated and modelled (**Figure 5**). The same target variogram as in the previous example was used. The corrected model has a nugget effect of 0.16 and a spherical model with sill contribution of 1.36 and range 10 nodes. The total sill is 1.52. Notice that this represents a scaling factor to get the correct target sill:

Target Sill	_	Corrected Variogram Sill
Biased PFSIM Sill	_	Target Sill
1.00		1.52
$\overline{0.66}$	=	1.00

Again, re-running the generation of p-fields with LUSIM and the corrected variogram yielded the target variogram after drawing the simulated values from the conditional distributions (Figure 6).



Figure 3: Performance of p-field simulation after correcting the variogram model to generate the p-fields. Top left: variograms of 100 LUSIM realizations with the corrected variogram. Top right: average variogram from 1000 corrected realizations with LUSIM. The match with the corrected model is perfect. Bottom left: variograms of 100 PFSIM realizations calculated with the LUSIM realizations as p-fields. Bottom right: average variogram from 1000 realizations with PFSIM. The matching is now excellent. Notice that in the top two figures the variogram model is the corrected one (to obtain the p-fields), but in the bottom two figures, the variogram model is the target one.



Figure 4: Performance of p-field simulation prior to correcting the variogram model to generate the p-fields. Top left: variograms of 100 LUSIM realizations used to generate the p-fields. Top right: average variogram from 1000 realizations with LUSIM. The match is perfect. Bottom left: variograms of 100 PFSIM realizations calculated with the LUSIM realizations as p-fields. Bottom right: average variogram from 1000 realizations with PFSIM. Notice the bias in variogram reproduction.



Figure 5: The calculated variogram to correct the bias generated by p-field simulation (dotted line with bullets). The original model is shown as a solid line (the sill is 1.0). The modelled variogram required to obtain unbiased realizations with PFSIM is shown as a dashed line (the sill is greater than 1.0).

First Example in Two Dimensions

A two-dimensional example similar to the previous one, that is, with mean values all equal to 0, is presented. A 25 by 25 grid nodes is simulated using the same variogram model as before.

Running p-field without any corrections yields realizations with biased variograms, as seen in **Figure 7**. The corrected variogram and model is shown in **Figure 8**. In this case, the nugget effect is 0.25, and the spherical structure has a sill contribution of 1.22 with the same range as the target variogram: 10 nodes.

Figure 9 shows again that the corrected variogram provides the correct variogram in the final simulated models.

Second Example in Two Dimensions

Six samples are available for this example. **Figure 10** shows the locations on a 10 by 10 units domain. An anisotropic variogram is used to calculate the kriging estimates and variances of the local distributions for all 100 nodes in the domain. The variogram has a 20% of nugget effect and a spherical structure with the remaining 80% of sill contribution and a Y range of 10 units and a X range of 6 units.

The means and variances of the conditional distributions are shown in Figure 11.

10000 unconditional realizations are generated with LUSIM. Variograms are calculated for each realization and the average is plotted against the model in **Figure 12**. These realizations are used as probability fields to draw from the conditional distributions defined previously with multi-Gaussian kriging, using the program PFSIM. The resulting variogram of the simulated models is biased, as shown in **Figure 13**. The model is corrected to account for the conditioning information and the realizations are generated again with LUSIM (**Figure 14**). The simulated values calculated using the LUSIM realizations as pfields and the conditioning information provided by the Gaussian conditional distributions obtained previously, reproduce the correct target variogram model **Figure 15**.

An interesting additional experiment consists of calculating the block average over the entire domain for every realization and plotting the distribution of this average. This is



Figure 6: Performance of p-field simulation after correcting the variogram model to generate the p-fields. Top left: variograms of 100 LUSIM realizations with the corrected variogram. Top right: average variogram from 1000 corrected realizations with LUSIM. The match with the corrected model is perfect. Bottom left: variograms of 100 PFSIM realizations calculated with the LUSIM realizations as p-fields. Bottom right: average variogram from 1000 realizations with PFSIM. The matching is now excellent. Notice that in the top two figures the variogram model is the corrected one (to obtain the p-fields), but in the bottom two figures, the variogram model is the target one.



Figure 7: Performance of p-field simulation prior to correcting the variogram model to generate the p-fields. Top left: variograms of 100 LUSIM realizations used to generate the p-fields. Top right: average variogram from 1000 realizations with LUSIM. The match is perfect. Bottom left: variograms of 100 PFSIM realizations calculated with the LUSIM realizations as p-fields. Bottom right: average variogram from 1000 realizations with PFSIM. Notice the bias in variogram reproduction.



Figure 8: The calculated variogram to correct the bias generated by p-field simulation (dotted line with bullets). The original model is shown as a solid line (the sill is 1.0). The modelled variogram required to obtain unbiased realizations with PFSIM is shown as a dashed line (the sill is greater than 1.0).

also done by directly simulating the point values with LUSIM (notice that in this case, the realizations are conditional to the data values and are not used as p-fields). LUSIM is a multi-Gaussian simulation method, hence it generates models with maximum entropy. Because of this feature, the average over the entire domain tends to average more quickly towards the mean, generating lower uncertainty in the average than when p-field is used (**Figure 16**). It must be emphasized that p-field simulation is not a multi-Gaussian method. One realization using p-field simulation and one using LUSIM are shown for visual comparison in **Figure 17**.

Conclusions

In this paper, we revisited the p-field algorithm for the case in which the probability fields and the conditional distributions are calculated under the multi-Gaussian assumption. We showed the origin for the bias in the variogram reproduction in p-field simulation and we proposed a procedure to correct the variogram used to generate the unconditional probability fields.

Performance of the proposed approach was illustrated with several examples and the difference of p-field simulation with a conventional Gaussian simulation method was demonstrated with a simple example. P-field simulation does not suffer from the maximum entropy character of Gaussian methods.

References

- C. V. Deutsch and A. G. Journel. GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press, New York, 2nd edition, 1998.
- [2] R. Froidevaux. Probability field simulation. In A. Soares, editor, *Geostatistics Tróia* '92, volume 1, pages 73–84. Kluwer, 1993.



Figure 9: Performance of p-field simulation after correcting the variogram model to generate the p-fields. Top left: variograms of 100 LUSIM realizations with the corrected variogram. Top right: average variogram from 1000 corrected realizations with LUSIM. The match with the corrected model is perfect. Bottom left: variograms of 100 PFSIM realizations calculated with the LUSIM realizations as p-fields. Bottom right: average variogram from 1000 realizations with PFSIM. The matching is now excellent. Notice that in the top two figures the variogram model is the corrected one (to obtain the p-fields), but in the bottom two figures, the variogram model is the target one.



Figure 10: Location map depicting the samples (in Gaussian units).



Figure 11: Map of the means and variances of the conditional distributions.



Figure 12: Target variogram model (dashed lines) and experimental variogram values (dashed lines with bullets).



Figure 13: Biased variogram generated with $\tt PFSIM$ (dashed lines with bullets). The target model is shown as dashed lines.



Figure 14: Corrected experimental variogram values (from GAMPFIELD) are shown as dashed lines with bullets, and model required to generate unbiased realizations (dashed lines).



Figure 15: Unbiased variogram generated with PFSIM (dashed lines with bullets) and the target variogram model (dashed lines).



Figure 16: Histogram of the average value over the entire domain calculated from PFSIM realizations (left) and LUSIM realizations (right). Both methods reproduce the target variogram model.



Figure 17: A realization using PFSIM (left) and one using LUSIM (right).

- [3] A. G. Journel. Probability fields: Another look and a proof. In *Report 8*, Stanford, CA, May 1995. Stanford Center for Reservoir Forecasting.
- [4] M. J. Pyrcz and C. V. Deutsch. Two artifacts of probability field simulation. *Mathe-matical Geology*, 33(7):775–799, 2001.
- [5] R. M. Srivastava. Reservoir characterization with probability field simulation. In SPE Annual Conference and Exhibition, Washington, DC, pages 927–938, Washington, DC, October 1992. Society of Petroleum Engineers. SPE paper number24753.

Appendix: Program GAMPFIELD

A Fortran program has been prepared to calculate the corrected variogram model for probability field simulation, given a target variogram model and the conditional means and variances obtained by multi-Gaussian kriging. The program requires the following input:

- DATAFL: Name of file with multi-Gaussian kriging output.
- COLM,COLV: Column numbers for means and variances of local distributions of uncertainty.
- TMIN, TMAX: Trimming limits.
- NX,XMN,XSIZ,NY,YMN,YSIZ,NZ,ZMN,ZSIZ: Grid definition of the multi-Gaussian kriging input. The number of nodes NX, NY, and NZ, coordinates of the first point in the grid XMN, YMN, and ZMN, and spacing of the three-dimensional grid XSIZ, YSIZ, and ZSIZ are required. These values follow the convention of all the GSLIB programs.
- OUTFL: Name of the output file generated. The file contains the variogram values corrected with the factors calculated from the means and standard deviations of the conditional distributions in the domain. This variogram must be modelled prior to re-simulation. It outputs the corrected variogram in the X, Y, and Z directions defined by the orientation of the grid.
- NLAGC: Specifies the number of lags to correct.
- NST(1),C0(1),IT(i),CC(i),ANG1(i),ANG2(i),ANG3(i),AA(i),AA1,AA2: The normal scores variogram model parameters. As with all programs in GSLIB, NST(1) corresponds to the number of structures, C0(1) is the nugget effect, IT(i) is the variogram type for the structure *i*, CC(i) is its sill contribution, ANG1(i), ANG2(i), and ANG3(i) are the rotation angles for the principal directions of anisotropy, AA(1), AA1, and AA2 are the ranges in the directions of maximum continuity (h_{max}) , minimum continuity (h_{min}) , and perpendicular to both (*vert*).

Parameters for GAMPFIELD **********

START OF PARAMETERS: kt3d.out -file MG kriging values 2 - columns for mean and variance 1 -1.0e21 1.0e21 - trimming limits 100 0.5 1.0 -nx,xmn,xsiz 100 0.5 1.0 -ny,ymn,ysiz 0.5 1.0 1 -nz,zmn,zsiz gampfield.out -file for output 20 - number of lags to correct 1 0.1 - nscores variogram: nst, nugget effect 0.9 0.0 0.0 0.0 it,cc,ang1,ang2,ang3 1 10.0 10.0 10.0 a_hmax, a_hmin, a_vert

Figure 18: Parameter file for program POSTMG.