A Program to Calculate Uncertainty in the Mean by Conditional Finite Domain

Martha E. Villalba and Clayton V. Deutsch

The Conditional Finite Domain (CFD) method assesses uncertainty in the mean based on the assumption of multivariate Gaussianity conditional to the available data and considering the spatial extent of the domain. A GSLIB-like program is presented based on LU Simulation. The parameters are explained in detail and examples are shown.

Introduction

The Conditional Finite Domain(CFD) technique of resampling permits quantification of uncertainty in the mean. CFD samples from rotated and translated configurations of the data in order to obtain possible mean and assemble a distribution of them. The CFD technique may be defined as an extension of the conventional bootstrap and spatial bootstrap techniques. CFD starts by creating new configurations *(J)* by random translation and rotation of the data locations relative to the centroid of the original data configuration. The Order (K) in CFD is defined as the series of simulation for each of the *(J)* configurations. The reference distribution is updated for every *order* of simulation. LU simulation is executed for each configurations *(J)* and reference distribution conditional to original data. The reference distribution is taken from the previous order of the same configuration. Simulated realizations of the *(J)* configurations will give J possible means that define the mean and uncertainty for every order, K. The uncertainty will stabilize with sufficient configurations and orders.

The number of realizations required is equal to the number of orders multiplied by the number of configurations. Each realization has *n* conditioning data and *n* locations to simulate. SGS would be relatively inefficient because it is commonly setup for a grid because of the search and covariance lookup table. The LU simulation approach is more efficient provided that 2*n* is less than 5000.

Simulation is done in Gaussian units after appropriate transformation. The normal scores transformation of the data becomes sensitive to tail extrapolation options as the order of simulation increases. Reasonable values must be chosen by the user.

Program

The code of Conditional Finite Domain follows the conventions of GSLIB, the name of the data file is asked in the first line, the location of coordinates X, Y (Z for implementing in the code), value and declustered weight is required.

```
data.dat           - input file with data
1  2  3  4  5        - columns for X, Y, Z, variable and weight
```

The trimming limits on the next line removes missing values.

```
-0.1E+04 0.1E+04  - trimming limits LTGT
```

The program will calculate the minimum and maximum tail values as small deviations from the minimum and maximum data values.

```
1                  - itail, permit calculate tails?(1=yes  0=no )
```

The next three parameters will be considered only when itail is set equal 0. These define the tail options in the standard GSLIB manner:

```
0  15.0               - data limits(tails)
1   0.0               - lower tail option, parameter
1  15.0               - upper tail option, parameter
```

The number of configurations (J) is a variable to be entered in the parameters code. The code offers the alternative to run only with translation (flag 0) or translation and rotations (flag 1).

```
100                    - nconf (number of new configuration)
1                      - permit rotation?(1=yes  0=no )
```

Translation is set respect the centroid of the original configuration data. A window around this point will limit the random translation; dimension window could be reference to the median space between data (0), domain (1) and value (2). trfrac is the fraction of the reference chosen. Sensitivity analysis shows robust CFD uncertainty in the interval [0.1-0.3]trfrac referent to domain size, equivalent to [0.2-0.6]trfrac referent to spacing data in the same example. If data is located in one sector of the domain, the suggestion is to use itrans (1) or itrans (2) with the idea of doing translation in the whole domain. Otherwise, use itrans (0) with trfrac [0.5] in domain where the sampling was done in its complete extension. trfrac * vector (referent option itrans) is the window dimension where translation respect to MPCD are randomly executed. The second line is used only when we want to do translation relative to some value and the code consider trfrac as 1, indeed itrans equal 2 give the option of entering the dimension window directly in this line.

```
 0 0.50           - itrans refer(space data=0,domain=1,value=2)and trfrac
 0.00             - if itrans = 2 dimension window value
```

The code generates an output file of configurations, uncertainty of the mean by order (k) and the means for all configurations *(J)* and orders *(K).* The file *cfdlu.out* is used to graphic the uncertainty in the mean(standard desviation) by every k order , that graphic permits to find the point of convergence of the uncertainty of the mean. Thus give a table of comparison between the different techniques of re-sampling in order to evaluate the uncertainty in the mean.

```
cfdlu_conf.out       - output file with new configuration
cfdlu_sum.out        - output file with uncertainty(conf x order)
cfdlu.out            - output file with uncertainty versus order
report.out           - output file_Summary table of uncertainty
```

The number or orders is another value that could be set and are the number of simulation that will run every configuration with the respective previous assembled reference distribution. The option of seed number is similar to traditional GSLIB simulation code.

```
100              - uncertainty order(number simulation of nconf)
112063           - random number seed
```

The variogram model, number of structure, nugget effect, sill, ranges and angles of anisotropy follow a GSLIB traditional code of simulation or interpolation.

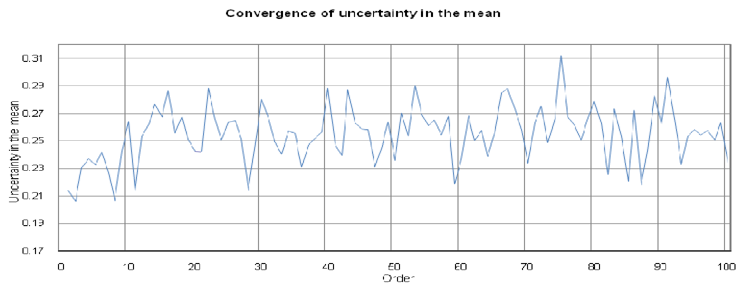
```
1   0.2                          - nst, nugget effect
1   0.8   0.0   0.0   0.0        - it,cc,ang1,ang2,ang3
  50.0  50.0  50.0               - a_hmax, a_hmin, a_vert
```

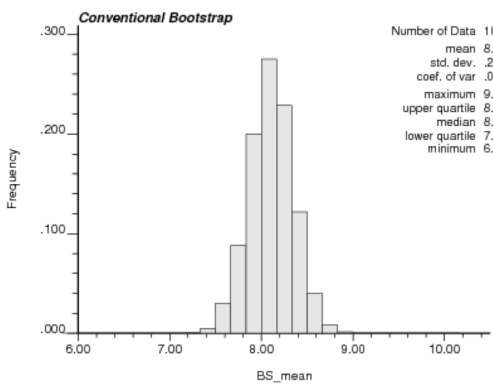
The code added routines of conventional bootstrap and spatial bootstrap. Therefore an additional output files without any specification in the parameter file of the code is generated. *Boot_avg.out* - the mean and standard deviation in *(J * K)* simulations using CB technique, *Spatial_bootstrap.out* - the mean and standard deviation in*(J * K)* simulations using SB technique, and *space_data.out* - information of the size domain and median space between data

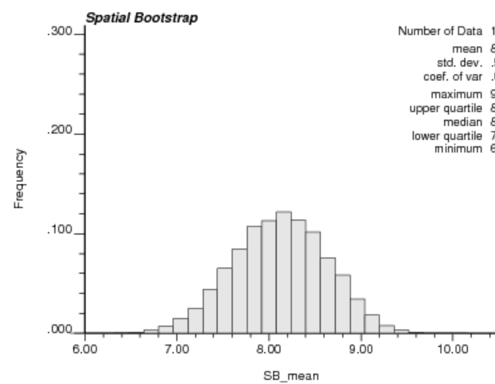
Application

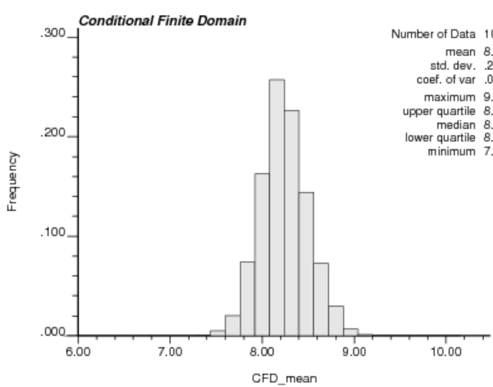
The validation of the code was done through a number of examples. Uncertainty in the mean has been evaluated for the value of porosity from Amoco data, gold in red data and values obtained as random function in the case of synthetic data. Variograms were modeled in every scenario and relevant tail extrapolation values were chosen.

The Amoco data has 62 wells with information of porosity located in two dimensions. The median distance between the samples is 1095 feet. The option of only translation and 100 orders were set to evaluate the uncertainty. The convergence of the uncertainty in the mean (standard deviation) is observed between an order of 20 and 40 order as a value of 0.25, see the chart below. This is higher than the bootstrap and lower than the spatial bootstrap, see histograms below.









The gold grade in the well used Red data were considered. The average spacing between samples is 28.43 meters. The option of translation was set because of anisotropy. Evaluation of uncertainty was done in 100 orders. Convergence of the uncertainty was observed between an order of 20 and 30 at a value of 0.13 , which is less than both the bootstrap and spatial bootstrap. The figure below shows The

Synthetic data were generated and used with both rotation and translation to see that the uncertainty was predicted in a reasonable manner.

Conclusions

A CFD program has been developed to obtain an easy and fast evaluation of uncertainty through the technique of conditional finite domain and the parameters have been explained in detail in order to avoid any mistake during the implementation of this code. A number of examples demonstrated how the program works. Care must be taken in presence of strong anisotropy: translation will lead to large uncertainty. The translation must consider the spacing of the data and size of domain. The CFD uncertainty will increase as translation increases. Translation about the spacing between the samples appears to work well

References

Babak, O., Deutsch C.V., *Reserves Uncertainty Calculation Accounting for Parameter Uncertainty*, Journal of Canadian Petroleum Technology, Vol. 47, No 8, pp. 41-49, 2008

Babak, O., Deutsch C.V. *Accounting for Parameter Uncertainty in Reservoir Uncertainty Assessment: Conditional Finite Domain Approach*, Natural Resources Research ,18(1), pp. 7-17, March 2009

Deutsch, C.V., , *Geostatistical Reservoir Modeling*, pages 65-70, Oxford University Press, New York, 2002

Deutsch, C.V., *A Statistical Resampling Program for Correlated Data, Spatial_Boostrap*, Department of Civil & Environmental Engineering, University of Alberta, September 2004.

Deutsch C.V., Journel A.G. *GSLIB: Geostatistical Software Library: and User's*

Guide Oxford University Press, New York, 2nd Ed, 1998

Leuangthong O. , Khan K.D., Deutsch C.V., *Solved problems in Geostatistics,* ,pp.47-57, 2008

Weiss E., *Bootstrapping With SAS,* The Winchendon Group Inc., 3907 Lakota Road, Alexandria, Virginia