

Programs for Debiasing and Cloud Transformation: bimodel and cltrans

Clayton V. Deutsch¹ and Thies Dose²

¹ Centre for Computational Geostatistics, Edmonton Alberta

² RWE Dea Aktiengesellschaft, Hamburg Germany

Debiasing is a procedure whereby a representative distribution of a petrophysical variable is derived based on a spatially exhaustive secondary variable and a calibration relationship. We often deal with continuous variables, but categorical variables could be used. The exhaustive secondary variable could be seismic or a geological map. The calibration comes from the available well data or forward geophysical modeling. This algorithm is important to create the required representative input to resource calculations and geostatistical modeling.

The Cloud Transform is a related technique. The goal is to construct a 3-D model of a petrophysical variable using a previously modeled variable and a bivariate calibration relationship. The probabilities to draw can be independent from each other or correlated in space. The technique is commonly referred to as P-field simulation if the probability values are correlated.

*Two GSLIB-like programs have been written. The first program models bivariate distributions from limited data and parametric/non-parametric conditional distributions (*bimodel*). This program also performs debiasing and soft-data declustering. The second program creates multiple realizations using the cloud transform (*cltrans*) and the bivariate distributions from *bimodel*.*

Debiasing/Modeling Bivariate Distributions

The *bimodel* program models a bivariate distribution using limited bivariate data. The conditional distributions may be fitted non-parametrically from data within bins or by fitting normal/lognormal distributions. The distributions may be constrained by user-defined quantiles. There are four outputs from the program: (1) a file containing the bivariate distribution in the format of a conditional cumulative distribution functions, (2) a file with conditional summaries such as conditional expectation, conditional standard deviation, and quantile curves for probability values of 0.05, 0.25, 0.5, 0.75, and 0.95, (3) debiased distribution of the dependent variable given a reference distribution of the independent variable and the modeled bivariate distribution, and (4) a PostScript plot showing the conditional distributions, the input paired data, and some conditional statistics.

An example parameter file for *bimodel* is shown below. Throughout these programs the X variable is the independent or secondary variable; Y is the dependent variable, for example, X could be porosity and Y could be permeability, which is to be predicted. The algorithm(s) used in the program will be explained with the parameters.

```

1 ../Examples/cluster.dat      -file with paired data (optional)
2 4 3 0                        - columns for X, Y, and weight
3 -1.0e21 1.0e21              - trimming limits
4 ../Examples/ydata.dat       -file with representative X (optional)
5 4 0                          - columns for X variable and weight
6 ../Examples/quantiles.dat   -file with conditional quantiles (optional)
7 1 2 3                        - columns for X value, ccdf, Y value
8 1.0                          - width for quantile values (in X units)
9 15 0.0 15.0 0              -X ndis, min, max, (0=arith, 1=log)
10 60 0.0 60.0 0             -Y ndis, min, max, (0=arith, 1=log)
11 1                            -type: 1 non param, 2 biv norm, 3 cond norm
12 0 4.0 12.0 4.0 12.0 0.5   -type 2 - 1 -> specify xm, xv, ym, yv, rho
13 10.0 25                    -bins: size, max data
14 0                            -increasing (1), decreasing (-1), no (0)
15 ../Examples/ex01-biv.out    -file for output bivariate distribution
16 ../Examples/ex01-cond.out   -file for output conditional statistics
17 ../Examples/ex01-ydist.out  -file for output Y distribution
18 ../Examples/ex01.ps        -file for output PostScript

```

The file containing paired data is specified in **Line 1** (standard GSLIB format). This file is not required if a bivariate (log)normal distribution is being used with user specified parameters (see **Lines 11** and **12**); otherwise, it is required. The columns for the X variable, Y variable, and declustering weight are specified on **Line 2**. The declustering weight is optional. Trimming limits are specified on **Line 3**. These limits apply to all of the input data variables.

```

4 ../Examples/ydata.dat      -file with representative X (optional)
5 4 0                        - columns for X variable and weight

```

The file containing a representative distribution of the X (independent) variable is specified in **Line 4** (standard GSLIB format). This file is not required. Debiasing will be performed if this file is specified and will *not* be performed if this file does not exist. The columns for the X variable and declustering weight are specified on **Line 5**. The declustering weight is optional.

```

6 ../Examples/quantiles.dat  -file with conditional quantiles (optional)
7 1 2 3                      - columns for X value, ccdf, Y value
8 1.0                        - width for quantile values (in X units)

```

The file containing user-defined quantiles is specified in **Line 6** (standard GSLIB format). This file is not required. The conditional distributions will be constrained to the specified quantiles if they are provided. The purpose of this is to account for qualitative constraints on the distributions. The columns for the X value, conditional cumulative distribution function, and Y values are specified on **Line 7**. For example, consider porosity as the X variable and permeability as the Y variable. An input data record/line of "0.20 1.00 1500.0" would specify that the maximum permeability for a porosity of 20% is 1500 mD. An input data record/line of "0.10 0.5 750.0" would specify that the median permeability for a porosity of 10% is 750 mD. The ordering of the data in this file is unimportant. The width of the quantile values is specified on **Line 8**. The Y quantile constraints will be applied over this width. Continuing the example, a value of 0.05 would mean that the maximum constraint of 1500 mD would be applied to the porosity range of 15% to 25%. The median constraint of 750 mD would be applied over the range 5% to 15%. The quantile constraint is applied after the conditional distributions have been fit (parametrically or non-parametrically).

```

9 15 0.0 15.0 0              -X ndis, min, max, (0=arith, 1=log)
10 60 0.0 60.0 0            -Y ndis, min, max, (0=arith, 1=log)

```

The discretization of the X and Y distributions is specified on **Lines 9** and **10**. These parameters must be passed to the cloud transform program. There are four parameters for each distribution: the number of bins/thresholds, the minimum value, the maximum value, and a flag specifying

whether or not the variable should be considered with a logarithmic scale. The spacing between the classes is equal (in arithmetic or logarithmic scale). The center of the first class (the threshold value) is $\frac{1}{2}$ of a class width into the range. Specifying “15 0 15 0” means that there will be 15 classes arithmetically spaced as 0.5, 1.5,... 14.5 with bin widths of 0.0-1.0, 1.0-2.0, ... 14.0-15.0. The thresholds are used on the Y axis (for CCDF calculation) and the bins are used on the X axis. The bins could be expanded to find data for CCDF determination (see later **Line 13**).

```
11 1                                -type: 1 non param, 2 biv norm, 3 cond norm
12 0 4.0 12.0 4.0 12.0 0.5 -type 2 - 1 -> specify xm, xv, ym, yv, rho
```

The CCDF distribution type is specified on **Line 11**. Option 1 is to fit each CCDF using data that fall into the X bin. Option 2 is to fit the entire bivariate distribution as a normal distribution. Option 3 is to fit each conditional distribution as a normal distribution. The distribution type is lognormal whenever the variable is specified to have a log scaling. **Line 12** could be used to specify the bivariate distribution parameters if Option 2 is selected. The first integer flag on **Line 12** says whether or not to get the parameters from the paired data (value of 0) or to read the values from the parameter file (value of 1). The values are the X mean, the X variance, the Y mean, the Y variance, and the correlation coefficient. These completely specify a bivariate normal, lognormal, or semi-lognormal distribution.

```
13 10.0 25                          -bins: size, max data
```

Line 13 specifies the bin width (1/2 window size) and the maximum number of data to use to establish the CCDFs if they are to be determined with options 1 or 3 (**Line 11**). The closest data to the center of the bin will be used within the expanded bin width. The data are used directly for the CCDF with type 1 and are used to get the mean and standard deviation for type 3 distributions.

```
14 0      T                          -increasing (1), decreasing (-1), no (0)
```

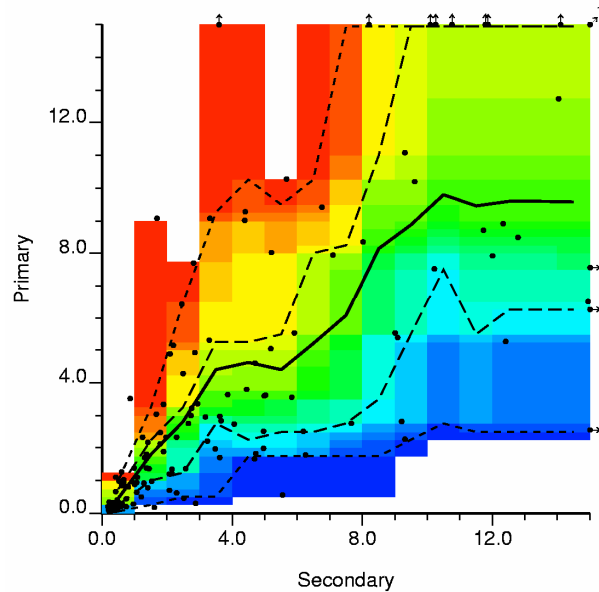
Line 14 specifies whether to enforce increasing distributions (often the case for positively correlated variables), decreasing distributions (for negatively correlated variables), or whether to leave the distributions alone – the default. The correction is similar to the order relations correction in indicator-related GSLIB programs. The distributions are corrected in an upward direction, a downward direction, and then averaged. This ensures some smoothness and consistency and avoids a saw tooth appearance to the distributions and conditional statistics.

```
15 ../Examples/ex01-biv.out          -file for output bivariate distribution
16 ../Examples/ex01-cond.out         -file for output conditional statistics
17 ../Examples/ex01-ydist.out        -file for output Y distribution
18 ../Examples/ex01.ps               -file for output PostScript
```

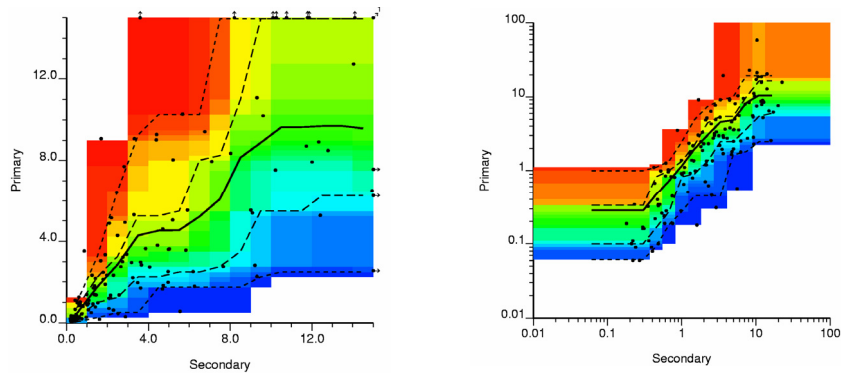
The four output files are specified on **Lines 15** through **18**. The bivariate distribution file contains $n_{xdis} \cdot n_{ydis}$ (**Lines 9** and **10**) lines of CCDF values that are needed in the cloud transform program. The conditional statistics file contains n_{xdis} lines of conditional statistics (conditional mean, conditional standard deviation, and 0.05, 0.25, 0.50, 0.75, 0.95 conditional quantile values). The output Y distribution file is for the debiased distribution corresponding to the modeled bivariate distribution and the input reference X distribution file. This file has a resolution of n_{ydis} .

The PostScript output file (see example below) shows the paired values as bullets. Data that are outside of the plotting limits are shown with arrows to indicate that the actual X/Y values are larger than shown. The number of values where both X and Y values are larger is shown in the upper right hand corner of the plot. The CCDFs are color coded from low (blue) to high (red)

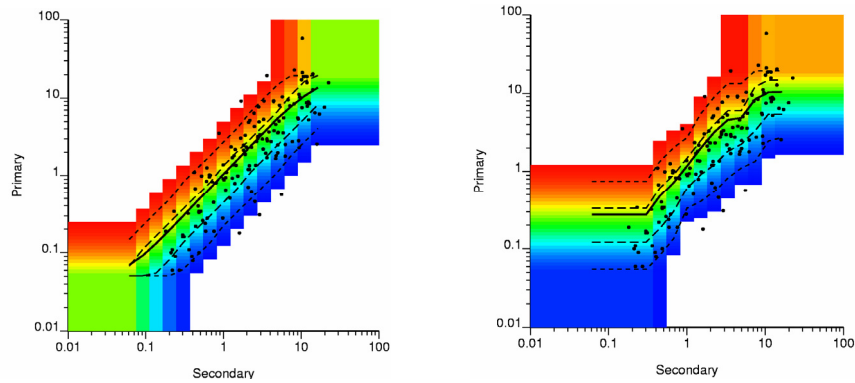
using the standard GSLIB color scale between limits of 0 and 1. Values less than 0.01 or greater than 0.99 are not shown. The conditional mean values are connected by a solid black line (normally in the middle, but may be clear of the centre in the case of skewed ccdf's). The 0.25 and 0.75 quantiles are shown by long dashed black lines. The 0.05 and 0.95 quantiles are shown by the dashed lines. The quantiles are constrained to be within the user-specified limits.



The example shown above is from the GSLIB cluster.dat dataset. There is some saw tooth variations that are unrealistic. Enforcing increasing distributions leads to the results shown below left. Specifying logarithmic scaling leads to the results shown below right.

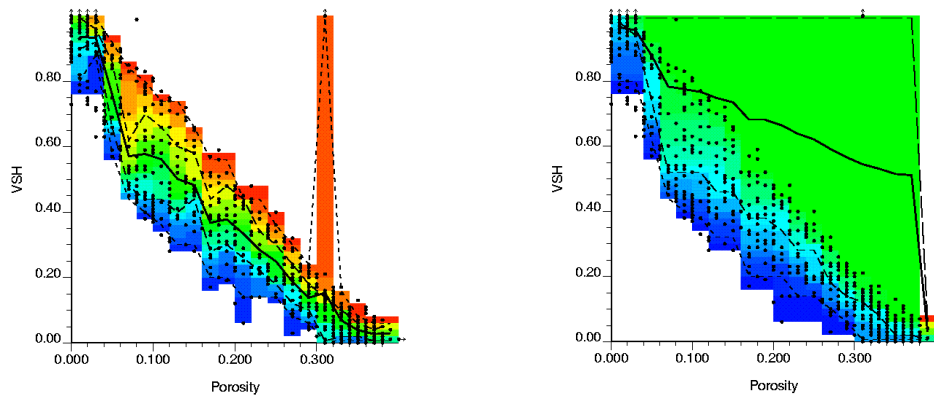


Following is the result with fitted bivariate lognormal (type 2 at right) and lognormal distributions (type 3 at left):



The color scale represents conditional cumulative distributions and not a bivariate probability distribution. These are used directly in the cloud transform.

Outliers will certainly cause a problem in the modeling of bivariate distributions. Following is an example with an outlier data that affects a limited number of conditional distributions if not corrected, but the influence is spread significantly when corrected. Care should be taken and the results always visually inspected for artifacts.



Cloud Transform

The `cltrans` program creates simulated geostatistical realizations using bivariate distributions modeled in the `bimodel` program. The probability values can be drawn independently in `cltrans` or correlated probability values (in probability or Gaussian units) obtained from `sgsim` or similar program can be read in. The program works within selected facies and/or with a keyout array by assigning all locations with missing X values with missing Y values too. There is no need for explicitly reading a keyout array. The resulting model could be additionally clipped by some other variable if required

An example parameter file for `cltrans` is shown below. This program requires a bivariate distribution file from `bimodel`. The same conventions as `bimodel` are used.

```

1 ../Examples/ex01-biv.out      -file with bivariate distribution
2 15 0.0 15.0 0              - X ndis, min, max, (0=arith, 1=log)
3 60 0.0 60.0 0              - Y ndis, min, max, (0=arith, 1=log)
4 ../Examples/ydata.dat       -file with independent X values
5 4                            - column for X variable
6 -1.0e21 1.0e21              - trimming limits
7 50 50 1                     - nx, ny, nz
8 1                            - number of realizations
9 1                            -simulation: 1 random, 2 correlated
10 69069                      - if 1: random number seed
11 ../Examples/sgsim.out       - if 2: file with probabilities
12 1                            - column for prob value
13 0                            - 0=Gaussian, 1=CDF
14 1                            -output: 0=sim values only, 1=all
15 ../Examples/ex01-cltrans.out -file for output realizations

```

The file containing the input bivariate distribution (output from `bimodel`) is specified in **Line 1**. The discretization parameters on **Lines 2** and **3** should match the parameters (see **Lines 11** and **12** in `bimodel`); otherwise, strange results will be created.

The file with the independent X values is specified on **Line 4**. This file could be a single grid of seismic in the case of simulating porosity conditional to seismic. It could be multiple realizations of porosity in the case of simulating permeability conditional to porosity. The file could contain one or many realizations. The column number is specified on **Line 5**. Trimming limits are specified on **Line 6**.

```

5 4                            - column for X variable
6 -1.0e21 1.0e21              - trimming limits
7 50 50 1                     - nx, ny, nz
8 1                            - number of realizations

```

The grid size is specified on **Line 7**. Any number of grid nodes on a structured or unstructured grid could be specified. The product of the three size parameters is used as a counter. Any number of grid nodes on a structured or unstructured grid could be specified. The number of realizations is specified on **Line 8**. If there is only one realization of X values and the number of realizations is greater than 1, the grid of X values will be reused for every realization.

```

9 1                            -simulation: 1 random, 2 correlated
10 69069                      - if 1: random number seed
11 ../Examples/sgsim.out       - if 2: file with probabilities
12 1                            - column for prob value
13 0                            - 0=Gaussian, 1=CDF
14 1                            -output: 0=sim values only, 1=all
15 ../Examples/ex01-cltrans.out -file for output realizations

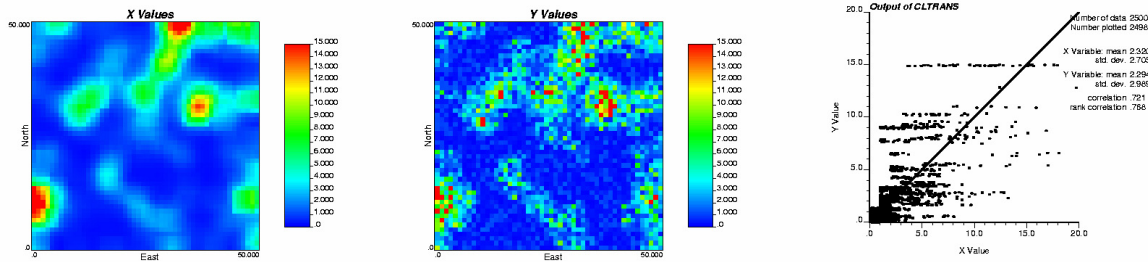
```

The type of cloud transform is specified on **Line 9**. Type 1 is random, that is, a random probability value is drawn at each location (a random number seed is required on **Line 10**). Type 2 is for correlated probability values coming from an external file. A file of probability values is specified on **Line 11**, the column number on **Line 12**, and a flag of whether the values are Gaussian or uniform on **Line 13**. Unconditional SGSIM would create Gaussian values.

The output file is specified on **Line 15**. The output option is specified on **Line 14**; option 0 means that only the Y simulated values are written to the output file; option 1 means that the input X values, the probability values, and the simulated Y values are all written to the output file. The latter option makes it easier to check the results.

Following is the results of one run. The X values are shown on the far left (they are input). The simulated values with random probability values are shown in the middle. The cross plot of the

two grids is shown at the right – this should match (approximately) the distribution created in bimodel.



Application Notes

Debiasing is often performed in presence of a clear geologic trend or a secondary variable. In such cases, there are two steps required: debiasing and modeling with a trend or secondary variable. It is not enough to apply just one technique.

There are two well documented artifacts associated with P-field simulation/cloud transform: (1) data appear as local minima and maxima, and (2) the covariance of the p-field values is non-stationary. Modelers should consider the impact of these two problems (Pyrzcz and Deutsch, 2002).

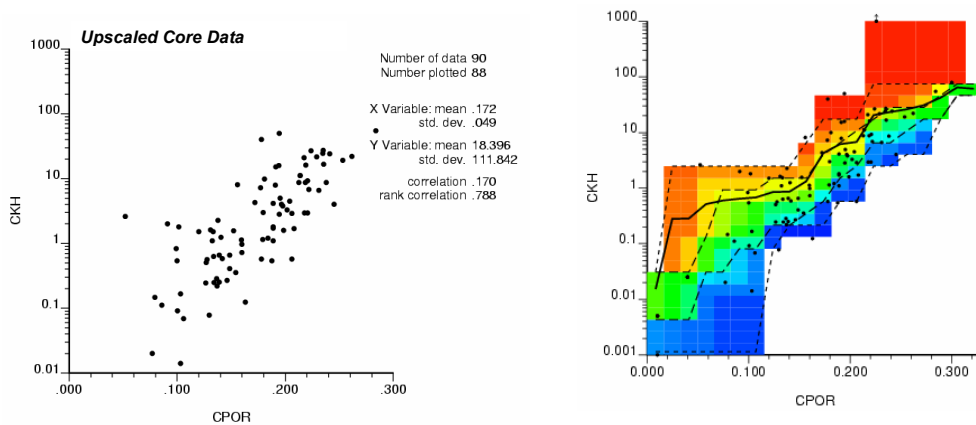
Two Practical Petroleum Examples

A common problem in modeling petroleum reservoirs is the prediction of permeability from core data. For there is no way to acquire permeability directly from wireline logs*, permeability is usually predicted from other properties (e.g. linear regression based on porosity, fuzzy logic) as a continuous curve. The downside of this is that no longer the permeability from core data itself, but the permeability predictor is modeled instead.

An attractive alternative to avoid a permeability predictor is to model the permeabilities directly from core data, eventually after sampling to a geologic grid. Of course, the situation of having fully representative core permeabilities is very rare. With the bimodel software described above, this gap can be filled easily.

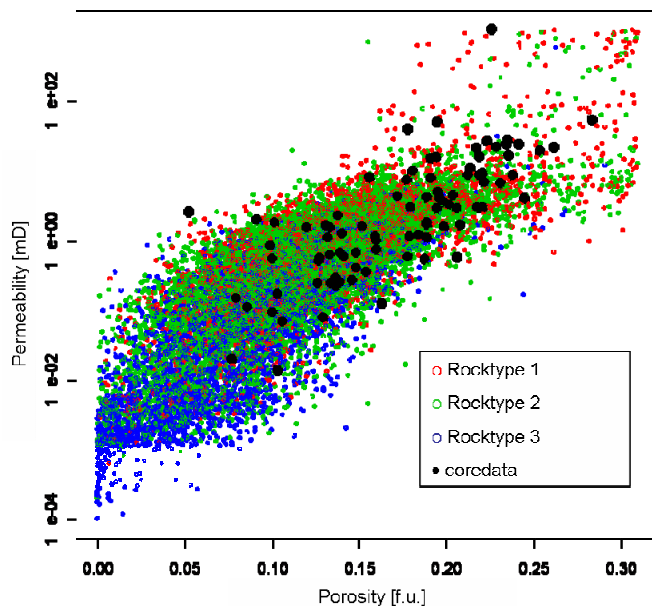
The following example shows a plot of core permeabilities versus core porosities of a gas reservoir after sampling to the geologic grid of resolution 30 x 30 x 1 m (left plot). The log porosities were matched carefully with the core samples, and are deemed to be correct within engineering accuracy. A bivariate, non-parametric relationship between porosity and core permeability was established. Because no core samples were taken for the low porosities, some artificial samples were added to provide stability for the low porosities (right plot).

* Expect for NMR logs in certain situation; however, these need to be calibrated at various representative NMR core measurements.



In this example, the scope of modeling the core permeabilities directly was not only to calibrate the bivariate relationship between porosity and permeability, but also to have the core permeabilities as conditioning data. Hence, the application of a p-field via cloud transform was not acceptable (see above). Instead, co-located cokriging was applied.

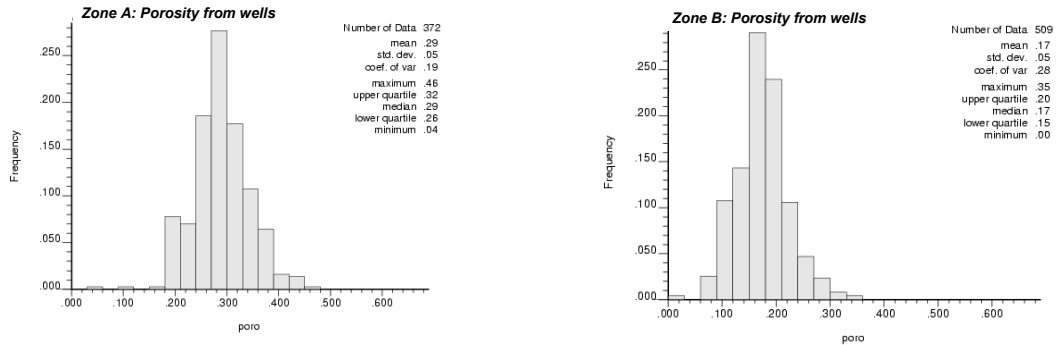
The next figure shows the combined crossplot of both core data and a realization of the modeled 3D-data. The scatter from the core data is well reproduced by the SGS/COLC algorithm. For porosities above 0.2 an outlier is present. However, this sample was confirmed as a reliable measurement by the core lab. The scatter here is reproduced quite well, a detail which would have been almost certainly smoothed out by any type of permeability prediction technique.



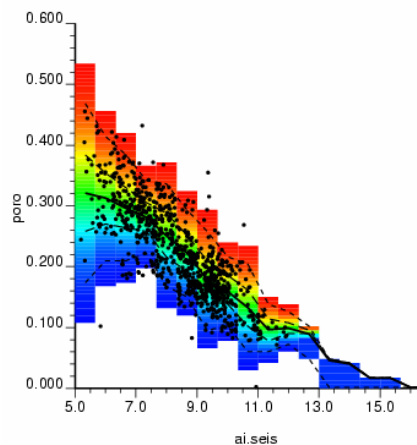
Gaussian methods to predict uncertainty need an exhaustive histogram of the variable of interest (e.g., porosity). In the case of only few and clustered wells drilled to a reservoir, errors in the

cumulative histogram based only on well data may bias the outcome substantially. One solution may be to debias porosity using a seismic attribute[†].

The following example shows the porosity histograms from two stratigraphic intervals from an carbonate exploration project, Zone “A” being more porous than Zone “B”. All of the four wells are evenly spaced and fully penetrating the reservoir, so declustering will have no effect.



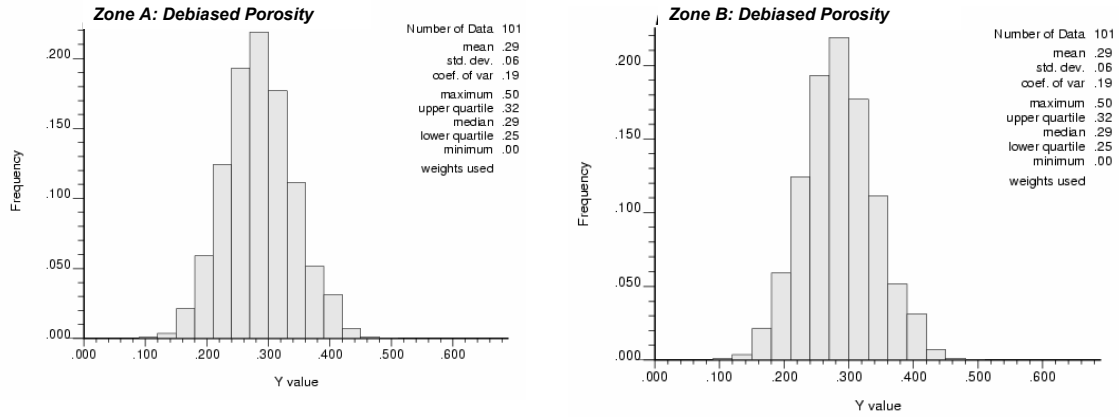
The related bimodel calibration plot shows porosities from logs and acoustic impedance from seismic inversion, sampled on a geologic grid of 150 m x 150 m x 4 m. Fluid substitution forward models proved, that there are no relevant dependencies of fluid content to the seismic signal and no tuning effects; as a consequence, this type of calibration is valid. For porosities of zero, the acoustic impedance of the clear CaCO₃ is 17 MRay (Schlumberger 2001). Because no calibration data is available for the low porosities, this information was added to the calibration via an additional quantile data file.



From this calibration plot, the porosity histograms of two geological zones were predicted (see below). For Zone “A”, debiasing does not provide any substantial changes to the well data histogram (left plot), but there is no alternative way to check this. In contrast to that, for Zone “B”

[†] In this situation it must be checked if there are any substantial artifacts caused by e.g., changes of fluid saturation, change of seismic data quality or tuning effects.

the debiasing results in a change of about 10% of the average and median porosity (right plot). This difference highlights the need of debiasing porosities for this particular reservoir.



Acknowledgements

We thank RWE Dea for funding the development of the software and the permission to publish the petroleum examples.

References

- Pyrcz, M.P., and Deutsch, C.V., "[Two Artifacts of Probability Field Simulation](#)," Math Geology, Vol. 33, No. 7, September 2001, pp 775-799.
- "Log Interpretation Charts", Schlumberger, Sugar Land, Tx, 2001, pp B5-6