Declustering with Seismic or "Soft" Geological Data

C. V. Deutsch (cdeutsch@civil.ualberta.ca)
Department of Civil & Environmental Engineering, University of Alberta

P. Frykman (pfr@geus.DK)
Geological Survey of Denmark and Greenland (GEUS)

Y-L Xie (ylxie@civil.ualberta.ca)
Department of Civil & Environmental Engineering, University of Alberta

Abstract

Data are rarely collected for their statististical representativity. This is particularly true in petroleum exploration and production where wells are often drilled in locations to maximize future production or to delineate the most productive portions of the reservoir. Nevertheless, geostatistical simulation methods require distributions of facies proportions and petrophysical properties that are representative of the entire *reservoir being modelled. Therefore, there is a need of declustering procedures, which adjust the histogram and summary statistics to be representative of the entire volume of interest.*

In presence of severely limited well data, conventional declustering algorithms such as the polygonal and cell-declustering methods are not effective; there are too few data to assign relative weights. Regions of poorer reservoir quality may be indirectly observed by seismic data or geological interpretation, but may not be sampled. In this case, there are no hard data to "weight." In such cases, we must obtain a representative distribution with the aid of soft secondary data and a calibration relationship between the soft data and the primary variable under consideration.

We review the conventional declustering techniques in an appendix, and develop a soft-data *based declustering approach in the main part of the report. A representative distribution is obtained by merging conditional distributions inferred from calibration with the soft data variable. A program* sddeclus *is presented for* soft-data *declustering. The performance of the proposed approach is illustrated by an example. The comparative study illustrates the importance of declustering.*

KEY WORDS: unbiased distribution, geostatistics, stochastic simulation, secondary data, reservoir characterization

Introduction

All geostatistical simulation methods require distributions of rock types and petrophysical properties that are representative of the *entire* area of interest. In practice, data are collected on the basis of economic criteria and *not* for statistical representativity. Wells are drilled

in areas with the greatest probability of high production rates and core measurements are taken preferentially from good reservoir quality rock. Therefore, the greatest number of data often come from portions of the reservoir that are the most important (or accessible). These sampling practices should not be changed; however, numerical modellers must take this into consideration in building geologic models.

Declustering techniques have been devised to account for the fact that data are clustered in important or easily accessible areas. Journel proposed cell declustering [5], Deutsch presented some refinements and software [1], and Isaaks and Srivastava illustrate the importance with the Walker Lake Data Set [4]. Programs for declustering are available in GSLIB [2]. The methodology is also explained in the new book by Goovaerts [3].

It is clear that the naive equal-weighted sample distribution / histogram and sample statistics are biased in presence of preferential sampling. It is critical to go beyond the sample data and infer a representative histogram and other summary statistics before geostatistical simulation. This paper describes procedures for such inference.

Background

Data in densely sampled areas should receive less weight than those in sparsely sampled areas. The declustering techniques mentioned above [1, 4, 5] assign each data a weight $w_j, j = 1, \ldots, n$, based on the closeness of the data to the surrounding data. The weights are greater than 0 and sum to 1. A recall of the conventional declustering methods is presented in an Appendix to this paper. For notation, consider n data $z(\mathbf{u}_j), j = 1, \ldots, n$, where the variable z may represent a continuous variable or indicator transform of a categorical variable and $\mathbf{u}$ represents a location coordinate vector. Summary statistics such as the mean and variance are then calculated with the declustering weights:

$$\overline{z} = \sum_{j=1}^{n} w_j z_j \quad \text{and} \quad s^2 = \sum_{j=1}^{n} w_j \left(z_j - \overline{z} \right)^2 \tag{1}$$

the cumulative histogram, required for stochastic simulation and data transformation, is also estimated with the declustering weights:

$$F^*(z) = \sum_{j=1}^{n} w_j \cdot i(\mathbf{u}_j; z) \tag{2}$$

where

$$i(\mathbf{u}_j; z) = \begin{cases} 1, & \text{if } z(\mathbf{u}_j) \leq z \\ 0, & \text{otherwise} \end{cases}$$

Conventional declustering techniques, which assign each data a relative weight only work when there are enough data to assign greater and lesser weights. In many cases, the limited number of data makes it impossible to use these declustering algorithms for deriving critical statistics. Figure 1 is a contour map of the Dan Field in Danish North Sea. A obvious domal structure exists. The physical properties of the constituent chalk is such that porosity and permeability decrease with increasing depth (at the margins of the reservoir). Cored wells (circled) are mostly found in crestal area having the best reservoir
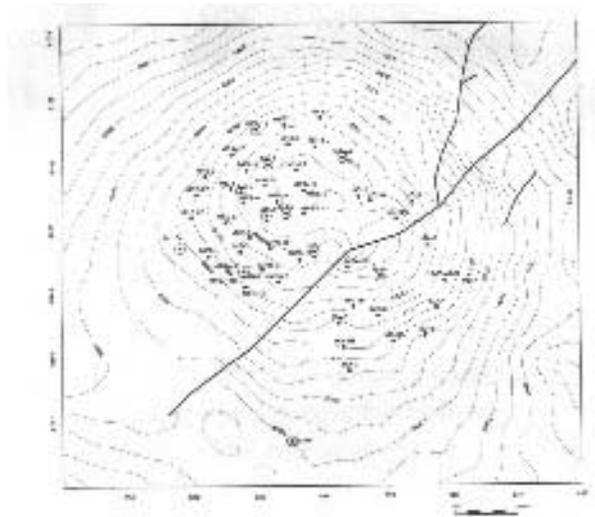
Figure 1: Dan Field, Danish North Sea. Domal structure on slat pillow.

properties, so the porosity and permeability distribution based on sampled core well data are *not* representative distributions of the *entire* reservoir.

There are too few data at the margin of the reservoir to use relative weights to infer a representative distribution. The result of conventional declustering techniques would also be biased. There is a need to use the secondary information, in this case depth, to aid in declustering.

If there is no secondary information provided by *soft-data*, the bias in the sample data and resultant statistics may *even* not be recognized. However, in many cases there are secondary *soft data* from geological interpretation (as in the case of the Dan Field) or seismic data. Soft data alone are of little help; the soft data must be complemented with a calibration relationship between the primary variable of interest and the soft data. Such a bivariate realtionship can be revealed based on limited measurements and supplementary knowledge of geophysics or geology.

In the Dan Field there exists a good relationship between depth of the reservoir and the reservoir quality. Seismic data provides depth at all locations throughout the reservoir. Generally, the deeper the reservoir, the smaller the value of porosity. This situation is illustrated by the sketch in Figure 2. The contour lines in Figure 2 represent a bivariate distribution of porosity and depth. The dots represent avialable data.

Obviously, the estimated distribution of porosity (black profile along *porosity* axis in Figure 3) based *only* on the *well data* (black dots in Figure 3) will be a biased estimate of the underlying *true* distribution of porosity (dotted blue profile in Figure 3). The lack of data in the margin of the reservoir makes it difficult to correct the distribution by the conventional declustering techniques. A more sophisticated method is thus needed.
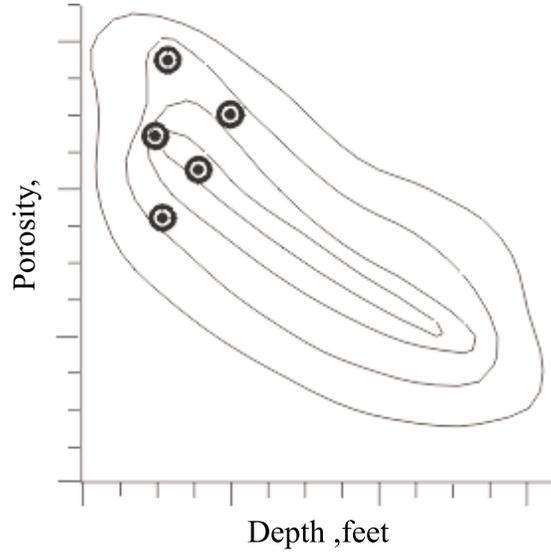
Figure 2: Schematic illustration of the bivariate relationship between porosity and depth.
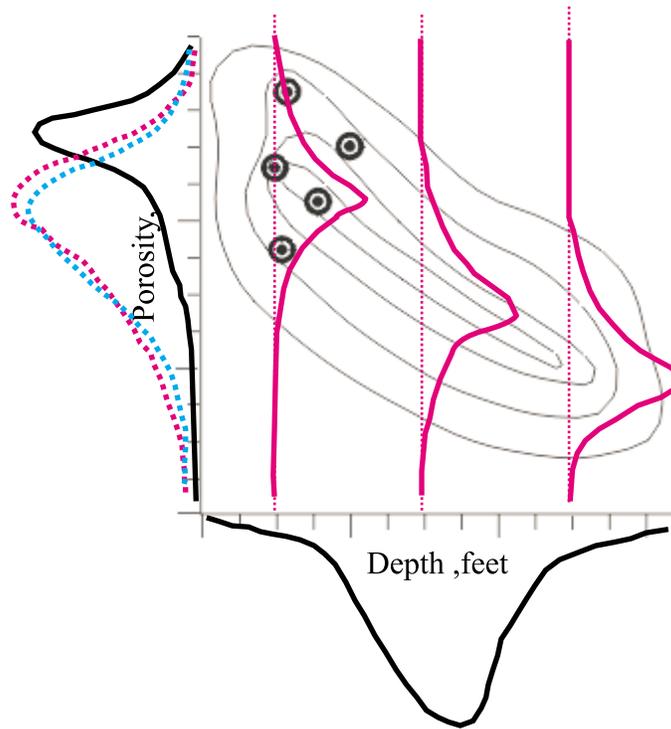

Figure 3: Schematic illustration showing naive distribution of porosity, conditional distributions of porosity given depth, representative distribution of depth, and the representative distribution of porosity.

Methodology

In this report, an approach is proposed to infer the distribution of a primary variable based on secondary *soft data*. The methodology has also been presented (within a detailed example and computer software) in a two page extended abstract [**?**]. The central idea is illustrated schematically on Figure 3. The exhaustive secondary data is representative of the entire domain of interest, see the black profile along *depth* axis in Figure 3). With the aid of the bivariate relationship between primary and secondary variables (contoured lines in Figure 3), distributions of the primary variable, conditional to the secondary variable, can be estimated (red profiles in Figure 3). Consequently, an estimate of the global distribution of the primary variable (dotted red profile along *porosity* axis) can be obtained by merging the local conditional distributions from the distribution of secondary variable.

Even before details of the proposed methodology are described, we can see that the bivariate relationship is key. The mechanics of merging conditional distributions into a representative distribution of the primary variable of interest depends entirely upon the goodness of the calibration relationship. Notwithstanding our dependence on this relationship, we show that using a poorly known calibration is better than ignoring an important trend. The following assumes that the calibration relationship is available for all possible values of the secondary variable. The problem of declustering in presence of soft data is transferred to inference of this calibration cross plot.

A distribution of the primary variable Z conditional to the secondary variable $y(\mathbf{u})$ at a particular location can be calculated from the calibration relationship. These local conditional distributions for all locations in the area of interest A may be combined with equal weight since the secondary variable is available at all locations in the area, that is, $y(\mathbf{u}), \mathbf{u} \in A$. In probabilistic notation, a representative distribution of the primary Z variable is then written:

$$f_Z^*(z) = \sum_{\text{all } \mathbf{u} \in A} \frac{1}{C} f_{Z|Y}\left(z | Y = y(\mathbf{u})\right) \tag{3}$$

where C is a normalization constant to ensure the resulting distribution integrates to 1.0.

Thus, the approach consists of the following steps:

- assemble a map of the secondary variable Y at all locations,

- develop the essential bivariate relationship between Y and the primary variable Z,

- calculate all local conditional distributions of Z given every Y value, and then

- add them to create a global Z distribution.

As already mentioned, the procedure to infer the calibration cross plot is critical. The program to perform the above steps requires the local conditional distributions expressed in the form of p-quantiles. The program `sddeclus`, written in FORTRAN 90, is used for the implementation of the proposed approach. The parameter file is shown on Figure 4. The parameters:

- **datafl:** input data file containing the exhaustive data of the secondary variable.

```
 Parameters for sddeclus
 **********************

START OF PARAMETERS:
..\data\DATAset.out                          -input exhausted secondary variable
..\data\GloblPDF.out                         -output file for representative distribution
4000                                         -number of grid points in soft data
0.00,40.00                                   -minimum and maximum of primary variable
201                                          -number of classes for output histogram
10,5                                         -# soft threshold, # primary cutoffs
      0.00, 0.25, 0.50, 0.75, 1.00           -cumulative probability of primary cutoffs
1000  22   26    28    31    34              -Interval, Quantiles at cutoffs
1020  12   18    20    22    29              -Interval, Quantiles at cutoffs
1040   9   15    16    18    25              -Interval, Quantiles at cutoffs
1060   7   13    15    16    22              -Interval, Quantiles at cutoffs
1080   6   11    13    14    19              -Interval, Quantiles at cutoffs
1100   5    9    10    12    16              -Interval, Quantiles at cutoffs
1120   4    7     9    10    14              -Interval, Quantiles at cutoffs
1140   3    6     7     8    12              -Interval, Quantiles at cutoffs
1160   2    4     5     6     9              -Interval, Quantiles at cutoffs
1170   0    2     3     4     7              -Interval, Quantiles at cutoffs
```

Figure 4: Example parameter file for `sddeclus`.

- **pdffl:** output file with the calculated distribution of the primary variable in the format of a probabilty density function (PDF),

- **nsize:** size of secondary data grid.

- **min, max:** minimum and maximum value of the primary variable.

- **no_class:** number of classes (*resolution*) for the output PDF of the primary variable

- **no_thresholds, no_cutoff:** number of *control* points along secondary variable axis and the number of cutoffs in the *quantile* function of the primary variable

- The cumulative probability cutoffs for the *quantiles* used for calibration.

- *control* points along secondary variable axis and the *quantiles* of primary variable at those *control* points. They contain information of the *well data* and the bivariate relationship between primary and secondary variables.

An Example

A cross section of a reservoir having a domal structure is used for an example, see the top plot of Figure 5. The reservoir section is gridded into a 200 (horizontal) by 20 (vertical) cells. The depth is available for every cell in the entire reservoir, which is in the range -1000

to -1170 meters. The thickness of the straigraphic layer is about 20 meters. Porosity data has been generated for every grid of the reservoir, which serves as the *reference* data. In the generation of data, a negative correlation between porosity and depth is assumed, that is, the larger the depth, the smaller the porosity. Five vertical wells are available in the center area of the cross section (vertical red lines in Figure 5). The porosity data and depth at well locations are taken as the well data.

Porosity is the primary variable of interest and depth is the secondary variable. The goal is to determine a representative distribution of porosity based on porosity measurements at the well locations and a bivariate relationship between porosity and depth.

The cross plot of porosity versus depth for the reference data is shown at the bottom right plot of Figure 5 shows. The reference histogram of porosityis shown at the bottom left of Figure 5. Of course, this information would be *inaccessible* in practice and is used here for comparison purposes only.

Figure 6 is the scatter plot of porosity versus depth of the well data, and Figure 7 is the histogram of porosity of the *well data*. Comparing to the histogram in Figure 5, it is obvious that the *well data* only contains high porosity values (mean value of 21.02 versus the reference mean 14.39) and the histogram derived from the well data is not representative. The reason is the domal structure of the reservoir, the trend of porosity with depth, and wells located in the center area of the reservoir. Figure 8 and Figure 9 are the histograms of the depth of the *reference data* and the *well data*, respectively.

Since there is no data avi)lable in the margins of the reservoir, specifically in the areas with greater depth, the conventional declustering techniques will be unable to get a set of proper weights. Figure 10 and Figure 11 are the histograms after *cell declustering* and *polygonal declustering*, respectively. As expected, there is little improvement in the estimate of the distribution and summary statistics by using these conventional declustering techniques.

Based on geological information, the larger the depth the smaller the porosity. A bivariate model describing the bivariate realtionship between porosity and depth must be derived. Figure 12 shows local distributions of porosity conditional to the depth. Based on the *well data* which are shown as black dots in the figure, local conditional distributions of porosity are calculated in the depth range covered by the *well data*. A set of *control* points is chosen along the *entire* depth axis including both extreme values. Based on the local conditional distributions derived from the *well data* and considering the bivariate structure suggested in Figure 2, the *quantile* functions on the *control* points are estimated by extrapolation and shown as colorful star signs in Figure 12. Different color represents different *quantile*. Then, the *whole* set of local conditional distributions of porosity on every depth value are estimated by linear interpolation from those of *control* points, and they are shown as the colorful lines in Figure 12.

In this way, the information embedded in the *well data* and the bivariate model is represented by the *quantile* functions, that is, the lines in Figure 12. They provide the input cutoffs, *quantiles* of the parameter file. It should be noted that usually one only has a conceptual bivaraite model *or* bivariate relationship with certain variability. The bivariate model constructed is only an approximate representation of the underlying bivariate distribution between primary and secondary variables. Generally, the more well data the better our understanding of the bivariate relationship between primary and secondary variables,
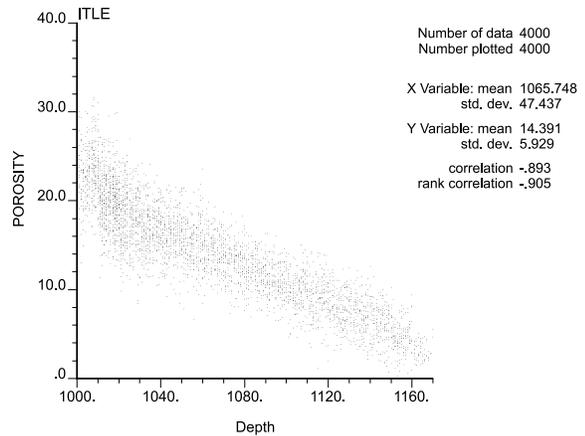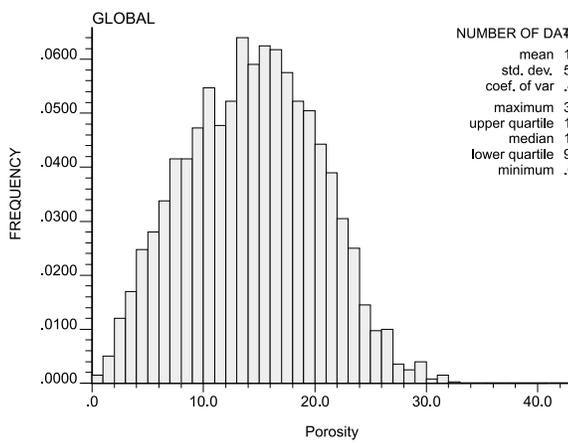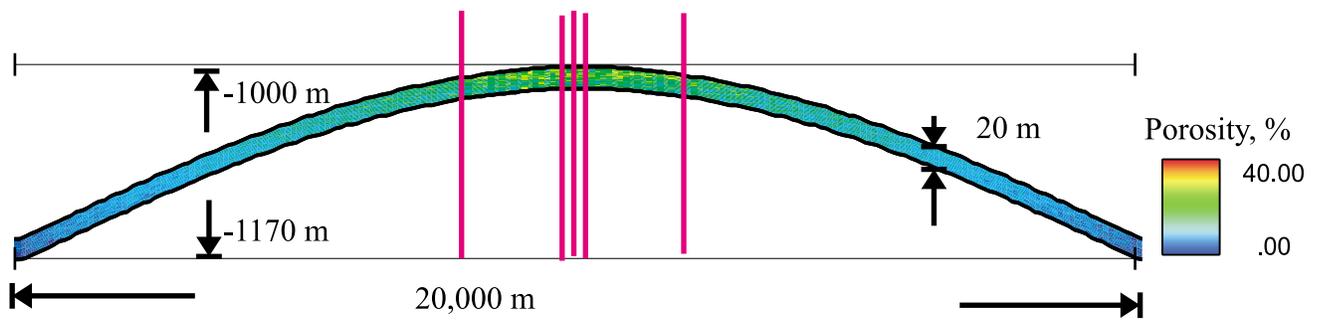
Figure 5: Cross section view of a reservoir (*top*). The color scale codes shows the underlying *true* distribution of porosity. The true histogram of porosity (*bottom left*) and cross plot of porosity versus depth (*bottom right*) are also shown.
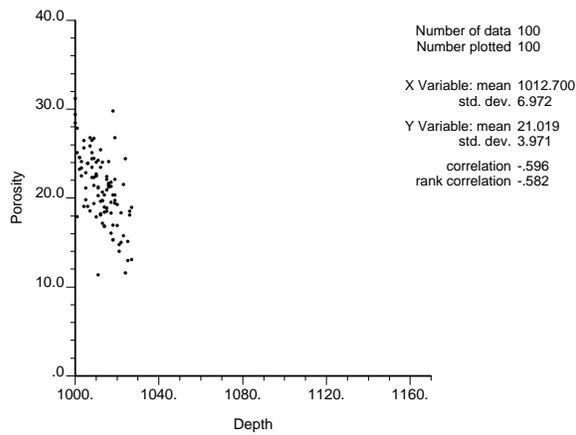
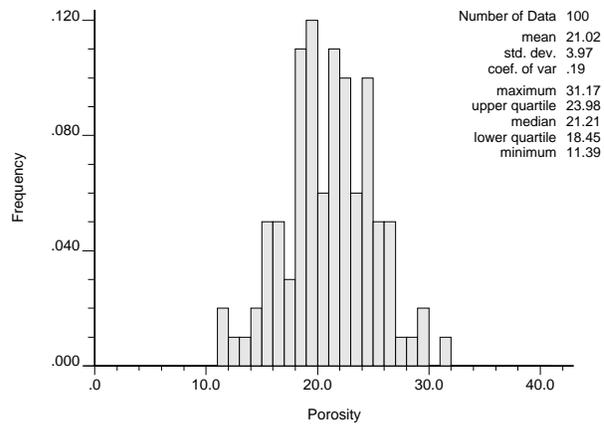Figure 6: Scatter plot of porosity versus depth of the *well data*.

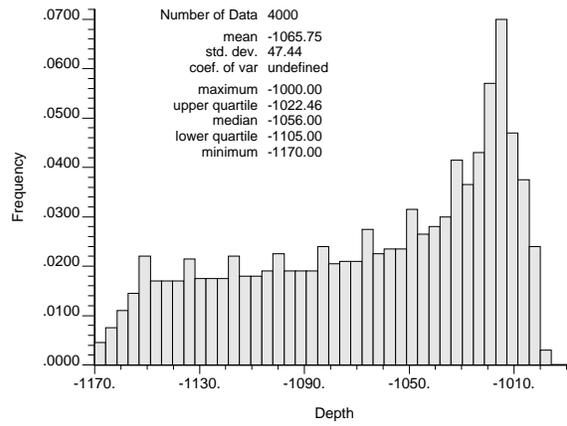Figure 7: Histogram of porosity derived from the *well data*.

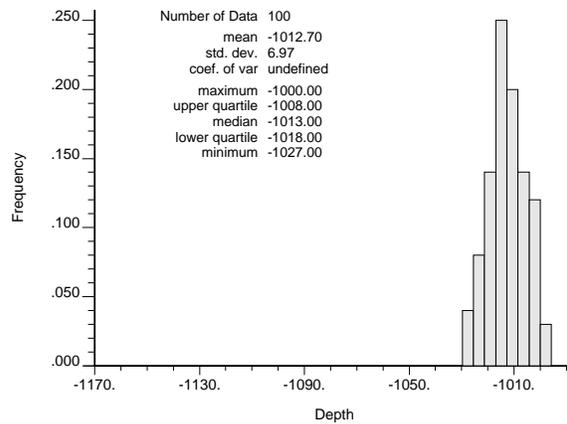Figure 8: Histogram of depth of the *entire* reservoir.
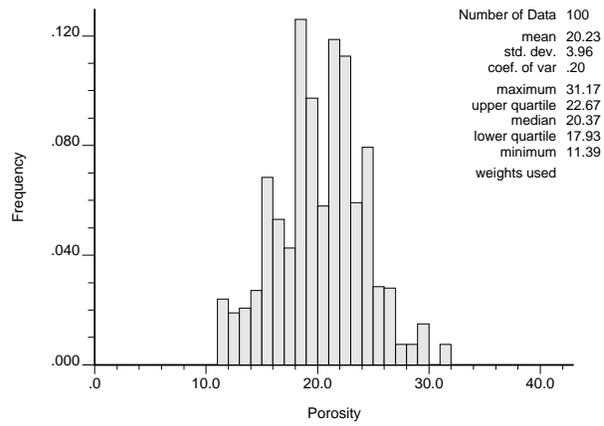

Figure 9: Histogram of depth of the *well data*.

Figure 10: Histogram of porosity from the well data after cell declustering.
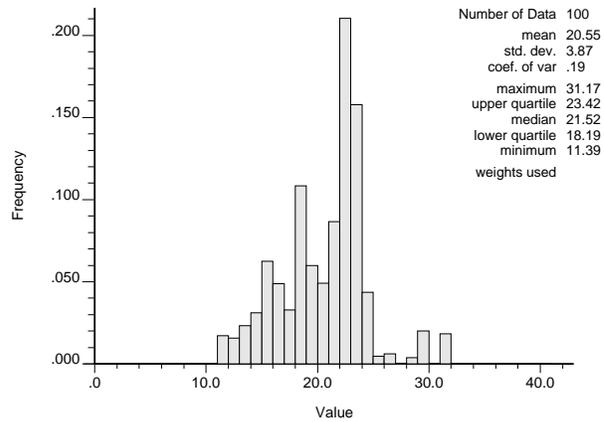
Figure 11: Histogram of porosity from the well data after polygonal declustering.
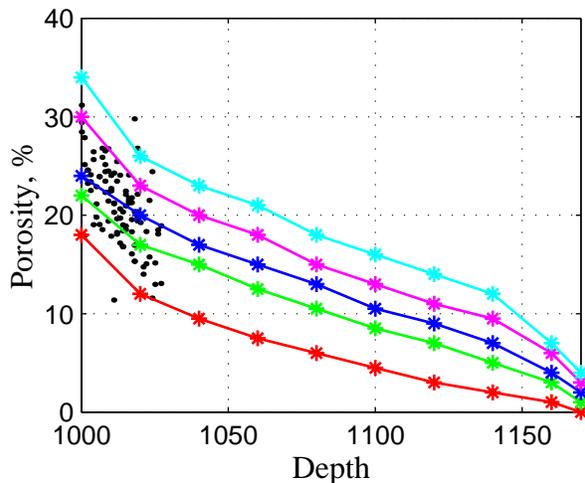
Figure 12: Local conditional distributions of porosity. The black dots are the well data and colorful star signs are control points. The red line is the minimum 0.0 quantile; green: 0.25 quantile; blue: 0.50 quantile; magenta: 0.75 quantile and cyan: 1.00 quantile.

and the better quantile functions will be constructed.

Simple linear interpolation and extrapolation were used to extend the estimates of local conditional distributions from range covered by the *well data* to the *entire* area of reservoir followed an imprecise *trend*.

Figure 13 is the estimated histogram from `sddeclus`. Comparing to the *reference* histogram in Figure 5 and the histogram derived from the *well data* in Figure 7, not only the mean (14.92%) is much closer to the *reference* mean (14.39%), the shape of the distribution is closer to the *reference* distribution. In particular, the distribution now covers the entire range including the low-valued area which is not covered by the *well data*.

Simulation with Different Distributions

As mentioned above, the purpose for declustering is to get a representative distribution for subsequent geostatistical simulations. We use sequential Gaussian simulation to generate numerical realizations of porosity based on the histogram distribution before and after correction. Since data should be transferred into normal score values for Gaussian simulation, the estimate of data distribution has a tremendous influence on the result of sequential Gaussian simulation.

The `sgsim` program in GSLIB is used for sequential Gaussian simulation. Four scenarios are considered: (1) simulation based on the distribution derived from the *well data* only (shown in Figure 7), (2) simulation based on the distribution derived from the *well data* and considering a *trend* of porosity with depth, (3) simulation based on the estimated
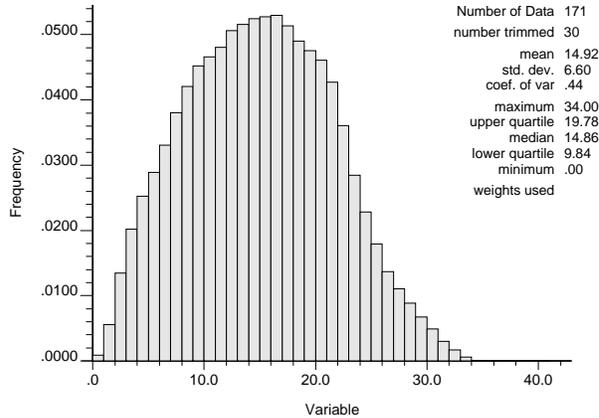
Figure 13: The estimated histogram of porosity.

distribution by the proposed approach (shown in Figure 13), and (4) simulation based on the estimated distribution by the proposed approach and considering the *trend*. One hundred realizations are generated for each scenario. Since there was little improvment in the estimate of the distribution by cell declustering or polygonal declustering, no simulation was conducted based on those distributions.

Figure 14 shows the histograms of mean porosity resulting from the one hundred realizations of each scenario. The red vertical line in the plots represents the mean value of the *reference data* and blue vertical line denotes the mean value of the *well data*. It can be seen from Figure 14 that the mean values of simulation are far away from the *reference* mean (positive biased) when simulations are based on the distribution derived from the *well data* only. Simulations based on the estimated distribution by the proposed approach give quite acceptable mean values.

Figure 15 to Figure 18 show cross section view, histogram of simulated porosity and scatter plot between simulated porosity and depth for the first realizations of each simulation scenario, respectively.

The histograms of simulated porosity shown in Figure 15 and Figure 16 have similar shapes with the histogram derived from *well data* (Figure 7). They are systematically biased from the *reference* distribution (Figure 5). However, the trend structure, *i.e.,* higher porosity in low depth area and lower porosity in higher depth area, is reproduced when trend is considered in the simulation (Figure 16).

When the corrected distribution by the proposed approach adapted in the simulation, the histogram of the simulated porosity (Figure 17) shows similar shape to that of the *reference* distribution (Figure 5). However, the porosity image on the top plot of Figure 17 is still far away from the reality (top plot of Figure 5) when the trend has not been considered in the simulation. A more realistic image is obtained when both corrected distribution and trend are considered in the simulation (Figure 18).
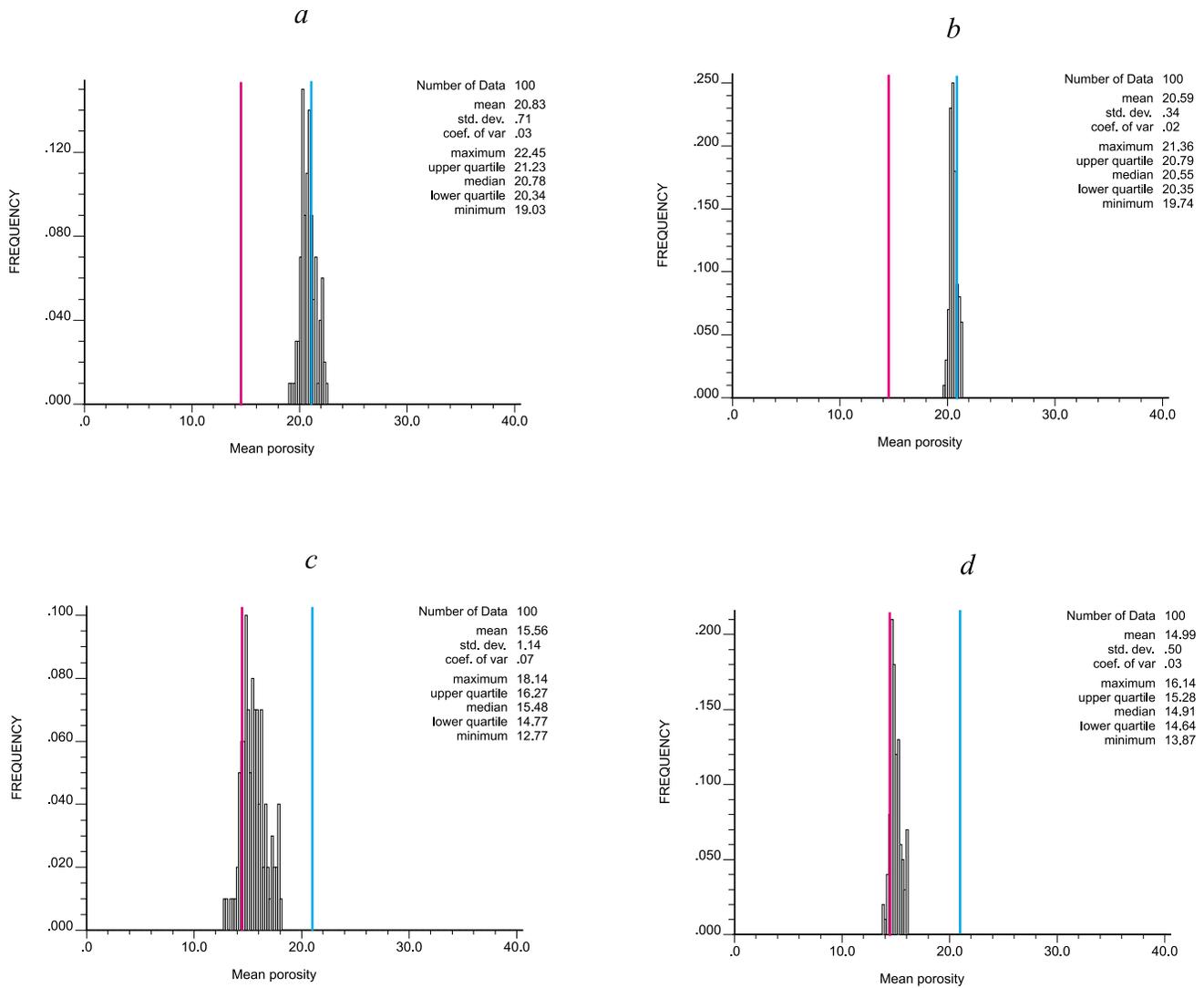
13

Figure 14: Histograms of means from 100 realizations; *a, b* refer to the simulations based on the *well data* without and with the consideration of the trend; *c, d* refer to the simulations based on the estimated distribution without and with the consideration of trend, respectively; red vertical line: mean of the *reference data*; blue vertical line: mean of the *well data*.
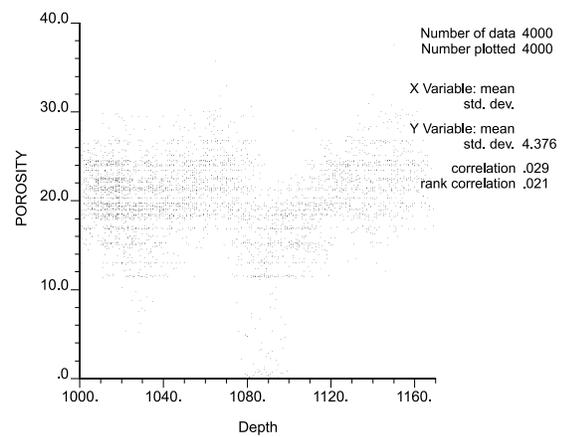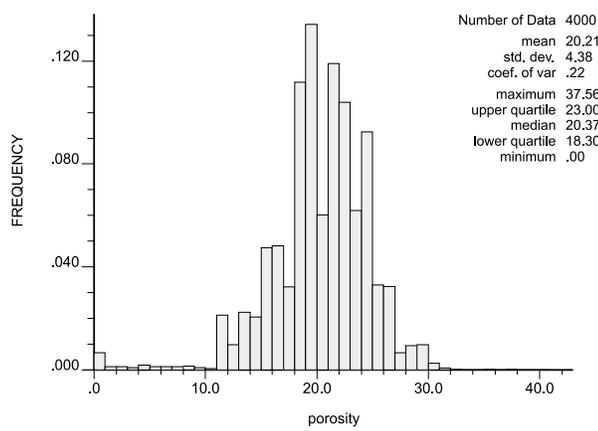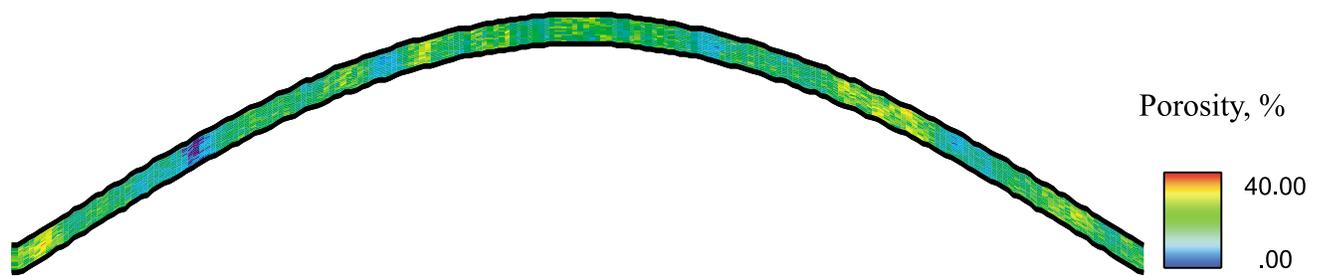
Figure 15: Results of one realization from simulation based on the well data without considering the trend; *top*: Cross section view; *bottom left*: histogram of simulated porosity; *bottom right*: Scatter plot of simulated porosity versus depth.
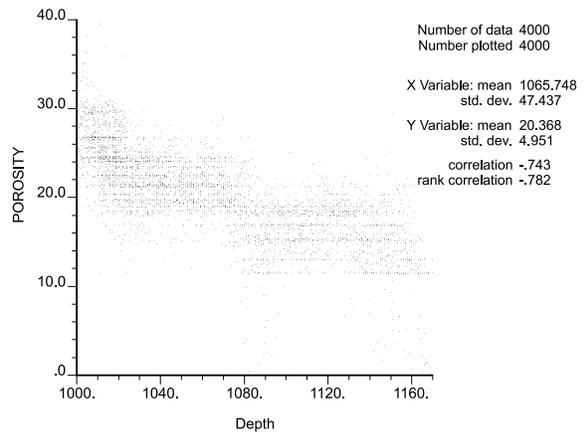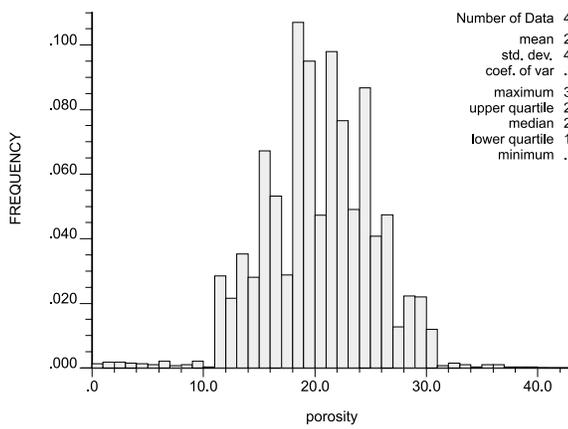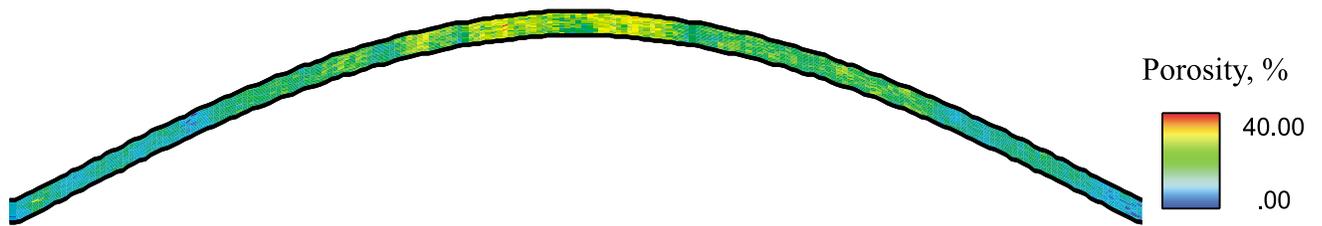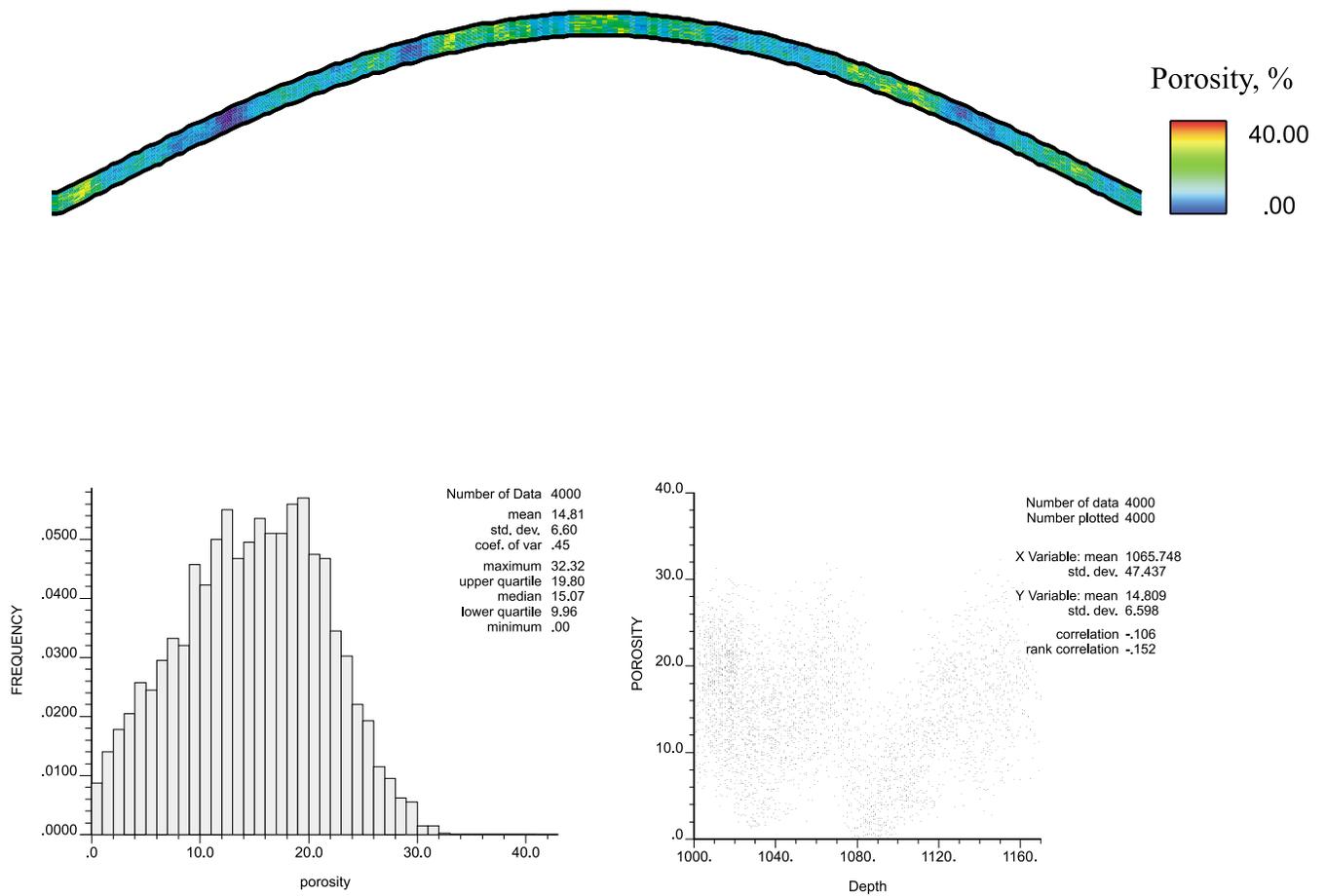
Figure 16: Results of one realization from simulation based on the well data with the consideration of the trend;*top*: Cross section view; *bottom left*: histogram of simulated porosity; *bottom right*: Scatter plot of simulated porosity versus depth.

Figure 17: Results of one realization from simulation based on the estimated distribution without considering the trend; *top*: Cross section view; *bottom left*: Histogram of simulated porosity; *bottom right*: Scatter plot of simulated porosity versus depth.
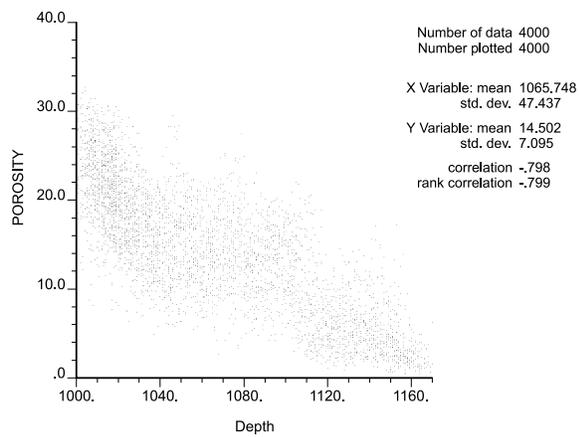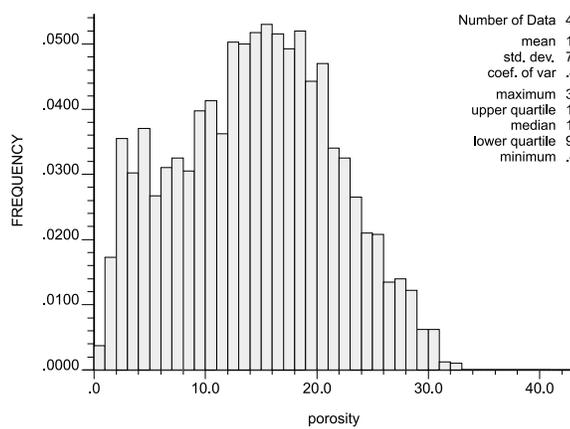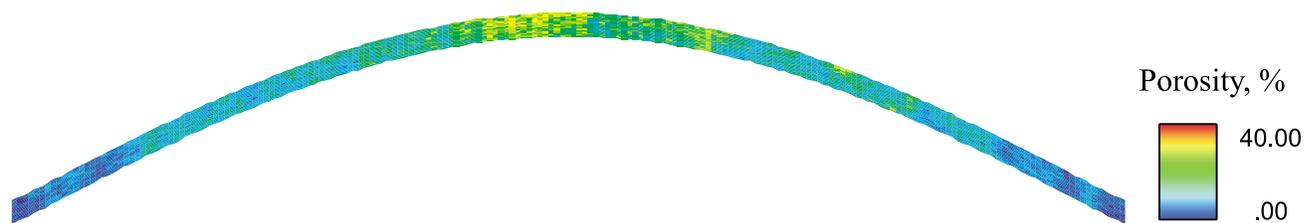
Figure 18: Results of one realization from simulation based on the estimated distribution with the consideration of the trend; *top*: Cross section view; *bottom left*: Histogram of simulated porosity; *bottom right*: Scatter plot of simulated porosity versus depth

To further compare the distributions of porosity generated from different simulations, the Q-Q plots between the simulated porosity of the first realization from each scenario and the *reference data* are shown in Figure 19. Q-Q plot *a* and *b* of Figure 19 suggest that simulated porosity based on the distribution derived from the *well data* is a distribution different from the *reference* distribution in terms of mean (higher mean), variance (not a 45° slope) and shape (non-linear character).

When the distribution is corrected by the proposed appraoch, the Q-Q plots between simulated porosity and the reference data are very close to the 45° line, which means the two distributions become similar.

In summary, simulation based on the distribution estimated by the proposed approach and considering the *trend* provides the best results, whereas, simulation based on the *well data* only and without considering the *trend* gives the worst results.

The bivariate model is critical to the success of the proposed approach. The more well data and the more supplementary knowledge about the bivariate relationship between primary and secondary variables, the better the bivariate model that will be built, subsequently, the better estimate of the distribution one will get. It is unlikely that considering the information provided by a secondary variable will provide worse results than the naive histogram.

Figure 20 shows a bivariate model which is obviously unreasonable. The bivariate model suggests that there will be zero porosity after a certain depth in the reservoir. Figure 21 is the estimated distribution of the global porosity. As the distribution derived from the *well data* (Figure 7) this estimate of the distribution does not represent the *reference* distribution (Figure 5) either, but it is still a better estimate than the one in Figure 7.

One hundred realizations are generated by using this estimated distribution and considering the trend. The means of the hundred realizations are close to the reference mean. Looking at section views, scatter plots between the simulated porosity and depth and the histogram of the simulated porosity, the result is still better than that only considering the well data.

The example is two-dimensional for the purpose of easy demonstration, but there is no problem to extend the approach to three dimension. The essence of the method is to construct decent local conditional distributions of primary variable on secondary variable, which depneds only on the pairwise relationship between primary and secondary variables and has nothing to do with the actual coordinate of the grid. So there will be no difference due to the dimensionality.

Conclusions

Geostatistical simulation algorithms reproduce the input target histogram. It is essential to have a representative histogram for the entire volume being simulated. A general-purpose algorithm for declustering in presence of soft data has been presented in this report. Conventional algorithms are recalled in an Appendix.

Another conclusion of this report is that it is essential to directly account for systematic trends, for example, a decrease in porosity with depth. It is insufficient to merely correct the histogram to be representative of the entire domain; the histogram must be corrected *and* the geostatistical simulation algorithm must directly account for the trend.
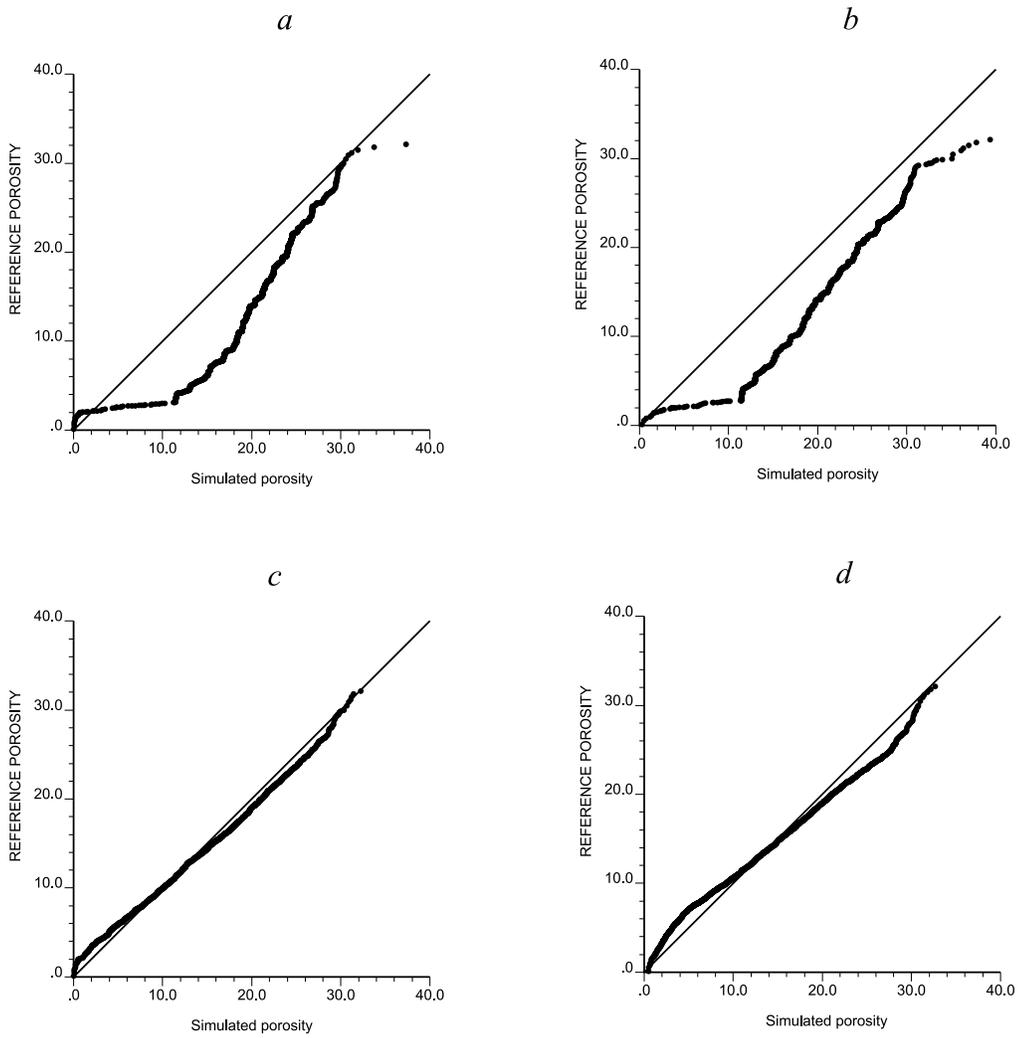
Figure 19: Q-Q plots between the simulated and the *reference* porosity
a, b, c, d refer to simulation based on the well data without and with the consideration of the trend and based on the estimated distribution without and with the consideration of trend, respectively.
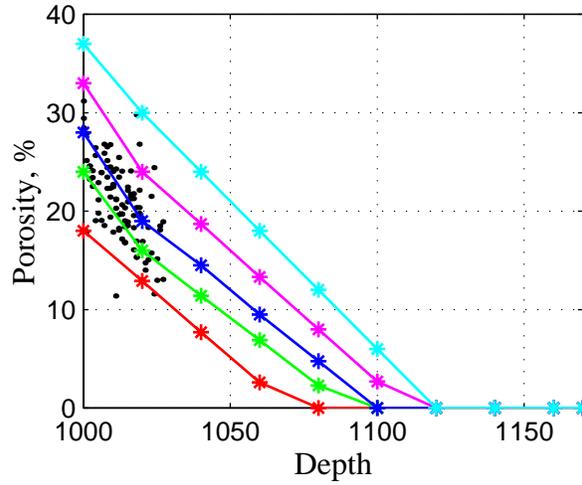
Figure 20: Local conditional distributions of porosity. The black dots are the well data and colorful star signs are control points, the red line: 0.00 quantile; green: 0.25 quantile; blue: 0.50 quantile; magenta: 0.75 quantile and cyan: 1.00 quantile.
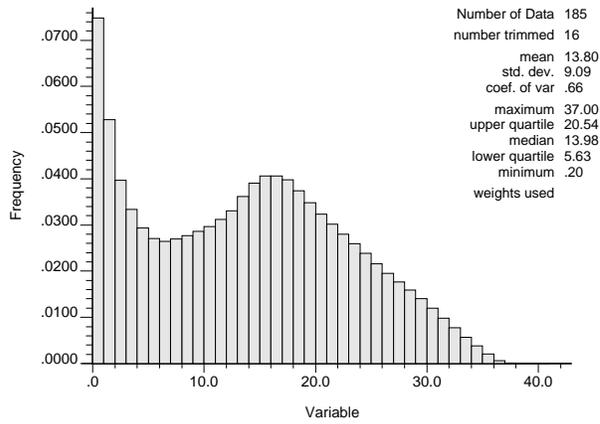

Figure 21: The estimated histogram of porosity corresponding to the bad calibration.

Many commercial software packages strive for the 80% solution, which does not include declustering. Declustering is not needed when the wells are equally spaced or when there is no evidence of trends. Preferential well locations and systematic geologic trends may be more common than 20% of the time!

References

[1] C. V. Deutsch. DECLUS: A Fortran 77 program for determining optimum spatial declustering weights. *Computers & Geosciences*, 15(3):325–332, 1989.

[2] C. V. Deutsch and A. G. Journel. *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press, New York, 2nd edition, 1998.

[3] P. Goovaerts. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York, 1997.

[4] E. H. Isaaks and R. M. Srivastava. *An Introduction to Applied Geostatistics*. Oxford University Press, New York, 1989.

[5] A. G. Journel. Non-parametric estimation of spatial distributions. *Mathematical Geology*, 15(3):445–468, 1983.

[6] H. Omre. The variogram and its estimation. In G. Verly et al., editors, *Geostatistics for natural resources characterization*, volume 1, pages 107–125. Reidel, Dordrecht, Holland, 1984.

APPENDIX: Polygonal and Cell Declustering

The polygonal method first determines *polygons of influence* for each data location [4]. The declustering weight is taken as inversely proportional to the area of the *polygons of influence*, A_j, centered at the data location:

$$w_j^{(p)} = \frac{A_j}{\sum_{j=1}^n A_j}$$

A synthetic example will be shown for illustration. Figure 22 shows location map of 122 wells. The gray scale code shows the underlying *true* distribution of porosity (*inaccessible* in practice). Figure 23 is the equally weighted histogram of 122 well data together with the *true* reference histogram shown as the black line (*inaccessible* in practice). Since samples are not sampled regularly but more densely taken from the high value area (25 to 30%) than low value area (0 to 20%), the equally weighted histogram from the data should only be a biased estimate of the *true* distribution.

Figure 24 shows the *polygons of influence* for 122 well locations. Clustered data with small *polygons of influence* receive less weight than isolated locations with large *polygons of influence*. The weights then are used as frequencies of occurence to generate a weighted histogram, which is shown on Figure 25. Now, the histogram is much closer to the underlying *true* distribution. The weight of data with a value between 25 to 30 % has been decreased and the weight of data with a value in the 0 to 20% porosity range has been increased. The declustered mean is 19.44% which is close to the *reference* mean, 18.72%.

It has been observed that polygonal declustering technique works well when the limits (boundaries) of the volume of interest are well defined and the polygons do not change in size by more than a factor of, say, 10 (largest area/smallest area).

Another widely used technique is called *cell declustering* [1, 2]. The *Cell declustering* algorithm first divides the volume of interest into a grid of cells, $l = 1, \ldots, L$, then count the number of occupied cells L_o, $(L_o \leq L)$, and the number of data in each occupied cell $n_{l_o}, l_o = 1, \ldots, L_o$ where $\sum_{l_o=1}^{L_o} n_{l_o}$ equals the number of data n.

Each occupied cell and each datum in a occupied cell will be equally weighted, then the cell declustering weight for a datum i falling in cell l', $l' \in L_o$ is:

$$w_i^{(c)} = \frac{1}{n_{l'} \cdot L_o}$$

Figure 26 shows the data locations of the same example and the area is gridded into 36 cells. Totally there are 33 cells containing data. For a cell containing two data, each of the two clustered data values receives a weight $1/(2 \times 33) = 0.0152$.

Two key parameters of the *cell-declustering* technique are the location (origin and orientation) of the grid and the cell size [1].

The shape of the cells depends on the geometric configuration of the data, which may be adjusted to conform to major directions of preferential sampling. The origin of the cell declustering grid and the number of cells L should be chosen such that all data are included within the grid network. The cell size also influence the weights. For the two extreme cell size, *i.e.,* all data falling into one cell and each datum in its own cell, the data will receive equal weight. When there are many data and it is known that the high- or low-valued
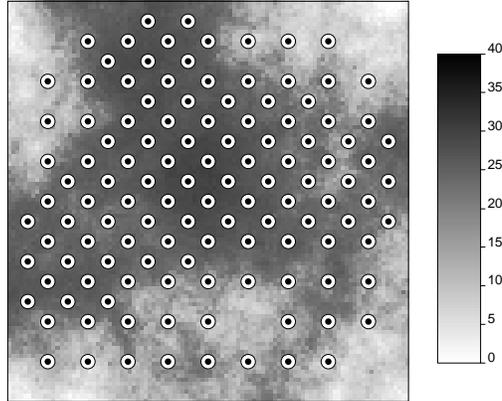
Figure 22: Location map of 122 wells. The gray scale code shows the underlying *true* distribution of porosity.

areas have been oversampled, then several cell sizes should be tried and the cell size that yields the smallest or largest declustered mean is retained according to whether the high- or low-valued areas were preferentially sampled.

Figure 27 is a plot of the declustered mean *versus* the cell size. Since the high-values areas have been oversampled, the cell size given the minimum declustered mean is chosen to generate weights for the data. Figure 28 shows the weighted histogram from *cell-declustering*. The declustered mean is 20.02%, which is much closer to the global mean of 18.72% than the equally weighted mean 22.26%. The proportion of data between 25 to 30 % has largely been corrected and the weight given to data in the 0 to 20% porosity range has been increased.

Besides the *polygonal declustering* and the *cell declustering* techniques, there are some other declustering techniques. For example, *global kriging* weights provided by a global kriging estimate can be used for declustering. However, as the *polygonal declustering*, one has to decide the range of influence for the data when using *global kriging* for decluster-ing. Furthermore, using kriging as a declustering procedure requires a variogram, but the calculation of a representative variogram may require declustering weights [6].
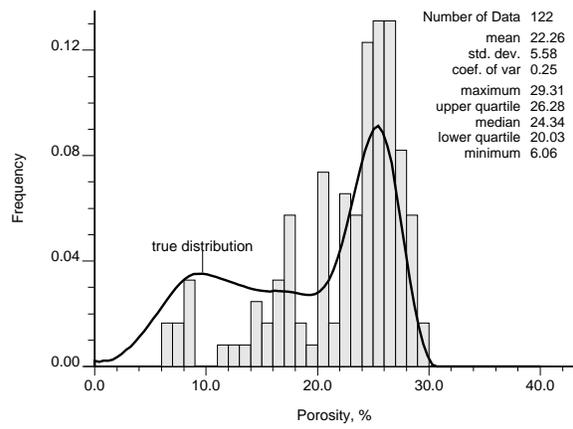
Figure 23: Histogram of porosity from 122 well data and the *true reference* histogram shown as the black line.
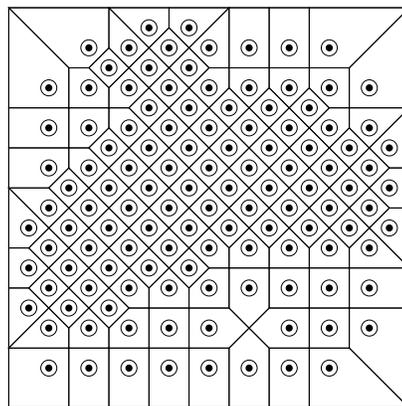

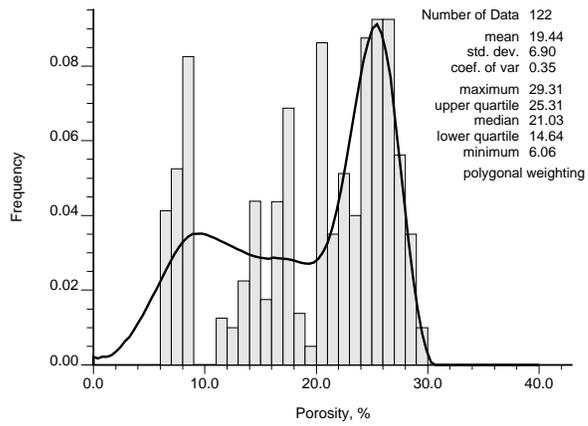Figure 24: Location map of 122 wells with polygonal areas of influence.

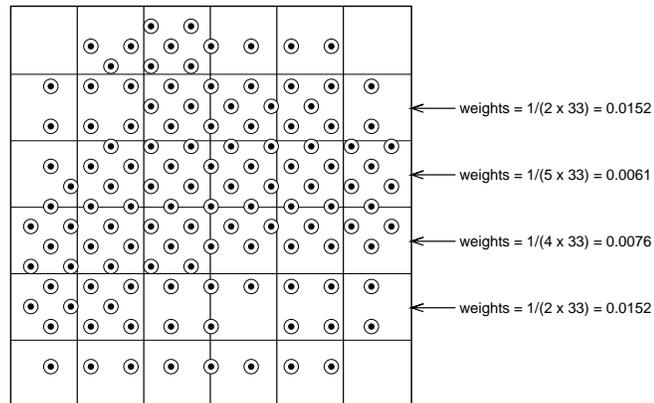Figure 25: Histogram of porosity after *polygonal declustering*.

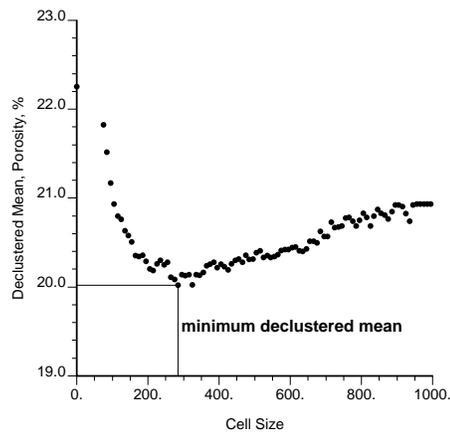Figure 26: 122 well data locations and 33 declustering cells.

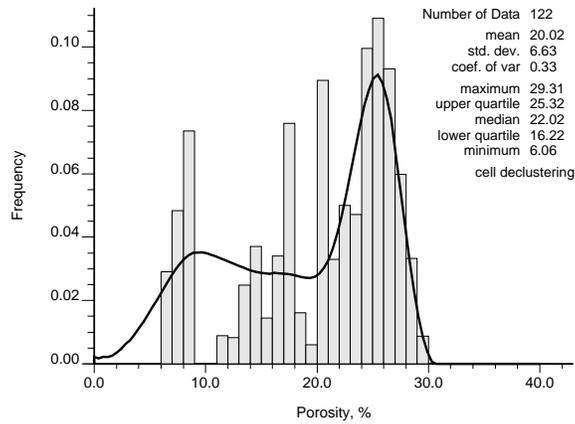Figure 27: The declustered mean versus the cell size.

Figure 28: Histogram of porosity after *cell declustering.*