# MPESIM2: A Multiple-Point Event Simulation Program 

Steve Lyster

The MPS-GS algorithm has been developed at the CCG for the last several years. The MPESIM2 program implements this algorithm using FORTRAN code in the style of GSLIB. This paper outlines the workflow used by MPESIM2, lists the parameters with explanations, and shows examples of other files used for simulation.

## Introduction

Simulation of complex geologic structures is an area of ongoing research. Multiple-point statistics (MPS) is a term used to describe high-order spatial moments or algorithms that use these high-order moments. In the last few years a new algorithm has been developed at the CCG, called MPS-GS, that uses MPS in a Gibbs sampler framework (Lyster et al, 2006, Lyster, 2007, Lyster and Deutsch, 20008). The use of a Gibbs sampler for facies simulation was first proposed by Srivastava, 1992, and the idea of using MPS was briefly touched upon in that paper.

For research purposes, a program was written to implement MPS-GS; this program is called MPESIM2. The name is a reference to the Multiple-Point Events used to calculate conditional probabilities and the SIMulation purpose of the program. This program can be used to generate facies models with statistics derived from a training image. The program uses GSLIB-style input and output formats as well as a parameter file that is specified in the command prompt.

## MPS-GS Algorithm

The MPS-GS algorithm is an iterative method for simulating facies. As put forward elsewhere, the MPSGS algorithm proceeds as follows:

1. Start with a randomly-populated field on the coarsest grid
2. At a random location:
a. Calculate the conditional probability density function (cPDF)
b. Adjust the cPDF to account for secondary information and global facies proportions
c. Apply noise reduction or cleaning and correct cPDF to sum to 1.0
d. Draw a new facies value from the corrected cPDF
e. Move to another unsampled location and repeat Step 2
3. After every location has been visited, check for convergence
a. If there is no convergence yet, repeat Step 2 over all locations
b. If convergence has been achieved, populate the next-coarsest grid
4. After the final grid has been simulated write out the results and start the next realization

The statistics used to calculate the cPDF in Step 2a are derived from a training image (TI); it is this part of the algorithm that uses MPS. The MPESIM2 program implements the MPS-GS algorithm using the following equation to estimate the conditional probability of each facies, $k$ :

$$
\begin{equation*}
P^{*}(k)=\sum_{i, \alpha} \lambda_{i, \alpha}^{k}\left[I\left(E_{i}^{\alpha}\right)-P\left(E_{i}^{\alpha}\right)\right]+P(k) \tag{1}
\end{equation*}
$$

Equation 1 is a linear estimate using non-linear indicators; $E_{i}^{a}$ represents a discrete multiple-point event, or MPE. Each event $i=1, \ldots, M$ is defined by $N$ points in space that are offset from the central point to be estimated; within each event there are $a=1, \ldots, K^{N}$ classes that define all of the possible combination of facies within the event. The optimal weights assigned to each indicator are found by solving the system of equations:

$$
\begin{equation*}
\sum_{j, \beta} \lambda_{j, \beta}^{k} \cdot \operatorname{Cov}\left\{E_{i}^{\alpha}, E_{j}^{\beta}\right\}=\operatorname{Cov}\left\{E_{i}^{\alpha}, k\right\} \tag{2}
\end{equation*}
$$

where all of the covariances between MPEs are derived from the TI. This system of equations is very large and is solved by using the LAPACK (Linear Algebra PACKage) suite of subroutines (Anderson et al, 1999). The solution to the equation may be found as:

$$
\left.[\lambda]=\left[\begin{array}{l} 
 \tag{3}\\
\end{array}\right] H S\right]^{-1}[R H S]
$$

where the LHS matrix is the covariances between the MPEs and the RHS is the covariance between the MPEs and the location being estimated. The LHS matrix is singular and very large; the maximum size is $M K^{N}$ which can easily be in the thousands. Those MPEs with no occurrences in the TI, or very few, are ignored. Even so, with more than two or three facies the size of the MPEs is limited by this system of equations.
Solving the singular system does not lead to a unique solution. The Moore-Penrose pseudoinverse (Anderson et al, 1999) is used, as this is the solution that leads to the minimum norm and therefore avoids any extreme weights. The pseudoinverse is found by decomposing the LHS matrix into its eigenvalues and eigenvectors:

$$
\left[\begin{array}{l}
\text { LHS }
\end{array}\right]=\left[\begin{array}{ll} 
& V
\end{array}\right]\left[\begin{array}{ll} 
&  \tag{4}\\
& \\
\end{array}\right]^{T}
$$

where $V$ is the orthogonal matrix of eigenvectors and the diagonal of $\omega$ is the corresponding eigenvectors (several of which are zero due to the singularity of the matrix). The pseudoinverse is found by taking the reciprocal of the eigenvalues, leaving those that are zero as zero. The optimal weights are then:

$$
\lambda]=\left[\begin{array}{ll} 
& V
\end{array}\right]\left[\begin{array}{ll} 
&  \tag{5}\\
& \\
& \\
&
\end{array}\right]
$$

Because the optimal weights do not depend on the MPE indicator values, the optimal weights only need to be found once and then may be stored and used as necessary. The estimator in Equation 1 may be broken into two parts, beginning with the "base" estimates:

$$
\begin{equation*}
P^{+}(k)=-\sum_{i, \alpha} \lambda_{i, \alpha}^{k} \cdot P\left(E_{i}^{\alpha}\right)+P(k) \tag{6}
\end{equation*}
$$

Equation 6 is the estimates for each facies assuming that all of the MPE indicators are zero (which will not be true, as one indicator per event must equal one). The "base" estimates are independent of the indicators and can be calculated for a given MPS template and TI without requiring any data input. Equation 6 can then be easily updated to account for local information:

$$
\begin{equation*}
P^{*}(k)=P^{+}(k)+\sum_{i, \alpha} \lambda_{i, \alpha}^{k} \cdot I\left(E_{i}^{\alpha}\right) \tag{7}
\end{equation*}
$$

There are $M$ indicators with a value of one and the rest are zero as per the exhaustive and exclusive properties of indicators. Equation 7 then only requires the addition of the base estimates and $M$ weights. This addition is very fast; much faster than solving a system of equations or looking up values in a table or search tree.
The MPS template that defines the spatial offsets of the MPEs is calculated automatically by the MPESIM2 program. The "optimal" template for a given TI is that which uses the lowest-entropy MPEs. The spatial entropy between an $N$-point event and the central point of the template is calculated as:

$$
\begin{equation*}
H_{i}=-\sum_{\alpha=1}^{K^{N}} \sum_{k=1}^{K} P\left(E_{i}^{\alpha} \cap k\right) \cdot \ln \left[P\left(E_{i}^{\alpha} \cap k\right)\right] \tag{8}
\end{equation*}
$$

Minimizing this entropy should lead to the MPEs that convey the most information about the central location that contains the facies of interest, $k$.
Calculation of the optimal template and estimation weights is done automatically by the MPESIM2 program. However, the program stops between each step to allow the user to inspect the output; template offsets should be looked at and the optimal weights checked for anomalous results before proceeding with simulation. The template and weights are stored in files that will be discussed later; once these files have been created the TI is no longer needed. Figure 1 shows a flowchart of this part of the workflow as implemented in MPESIM2.

Figure 2 shows a flowchart of the MPESIM2 program for the data input/preparation steps; this is after the statistics have been calculated but before simulation commences. Both hard data (wells, drillholes, etc) and soft data (probability cube) may be used. Hard data are frozen in place and not visited in the random path; soft data are honoured by using a multiple servosystem approach as shown in Equation 9:

$$
\begin{equation*}
P^{\prime}(k)=P^{*}(k)+\mu \cdot\left[P^{L O C}(k)-P^{S I M}\left(k, B_{k}\right)\right] \tag{9}
\end{equation*}
$$

A servosystem is a typical approach to use for matching facies proportions in simulation (Liu, 2006). The value of $\mu$ is a control parameter and is usually set to a default value of 1.0. Using multiple servosystems, the global proportions of facies within the local bins, $B_{k}$, are matched. This reproduces the local proportions without giving them undue weight. The local bins are calculated as:

$$
\begin{equation*}
B_{k}=\operatorname{int}\left(\frac{P^{L O C}(k)-P^{M I N}(k)}{P^{M A X}(k)-P^{M I N}(k)+\varepsilon} \cdot N_{B}\right)+1 \tag{10}
\end{equation*}
$$

After the servosystem update in Equation 9 has been applied, the updated probabilities of the facies are modified to reduce the overall noise in the realization by adding a connectivity correction:

$$
\begin{equation*}
P^{\prime \prime}(k)=P^{\prime}(k)+\eta \cdot\left[P\left(C_{k}\right)-E\left\{P\left(C_{k}\right) \mid P\left(C_{k}\right)>0\right\}\right] \tag{11}
\end{equation*}
$$

In Equation $11 C_{k}$ is the connectivity of facies $k$; this is the number of adjacent cells that have the value $k$. $P\left(C_{k}\right)$ is the probability of the connectivity values and is calculated from the TI. This correction causes the short-range connectivity to be matched to that of the TI; very continuous facies have higher probabilities of large $C_{k}$ values. The possible values of $C_{k}$ range from 0 (no connectivity) to 7 (all six adjacent cells in 3D are facies $k$ ). The $\eta$ parameter is a controlling value, usually set at about 0.1 .

Figure 3 shows a flowchart of the steps followed by the MPESIM2 program during simulation. When moving from one grid to the next, all cells in the current grid must be populated. The MPESIM2 program does this by assigning the facies values randomly (for the first grid only) or by assigning the nearest value from the previous grid (for all finer multiple grids).

## Parameter File

Several text files are used by the MPSEIM2 program. The first is a GSLIB-style parameter file (Deutsch and Journel, 1998). If no parameter file is found on the program execution then a default parameter file is created; Figure 4 shows a default parameter file. Line 1 is the name of the file containing the MPS; if this file exists it will be read in, if not it will be created (see Figure 1). Lines 2 through 4 define the MPS to use in simulation; the number of grids to use $(G)$, the number of MPEs for the template $(M)$, and the points per event ( $N$ ). The $M$ and $N$ values must agree with the MPS file, and $G$ must be no greater than that contained in the file (the simulation cannot use more grids than the MPS have been calculated for). Line 5 defines the minimum frequency of a MPE class to consider it in the system of equations (Equation 2). A higher value on Line 5 will reduce the size of the system to solve but will also ignore more MPEs.

Lines 6 to 8 define the TI file. Line 6 is the filename; Line 7 is the column containing the facies information; and Line 8 is the size of the TI field (X, Y, and Z). No TI file is needed if the MPS file (Line 1) already exists.

Lines 9 through 11 defines the template to calculate the MPS within, if necessary. Line 9 is the name of the file containing the template offsets; if this file exists it will be read, otherwise it will be created. Line 10 is the columns containing the $\mathrm{X}, \mathrm{Y}$, and Z offsets for each point in the template. Line 11 is the maximum offsets in the $\mathrm{X}, \mathrm{Y}$, and Z directions if a template is to be created. These lines are irrelevant is the MPS file already exists, but still must exist in the parameter file. If the simulation is 2 D , the Z offset column may be set to 0 .

Lines 12 through 14 define the indicators for the facies. Line 12 specifies the number of facies, $K$; Line 13 provides the integer codes defining the facies and must have $K$ values; and Line 14 defines the global target proportions of the facies and must also have $K$ values. If there are fewer than $K$ parameters on Lines 13 or 14 the program may crash.

Lines 15 and 16 specify the hard conditioning data file. Line 15 is the file name, and Line 16 identifies the columns for the $\mathrm{X}, \mathrm{Y}, \mathrm{Z}$, and facies information. If the data set is 2 D the Z column may be set to 0 . If the conditioning data file is not found then the realizations will be unconditional.
Lines 17 through 23 define the output. Line 17 is the name of the file to write the results to; this file will be in standard GSLIB format (Deutsch and Journel, 1998) with one column. Line 18 is the number of realizations to generate. Lines 19 to 21 are the grid definitions in the standard GSLIB format. Line 22 specifies the stopping criteria for the realizations, with a stopping number (typically left as 1 except for research); a minimum proportion of cell values to change in each loop (if less than, say, $0.1 \%$ of all values change the image is assumed to have converged); and the maximum number of loops to allow before stopping. Line 23 is the debugging output file; this file will provide a record of the output to the screen during simulation and will contain the parameters used, the number of changes in each loop, and the univariate proportions in each loop.

Lines 24 to 26 define the local proportions (or probability cube) information. Line 24 is the file name (if the file does not exist no local proportions will be used); Line 25 provides the $K$ columns for the local data; and Line 26 specifies the discretization of the local PDFs or number of bins $B_{k}$ in Equations 9 and 10. If local PDF data exists then the global proportions are not used.

Line 27 specifies the servosystem and connectivity correction control parameters. Line 28 is the random number seed value.

## Template and MPS Files

Using the TI and the defined parameters, the MPESIM2 program will calculate the optimal MPS template and weights used in Equation 1. These two calculations are stored in files for easy inspection by the user and for use later without having to be recalculated.

Figure 5 shows an example of a template file. For this template, a simple 2D TI was used. Line 1 defines the parameters for the template: $G, M$, and $N$. Line 1 is also the GSLIB-style header line for the file. Line 2 is the number of columns in the file; if the template is created by MPESIM2 this is always 3 . The Lines numbered 3 in Figure 5 are the header lines and there is one for each column in the file; in this case, one
each for the $\mathrm{X}, \mathrm{Y}$, and Z offsets. The Lines numbered as 4 defined the offsets in the template from the central location of $(0,0,0)$. There are $G M N$ lines in the file, with $N$ cycling the fastest, then $M$, then $G$, so the first $N$ lines define the first MPE in the first grid and the $(M N+1)$ to $(M N+N)$ lines define the first MPE in the second grid. In the example template file two grids were used with four, four-point MPEs in each grid for a template size of 16 points and 32 lines in the file.

Figure 6 shows an example of an MPS file. The dots "..." represent lines that have been removed to shorten the file to a single page for this example. Line 1 is the header line; note the MPS file is no longer GSLIB-standard, but is unique to the MPESIM2 program. Line 2 contains the parameters for the MPS in the file, with four numbers: $G, M, N$, and $K$ (these letters are shown in order on Line 1). These parameters must agree with the parameter file used, though grids may be dropped from MPS calculation to simulation.

The Lines labeled as 3 are the headers for each grid and there are $G$ of these lines. After the grid header there is a Line 4 (again one for each grid), which defines the global proportions for this grid, $P_{k}, k=1, \ldots, K$. Note that these numbers vary slightly from grid to grid due to the range of the search for each grid; these very small fluctuations can have a significant impact on the multiple-point covariance matrix. After the Line 4 there is a group of Lines, labeled 5 here, for each grid. There are seven lines and they define the connectivity values $P\left(C_{k}\right) k=1, \ldots, K$; each line defines one connectivity value from $C_{k}=0$ to $C_{k}=6$. These are the connectivity values as used in Equation 11.

After the grid header lines, there is the header for the first MPE on that grid. Line 6 is the header for each MPE, and there are $M$ of these lines for each grid. After Line 6 is the group of Lines 7 that specify the point offsets for the MPE. There are $N$ of these lines for each event. The set of Lines 8 are the statistics calculated for the indicators in the MPE. Each line specifies the class $\alpha$, the global proportion of that class, $P\left(E_{i}^{a}\right)$, and weight assigned to the indicator of that class for each facies, $\lambda_{i, a}^{k}, k=1, \ldots, K$. There are a maximum of $K^{N}$ of these lines for each event, one for every possible class. Classes that are below the cutoff proportion in the parameter file are not included.

The Lines 6 through 8 are repeated $M$ times for each grid. At the end of each grid are the weights assigned to univariate indicators. MPESIM2 incorporates single-point indicators into the estimation; this is shown in Equation 12, which is Equation 1 with an additional summation.

$$
\begin{equation*}
P^{*}(k)=\sum_{i, \alpha} \lambda_{i, \alpha}^{k} \cdot\left[I\left(E_{i}^{\alpha}\right)-P\left(E_{i}^{\alpha}\right)\right]+\sum_{u, k^{\prime}} \chi_{u, k^{\prime}}^{k} \cdot\left[I\left(k^{\prime} ; u\right)-P\left(k^{\prime}\right)\right]+P(k) \tag{12}
\end{equation*}
$$

These univariate indicators are redundant with the MPEs and do not change the estimates if all of the MPEs are informed; however, for large events or $4+$ facies many MPEs are not found in the TI and using univariate indicators speeds up convergence of the algorithm. Line 9 is a header for the univariate indicators; each of the Lines labeled 10 contain the $\mathrm{X}, \mathrm{Y}$, and Z offset of the indicator, the facies value at that offset, and the $K$ weights associated with that indicator. There are $M N K$ lines in this part of the MPS file.

After the univariate indicators portion of the file, Lines 3 through 10 repeat again for each of the remaining grids. Line 11 defines the end of the MPS file.

## FORTRAN Files

There are 81 files of FORTRAN code needed to compile the MPESIM2 program. The majority of these are from the LAPACK suite of subroutines (Anderson et al, 1999). A full list of these files is shown in Figure 7. There is the main file, mpesim2. f; a group of GSLIB and similarly-styled files; and a variety of LAPACK and BLAS (Basic Linear Algebra Subprograms) subroutines.

## Comments

The results of the MPESIM2 program are robust with respect to the parameters; typically good results can be obtained by using 4-5 grids, and 8-12 four-point MPEs or 4-8 eight-point MPEs (Lines 2-4). The minimum MP frequency (Line 5) should be adjusted based on the size of the TI and number of points and facies to be used in the simulation; typically a frequency of $10^{-5}$ (one in 100,000 ) is suitable for large TIs,
though this may be adjusted to reduce the size of the matrix to invert. A servosystem parameter of 1.0 and connectivity factor of 0.1 normally gives acceptable results (Line 27). A stopping number of 1 , threshold of 0.01 and maximum number of loops of 30 is also a good starting point for simulation (Line 22).

Currently the maximum size of the MPEs is limited by the memory requirements, as the entire LHS matrix of Equation 2 is initialized even for the classes that are not used. This simplifies the programming but drastically increases the RAM needed to run the program for more than two facies. The program runs best using a TI that is several times the size of the largest structure to be reproduced. For long-range channel connectivity this means having a TI that is several times the size of the simulation grid. Using TIs with more than 3-4 facies does not always give good results; in these cases grouping facies and simulating hierarchically would be necessary.

The MPESIM2 program was created primarily as research code for implementing and testing the MPS-GS algorithm. As such, there are a number of improvements that could be made in the code. The program is fast enough to be used for large geomodels but could be optimized further. Solving the system of equations uses a set of very efficient subroutines, but the calculation of covariances from the TI could be improved.

## References

Anderson, E., Bai, Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J., Du Croz, J., Greenbaum, A., Hammarling, S., McKenney, A., Sorensen, D. (1999) LAPACK User's Guide, $3^{\text {rd }}$ Ed. Society for Industrial and Applied Mathematics, Philadelphia, PA, 429 p.
Deutsch, C.V. and Journel, A.G. (1998) GSLIB: Geostatistical Software Library and User's Guide, $2^{\text {nd }}$ Ed. Oxford University Press, New York, 369 p.
Liu, Y. (2006) Using the Snesim Program for Multiple-Point Statistical Simulation. Computers \& Geosciences, Vol. 32, No. 10, Dec. 2006, pp 1544-1563.

Lyster, S. (2007) A Second Generation Gibbs Sampler for Simulation with Multiple-Point Statistics. Centre for Computational Geostatistics, No. 9, 26 p.
Lyster, S., and Deutsch, C.V. (2006) Short Note: An Entropy-Based Approach to Establish MPS Templates. Centre for Computational Geostatistics, No. 8, 10 p.
Lyster, S. and Deutsch, C.V. (2008) MPS Simulation in a Gibbs Sampler Algorithm. $8^{\text {th }}$ International Geostatistics Congress, 10 p.

Lyster, S., Deutsch, C.V., and Dose, T. (2006) A New MPS Simulation Algorithm Based on Gibbs Sampling. Centre for Computational Geostatistics, No. 8, 16 p.
Srivastava, M. (1992) Iterative Methods for Spatial Simulation. Stanford Center for Reservoir Forecasting, No. 5, 24 p.


Figure 1: Workflow of the MPESIM2 program for statistical calculation.


Figure 2: Workflow of the MPESIM2 program for data input and preparation to simulate.


Figure 3: Workflow of the MPESIM2 program for simulation.

| 1 | $\qquad$ | -File for MPS <br> -Number of grids to use <br> -Number of MP events to consider <br> -Number of points per event |
| :---: | :---: | :---: |
| 2 | 1.0e-6 | -Minimum MP frequency to consider |
| 3 | tifile.dat | -Training Image file |
| 4 | 1 | -Column for TI data |
| 5 | 256256128 | -Size of TI field |
| 6 | template.temp | -File containing point offsets |
| 7 | 123 | -Columns for $\mathrm{X}, \mathrm{Y}, \mathrm{Z}$ offsets |
| 8 | $\begin{array}{lll}5 & 5 & 3\end{array}$ | -Maximum point offsets |
| 9 | 5 | -Number of codes/facies |
| 10 | $\begin{array}{llllll}0 & 1 & 2 & 4 & 5\end{array}$ | -Indicators for codes/facies |
| 11 | $\begin{array}{lllll}0.4 & 0.2 & 0.2 & 0.1 & 0.1\end{array}$ | -Global pdf values |
| 12 | data.dat | -Conditioning data file |
| 13 | $1 \begin{array}{llll}1 & 2 & 3\end{array}$ | -X,Y,Z, Data columns |
| 14 | mpesim2.out | -File for output |
| 15 | 1 | -Number of realizations |
| 16 | $\begin{array}{lll}256 & 0.5 & 1.0\end{array}$ | -nx, xmin, xsiz |
| 17 | $\begin{array}{llll}256 & 0.5 & 1.0\end{array}$ | -ny, ymin, ysiz |
| 18 | $128 \quad 0.51 .0$ | -nz,zmin, zsiz |
| 19 | 10.0530 | -Stopping number, threshold, max loops |
| 20 | mpesim2.dbg | -File for debugging |
| 21 | pdf.dat | -Local pdf file |
| 22 | $\begin{array}{lllll}1 & 2 & 3 & 4 & 5\end{array}$ | -Columns for all indicators |
| 23 | 4 | -Discretization of local pdf |
| 24 | 1.00 .1 | -Servosystem, connectivity factors |
| 25 | 6744889 | -Random number seed value |
| 26 |  |  |
| 27 |  |  |
| 28 |  |  |

Figure 4: Default parameter file for the MPESIM2 program.


Figure 5: Example template file for the MPESIM2 program.


Figure 6: Example MPS file for the MPESIM2 program.

| Main File |
| :--- | :--- | :--- | :--- | :--- |
| mpesim2.f |

Figure 7: List of FORTRAN files required to compile the MPESIM2 program.

