Paper 410, CCG Annual Report 12, 2010 (© 2010)

An Update to the WA_Declus Program

Brandon Wilde and Clayton V. Deutsch

A program for performing a wide array of declustering techniques was provided in the 2007 CCG report. A number of modifications and improvements have been made to this program. The program is now more robust and provides improved output format. The modifications to the program are explained. Example of the improved output is provided.

1 Introduction

Declustering has proven to be a useful technique for obtaining the true univariate distribution of a spatial variable (Goovaerts, 1997; Deutsch, 2002). Declustering is performed to correct the univariate distribution when it has been preferentially sampled. A variable is said to have been preferentially sampled when the data locations are neither regularly nor randomly distributed over the study area. Declustering is a general term applied to a variety of techniques which weight the data by considering clustering. Data in densely sampled areas receive less weight than data in sparsely sampled areas.

Consider the porosity data configuration shown in Figure 1 right and the true reference histogram shown as the black line in Figure 1 left. The true porosity distribution has been sampled at 122 locations leading to the sample histogram shown as bars in Figure 1 left. The range of porosity values between 25% and 30% has been oversampled while the 0 to 20% range has been under sampled. The goal of declustering is to weight the data such that the true distribution is approximated. The data are weighted by considering their proximity to other data. Data close to other data will receive less weight than data far from other data. There are a number of techniques for determining these weights. Weighting the porosity data by declustering results in the distribution shown in Figure 2. The under sampled low values are given more weight while the oversampled high values are given less weight leading to a better approximation of the true distribution by the sample distribution.

A program for performing a number of declustering techniques was prepared previously (Wilde and Deutsch, 2007). A number of improvements have been made to this program. This short note documents these improvements.

2 Polygonal Declustering

The polygonal method of declustering determines the area of influence of each datum location $\mathbf{u}_\alpha$. This is the area (or volume in three dimensions) constituted by all locations $\mathbf{u} \in A$ closer to $\mathbf{u}_\alpha$ than any other datum location. The area or volume of the polygon centered at $\mathbf{u}_\alpha$ is used as a declustering weight for the datum value $z(\mathbf{u}_\alpha)$. Determining the area of influence in two dimensions is straightforward (see the GSLIB-style program `polydec`). The result is termed the Voronoi diagram in mathematics. Determining volumes of influence in three dimensions is a substantially more difficult problem. A brute-force method for performing polygonal declustering in two or three dimensions is implemented in the `wa_declus` program. It involves discretizing the domain, *A*, by a dense grid. The declustering weight for datum value $z(\mathbf{u}_\alpha)$ is then proportional to the number of grid nodes closer to datum location $\mathbf{u}_\alpha$ than any other datum location.

There are two primary practical considerations for the application of this method of polygonal declustering: boundary control and dense grid specification. Weights determined by polygonal declustering are sensitive to the specified boundary. The choice of boundary affects the weights of those data whose locations are near the edge of the study area. The previous implementation of this method in the wa_declus program provided two options for specifying the boundary. One was a distance limit, where a node is only assigned to a data location if it is within a specified distance tolerance, while the other was an expanded cube where the grid extents were determined by expanding the cube containing the data by some factor. This latter boundary control option has been removed and replaced with the option of a clipping grid. This clipping grid is a binary grid such that those nodes which fall within the volume of interest are given a value of 1 while whose nodes outside the volume have a value of 0. This grid must be created prior to declustering and is accepted as input when this option is selected.

The dense grid specification has also been modified. The previous implementation determined the smallest spacing between two data locations and used this distance as the basis for creating a dense grid. The possibility exists for a large domain to have two data locations spaced closely together causing the discretization of the volume to have tens or even hundreds of millions of nodes. This computational expense is not practical warranting the explicit control of the volume discretization now present in the program. A grid is specified in the program which is used for inverse distance, ordinary kriging, and global kriging declustering. This same grid is used in polygonal declustering. As this grid may not provide the discretization required for good polygonal declustering results, there is an option to discretize the grid further providing an appropriate number of nodes for the declustering. This denser grid is used only for polygonal declustering.

3 Inverse Distance Declustering

Various estimation techniques double as a method for declustering. This is accomplished by performing estimation and retaining a cumulative sum of the weights each data location receives. These cumulative weight sums can be used as declustering weights as those data located in a densely sampled area will not receive as much weight as those data located in sparsely sampled areas. The previous implementation of `wa_declus` included the option of performing ordinary kriging and/or global ordinary kriging estimations as declustering techniques. The additional option of using inverse distance estimation as a declustering technique has been added to the program. The same grid that is used for ordinary kriging and global ordinary kriging is used.

4 Search Parameters

Performing ordinary kriging and inverse distance estimation requires specification of a search neighborhood as well as limits on the minimum and maximum number of data for performing estimation. These search parameters have an impact on the declustering weights. The previous implementation performed ordinary kriging declustering for a number of different maximum numbers of data. This allowed the impact of varying search parameters to be understood. This functionality has been removed from the current implementation of `wa_declus`. This is done for both computational efficiency and improved aesthetics of the output results. Rerunning the kriging for different numbers of data can be computationally expensive. Also, displaying the results is difficult. A semi-transparent gray square was used to indicate the range of declustered means achieved using different numbers of data, but this square was found to be unclear and confusing. The current implementation allows for one minimum and maximum number of data to be specified. These values are applied to both the ordinary kriging and inverse distance declustering methods. To determine the effect of the search parameters on declustering, the user can run `wa_declus` for as many search specifications as deemed necessary.

5 Parameter File

The parameter file contains a number of subsections, each applying to different declustering methods. A number of the modifications discussed required changes to be made to the parameter file. As such, the parameter file has been rearranged to follow a more logical flow. The current parameter file is shown in Figure 3. The section of parameters for polygonal declustering has been moved and now falls between the grid parameters and the search parameters. The inverse distance parameters have been added before the variogram parameters. Within the section of parameters for the grid, the file name for a clipping grid has been added as well as the column for the clipping information. If this file does not exist, no clipping is performed for any of the estimation declustering methods. If this file does not exist and the clipping grid has been set as the boundary control for polygonal declustering, an error occurs. The polygonal declustering parameters have been updated to read in a grid discretization as well as incorporation of the new boundary control options previously discussed.

6 Postscript Output

The `wa_declus` program generates a visual display of the declustering results. The display is a modification of the declustered mean vs. cell size plot commonly used when performing cell declustering. The postscript output has been updated for greater clarity as shown in Figure 4. Declustering results for the estimation methods and polygonal declustering consist of one declustered mean as opposed to multiple declustered means from cell declustering. The declustered mean for these methods is shown on the plot as a horizontal line. Each line can be labeled along the right side of the plot as to the declustering method it referred to. These labels can be removed as different colors are used to identify each line. The color of the line relates to the color of the text in the summary table below the mean vs. cell size plot. This modification can create a more aesthetic plot.

7 Conclusion

Declustering is an important technique for obtaining the true univariate distribution of a random variable when it has been preferentially sampled. It is of particular importance when geostatistical simulation will be performed as it is an important input parameter of a multivariate distribution. A number of techniques have proven useful for determining weights for data locations based on the representivity of that location. The program `wa_declus` has been created to implement multiple techniques simultaneously. A number of modifications were necessary to improve the ease of use of the program and generate robust results for large areas with many data; these modifications have created a more stable program with improved flow in the parameter file and more intuitive outputs.

References

Deutsch CV, 2002. Geostatistical Reservoir Modeling. Oxford University Press, New York, 376p.

Goovaerts P, 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press. 496p.

Wilde BJ and Deutsch CV, 2007. Wide array declustering for representative distributions (the ultimate declus program). Centre for Computational Geostatistics 9th Annual Report, University of Alberta.

Figures

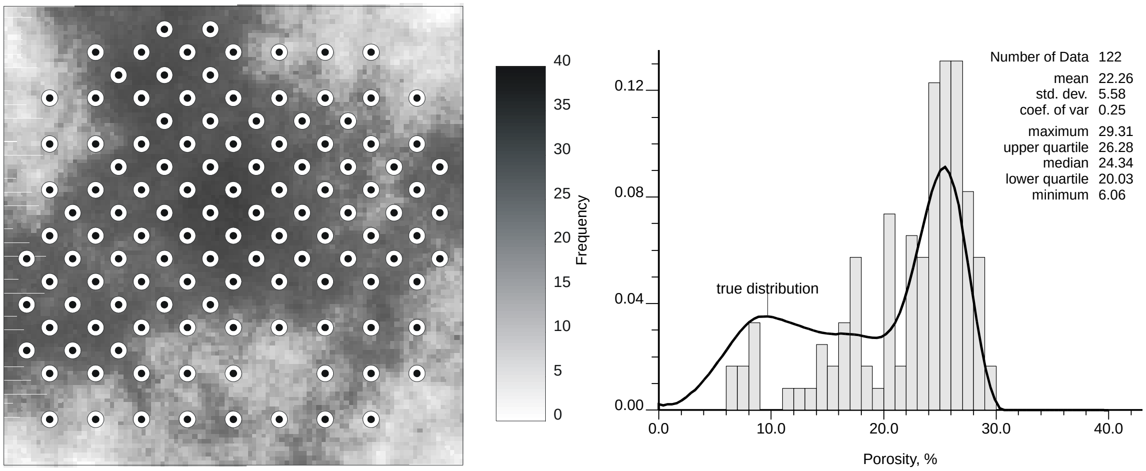


Figure 1: Configuration of porosity data and the sample (bars) and true (line) histograms.

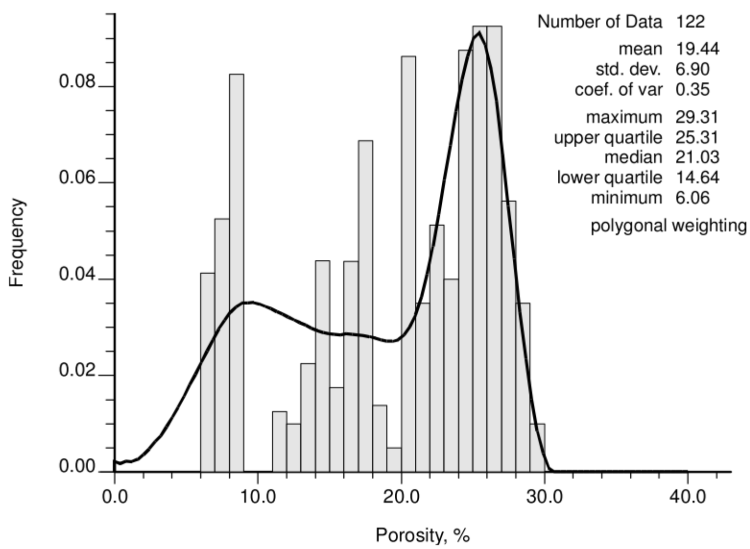
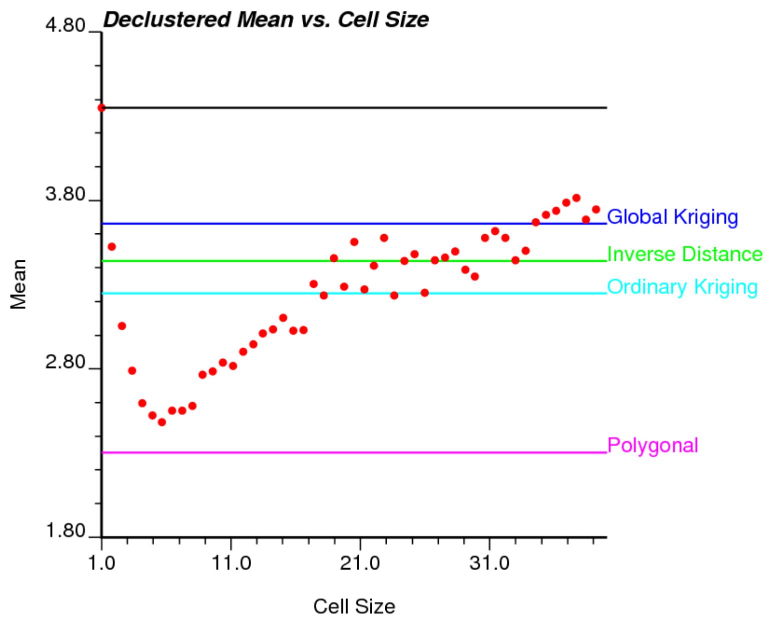


Figure 2: Declustered porosity histogram.

```
                  Parameters for WA_DECLUS
                  ************************

START OF PARAMETERS:

Common Parameters:
cluster.dat                       -file with data
0 1 2 0 3                         -   columns for DH,X,Y,Z,var
-1.0e21   1.0e21                  -trimming limits
1  1  1  1  1                     -Perform: Polygonal?,Cell?,OK?,ID?,Global OK?
0                                 -0=look for minimum declustered mean (1=max)
1                                 -plot labels? (1=yes)
wa_declus.res                     -results file
wa_declus.out                     -file for data with weights output
wa_declus.ps                      -file for histogram/scatterplot output

Parameters for Cell Declustering:
1.0   .1                          -Y and Z cell anisotropy (Ysize=size*Yanis)
50   1.0   40.0                   -number of cell sizes, min size, max size
5                                 -number of origin offsets

Grid Specification for Polygonal, ID, OK, and GK:
104  .25  .5                      -nx,xmn,xsiz
104  .25  .5                      -ny,ymn,ysiz
1    0.0 2.0                      -nz,zmn,zsiz
clipgrid.in                       -file for clipping grid
1                                 -column with clipping info

Parameters for Polygonal Declustering
1                                 -grid discretization
2                                 -boundary control (1=clipping grid, 2=distance limit)
40                                -if=2, distance limit

Data and Search Parameters for ID and OK
4   8                             -min and max data to use
2000.0  2000.0  20.0              -maximum search radii
 0.0   0.0   0.0                  -angles for search ellipsoid

Parameters for Inverse Distance
1                                 -variogram structure with applicable anisotropy
1                                 -small constant in denominator
0.5                               -exponent in denominator

Variogram Parameters for OK and GK:
1    0.2                          -nst, nugget effect
1    0.8  0.0   0.0   0.0         -it,cc,ang1,ang2,ang3
         1000.0  1000.0  10.0         -a_hmax, a_hmin, a_vert
```

Figure 3: Parameter file for current implementation of wa_declus.



Method	*Mean*	*% Change*
Equal Weighted	4.35	—
Cell	2.49	-42.88
Polygonal	2.30	-47.08
Inverse Distance	3.44	-20.88
Ordinary Kriging	3.25	-25.38
Global Kriging	3.66	-15.79

Figure 4: Updated postscript display applied to the cluster.dat file.