Gaussian Simulation for Porosity Modeling

- Petrophysical Property Simulation
- Gaussian Simulation
- Sequential Gaussian Simulation
- More Comments on the Steps in Sequential Simulation
- SGSIM Program

Petrophysical Property Modeling: Prerequisites

- Work within “homogeneous” lithofacies/rock-type classification $\mapsto$ may require a first step to model lithology
- Sequence stratigraphic framework $\mapsto$ Z_{rel} vertical coordinate space
- Clean data: positioned correctly, manageable outliers, grid spacing is appropriate
- Need to understand special features and “special” data:
  - trends
  - production data
  - seismic data
- Considerations for areal grid size:
  - practical limit to the number of cells
  - need to have sufficient resolution so that the upscaling is meaningful
  - this resolution is required even when the wells are widely spaced (simulation algorithms fill in the heterogeneity)
- Work with “grid nodes”. We assign a property for the entire cell knowing that there are “sub-cell” features

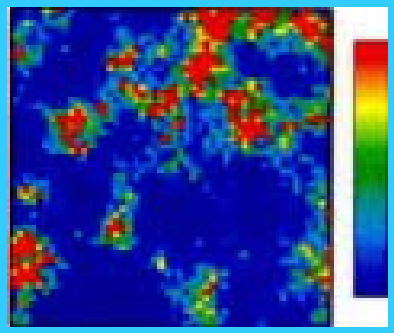
Different Petrophysical Property Simulation Algorithms

- **Matrix Approach** (LU Decomposition): not used because of the size of the problem (an *N* x *N* matrix must be solved where *N* could be in the millions) `lusim`
- **Turning Bands**: simulate the variable on 1-D lines and combine in 3-D. Not used because of artifacts and difficulty in conditioning to local data `tb3d`
- **Fractals**: not extensively used because self-similarity is only valid for a limited range of scales and difficulty in conditioning to local data (see Tom Hewett)
- **Annealing**: becoming popular $\mapsto$ recommended for permeability (talk about later)
- **Sequential Simulation**: widely used and recommended

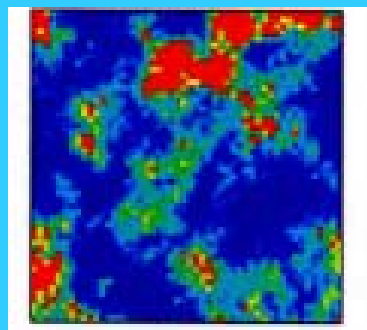
$\Rightarrow$ main purpose is to describe Sequential Gaussian Simulation (SGSIM)

Estimation versus Simulation

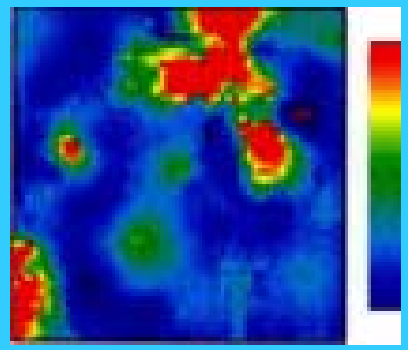
True Distribution



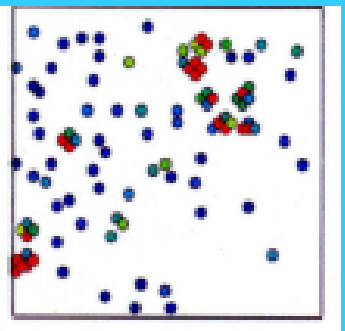
Simulation



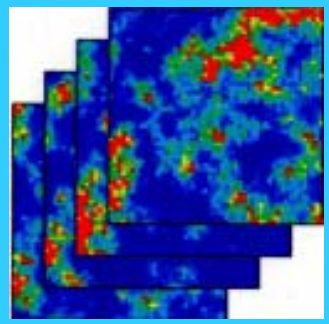
Kriging



Location Map of Sample Data



Multiple Simulated Realizations



- Estimation is locally accurate and smooth, appropriate for visualizing trends, inappropriate for flow simulation where extreme values are important, and does not assess of global uncertainty
- Simulation reproduces histogram, honors spatial variability (variogram), $\mapsto$ appropriate for flow simulation, allows an assessment of uncertainty with alternative realizations possible

Sequential Simulation: Theory

- Recall the kriging estimator: $$Y^*(u) = \sum_{\beta=1}^{n} \lambda_\beta \cdot Y(u_\beta)$$

  and the corresponding kriging system: $$\sum_{\beta=1}^{n} \lambda_\beta C(u_\alpha, u_\beta) = C(u, u_\alpha), \quad \forall u_\alpha$$

- The kriging system forces the covariance between the kriged estimate and the data values to be correct: $$Cov\{Y^*(u), Y(u_\alpha)\} = \sum_{\beta=1}^{n} \lambda_\beta C(u_\alpha, u_\beta) = C(u, u_\alpha) = C\{Y(u), Y(u_\alpha)\}$$

- Although the covariance between the estimates and the data is correct, the variance is too small: $$Var\{Y^*(u)\} = C(0) - \sigma^2_{SK}(u)$$

  correct the variance without changing the covariance by adding an independent (random) component with the correct variance: $$Y_s(u) = Y^*(u) + R(u)$$

  where $R(\mathbf{u}\})$ corrects for the missing variance.

- Covariance between kriged/simulated values is not correct: $$Cov\{Y^*(u), Y^*(u')\} \neq C\{Y(u), Y(u')\}$$

- The idea of sequential simulation is to use previously kriged/simulated values as data $\mapsto$ reproduce the covariance between all of the simulated values!

Why Sequential Gaussian Simulation? (1)

Gaussian distribution is used because it is extraordinarily straightforward to establish conditional distributions: shape of all conditional distributions is Gaussian (normal) and the mean and variance are given by kriging

1. Transform data to normal scores in the beginning (before variography)
2. Simulate 3-D realization in "normal space"
3. Conditional distributions are calculated by kriging to honor:
   - global histogram: $N(0,1)$
   - local data
   - secondary data (seismic, production data, **...**)
4. Back-transform all of the values when finished

$\Rightarrow$ Price of mathematical simplicity is the characteristic of maximum spatial entropy, i.e.,

low and high values are disconnected. Not appropriate for permeability.

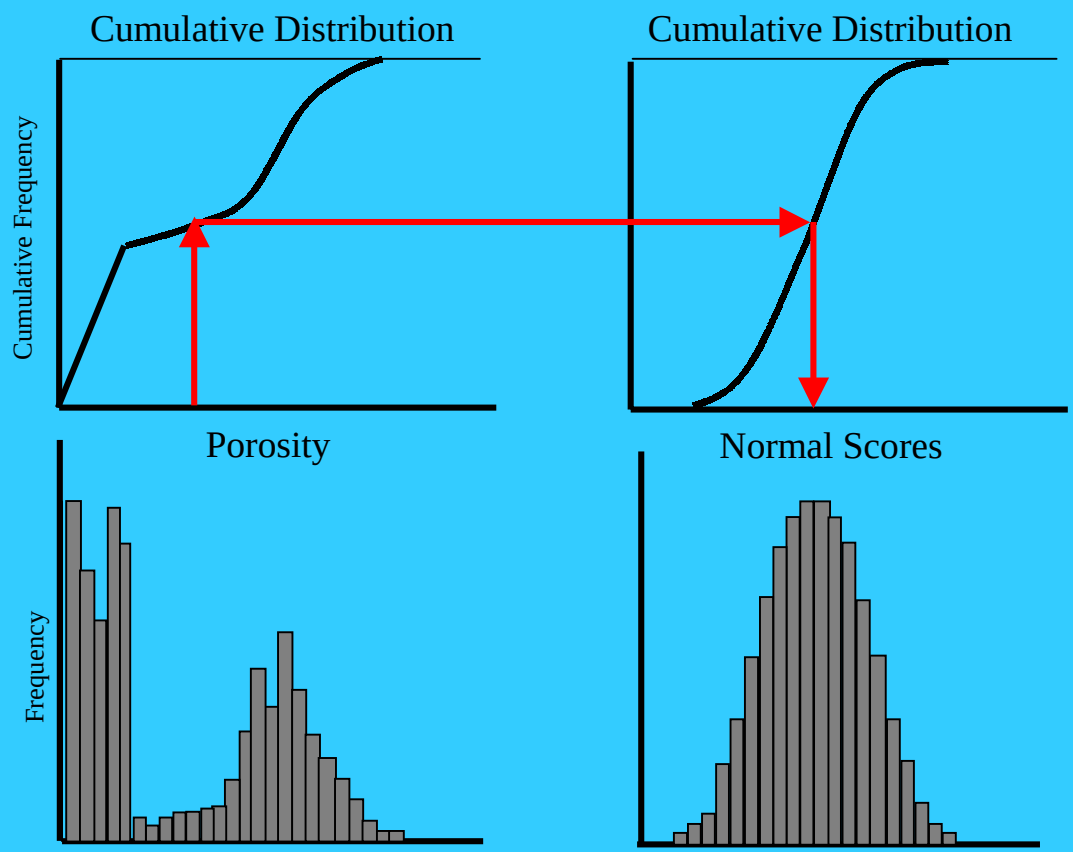
Why Sequential Gaussian Simulation? (2)

Steps in SGSIM:

1. Transform data to “normal space”
2. Establish grid network and coordinate system (Z_{rel}-space)
3. Decide whether to assign data to the nearest grid node or keep separate
4. Determine a random path through all of the grid nodes
   (a) search for nearby data and previously simulated grid nodes
   (b) construct the conditional distribution by kriging
   (c) draw simulated value from conditional distribution
5. Back transform and check results

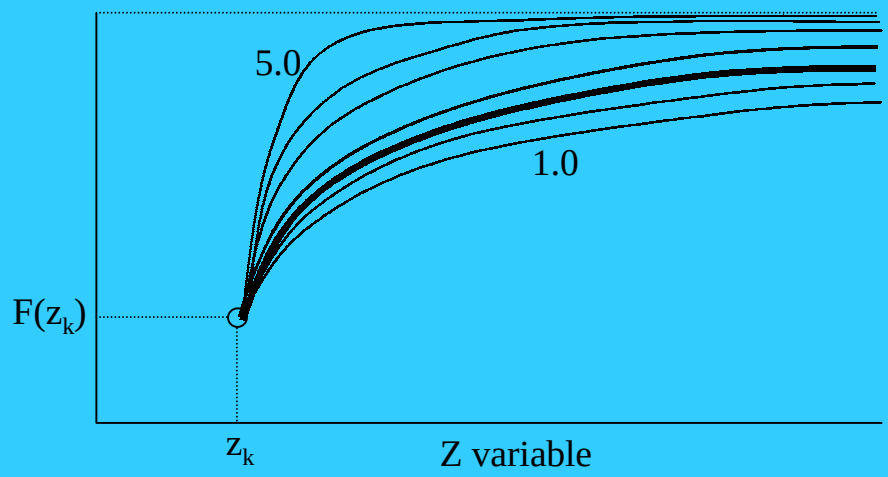
Normal Scores Transformation

"graphical" one-to-one (rank preserving) transform



Transformation to/from Normal Space

- Declustering weights are necessary
- Use “global” transformation table for small-area or cross-sectional models ↦ there may be too few data in the area of interest
- “Tail” options:
  - typically simulate the property at many more grid nodes than there are data ↦ expect higher values than observed in the data (and lower).
  - almost always a linear interpolation to user-defined minimum and maximum values is acceptable
  - can account for skewness in the upper tail with a hyperbolic model

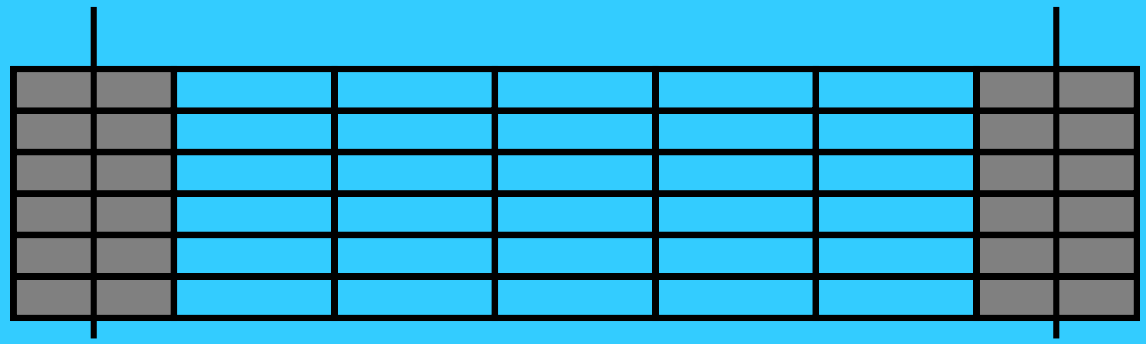


$$F_{\omega,\lambda}(z) = 1 - \frac{\lambda}{z^{\omega}}, \quad \omega \geq 1$$

Establish Data and Grid Network

- Work within “homogeneous” lithofacies/rock-type classification $\mapsto$ may require a first step to model lithology
- Z_{rel} vertical coordinate space
- Clean data: positioned correctly, manageable outliers, grid spacing is appropriate
- Considerations for areal grid size:
  - practical limit to the number of cells
  - need to have sufficient resolution so that the upscaling is meaningful
  - this resolution is required even when the wells are widely spaced (simulation algorithms fill in the heterogeneity)
- Work with “grid nodes”. We assign a property for the entire cell knowing that there are “sub-cell” features

Search: Two-Part or Assign Data to Grid Nodes



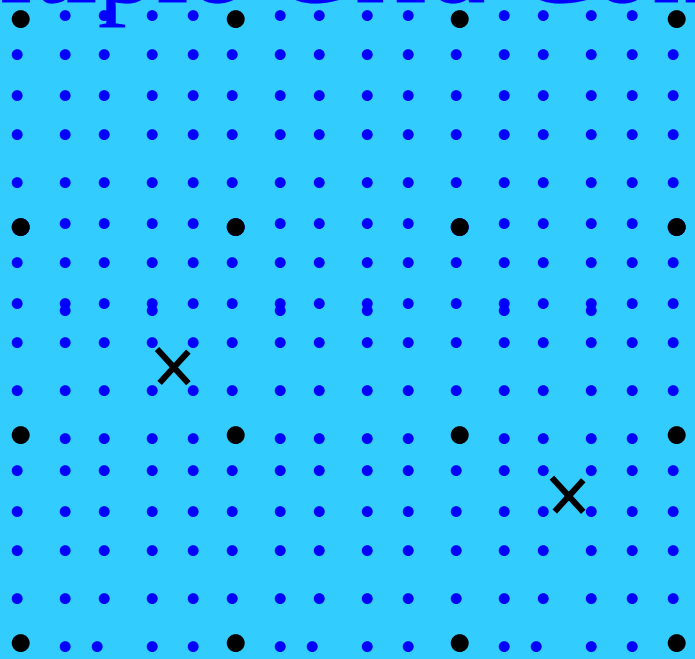
Two-Part?

- search for previously simulated nodes and then original data in two steps $\mapsto$ then treat the same when constructing the conditional distribution
- honor the data at their locations even if they can not be seen in the final model
- necessary for cross-sectional or small-area models

Assign Data to Grid Nodes:

- explicitly honor data - data values will appear in final 3-D model
- improves the CPU speed of the algorithm: searching for previously simulated nodes and original data is accomplished in one step

Multiple Grid Concept



- We typically limit ourselves to the nearest 12-48 grid nodes
- As the sequential simulation proceeds (fine grid) only the close samples will be used
- Variogram reproduction can be poor, especially if the variogram range is large with respect to the grid node spacing
- Multiple grid concept: simulate a coarse grid first and then refine one or more times
- Searching for data is more complicated:
  - two-part search at coarse levels
  - perhaps relocate to nearest fine grid node first (unless a two-part is used at the finest level)
- A number of implementation considerations

Number of Data to Consider

Reasons for more:

- theoretically better
- more accurate estimate of the conditional mean and variance
- better reproduction of the variogram

Reasons for less:

- CPU time is proportional to N^3
- memory requirements proportional to N^2
- negative weights are commonly encountered when data are screened
- using fewer data places less emphasis on the assumption of stationarity

So, choose between 12 to 48 depending on:

- 2-D versus 3-D
- range of variogram relative to grid node spacing
- CPU time restrictions

Type of Kriging

- Simple Kriging (SK):

$$m^*_{SK} = \sum_{i=1}^{n} \lambda_i \cdot z(u_i) + \left[1 - \sum_{i=1}^{n} \lambda_i\right] \cdot m_{global}$$

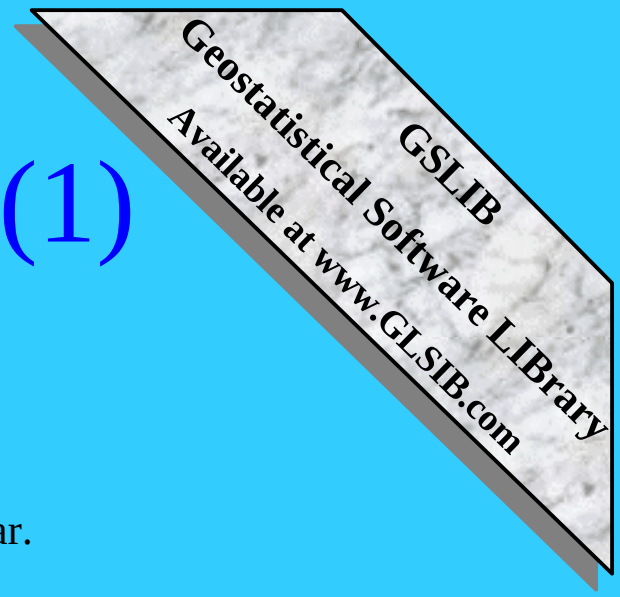
- Ordinary Kriging (OK) - constrain

$$m^*_{OK} = \sum_{i=1}^{n} \lambda_i \cdot z(u_i)$$

- Other Types:
  - Universal Kriging (UK) $\mapsto$ accounts for simple trends
  - External Drift $\mapsto$ accounts for more complex trends
  - Locally Varying Mean $\mapsto$ accounts for secondary information

Detailed Steps in SGSIM

1. Transform data to “normal space”
2. Establish grid network and coordinate system ($\boldsymbol{Z}_{rel}$-space)
3. Assign data to the nearest grid node (take the closest of multiple data assigned to the same node)
4. Determine a random path through all of the grid nodes
   (a) find nearby data and previously simulated grid nodes
   (b) construct the conditional distribution by kriging
   (c) draw simulated value from conditional distribution
5. Check results
   (a) honor data?
   (b) honor histogram: $N(0,1)$ - standard normal with a mean of zero and a variance of one?
   (c) honor variogram?
   (d) honor concept of geology?
6. Back transform



SGSIM Program (1)

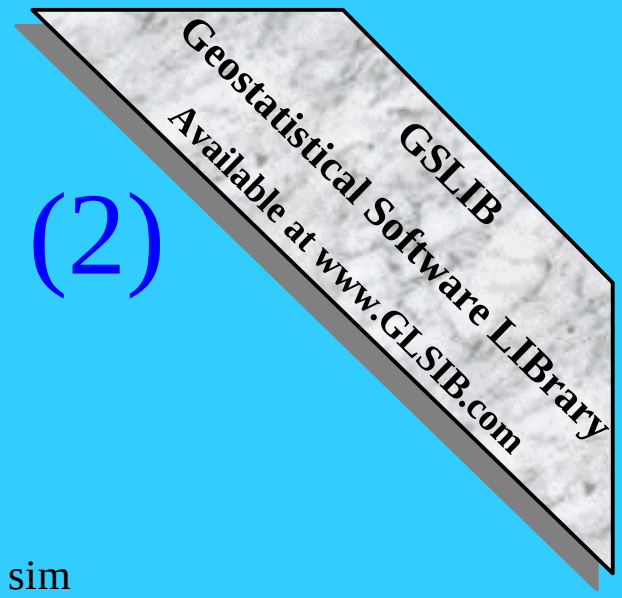
Parameters for SGSIM

****

START OF PARAMETERS:

../data/cluster.dat	\ file with data
1 2 0 3 5 0	\ columns for X,Y,Z,vr,wt,sec.var.
-1.0 1.0e21	\ trimming limits
1	\ transform the data (0=no, 1=yes)
sgsim.trn	\ file for output trans table
0	\ consider ref. dist (0=no, 1=yes)
histsmth.out	\ file with ref. dist distribution
1 2	\ columns for vr and wt
0.0 15.0	\ zmin,zmax(tail extrapolation)
1 0.0	\ lower tail option, parameter
1 15.0	\ upper tail option, parameter
1	\ debugging level: 0,1,2,3
sgsim.dbg	\ file for debugging output
sgsim.out	\ file for simulation output
5	\ number of realizations to generate

...



SGSIM Program (2)

```
50    0.5    1.0              \ nx,xmn,xsiz
50    0.5    1.0              \ ny,ymn,ysiz
1     0.5    1.0              \ nz,zmn,zsiz
69069                         \ random number seed
0     8                       \ min and max original data for sim
12                            \ number of simulated nodes to use
1                             \ assign data to nodes (0=no, 1=yes)
1     3                       \ multiple grid search (0=no, 1=yes),num
0                             \ maximum data per octant (0=not used)
10.0  10.0  10.0              \ maximum search radii (hmax,hmin,vert)
0.0   0.0   0.0               \ angles for search ellipsoid
4     0.60                    \ ktype: 0=SK,1=OK,2=LVM,3=EXDR,4=COLC
../data/ydata.dat             \ file with LVM, EXDR, or COLC variable
4                             \ column for secondary variable
1    0.1                      \ nst, nugget effect
1    0.9  0.0   0.0   0.0     \ it,cc,ang1,ang2,ang3
          10.0  10.0  10.0     \ a_hmax, a_hmin, a_vert
```