

Developments in Semiautomatic Variogram Fitting

Chad Neufeld and Clayton V. Deutsch

Centre for Computational Geostatistics
Department of Civil and Environmental Engineering
University of Alberta

Abstract

The varfit program developed by CCG has received widespread attention because it greatly simplifies variogram fitting, particularly in 3D and with multiple variograms making up a linear model of coregionalization. A number of enhancements have been made: (1) a new gamv output format is considered as input, which simplifies the input parameters and makes it straightforward to input directional, indicator, and cross variograms, (2) any number of directions can be optimized simultaneously, not just the three principal directions, (3) the angles can be optimized independently, optimized and fixed between structures, or frozen at user input values, and (4) indicator direct and cross variograms can be fit with constraints on reasonable variations between thresholds.

Introduction

Fitting variogram models is an integral part of the geological modelling process. However, fitting experimental variograms can take a large proportion of time during the building of numeric geological models. This is particularly true if there are a significant number of facies and/or variables to model. The original varfit program received a great deal of attention because it simplifies the variogram fitting process.

During the previous year a number of enhancements have been made to varfit. These enhancements were aimed at making varfit easier to use and to reduce the number of errors that occur. Some of the changes are that the experimental variogram information is passed to varfit using a new gamv output format and indicator variograms can now be fit with constraints on parameter variation between thresholds.

The new gamv2004 Program

The limitation of the gam/gamv formats in GSLIB is that little information is passed regarding the parameters used to calculate the variogram. This requires the user to keep those parameters separately and pass them into other programs such as varfit. This is inefficient and error-prone. The same basic format will be retained, but now there will be the possibility of additional lines of input parameters. The input parameter file has been modified to the following:

```

START OF PARAMETERS:
testing/cluster.dat      -file with data
1  2  0                  - columns for X, Y, Z coordinates
2  3  4                  - number of variables, column numbers
-1.0e21    1.0e21       - trimming limits
testing/gamv.out        -file for variogram output
3                        -number of directions
  0. 90. 9999.    0. 90. 9999. -Dir 01: azm,atol,bandh,dip,dtol,bandv
10 5.0 3.0        - nlag,xlag,xtol
  0. 20. 9999.    0. 90. 9999. -Dir 02: azm,atol,bandh,dip,dtol,bandv
10 5.0 3.0        - nlag,xlag,xtol
  90. 20. 9999.   0. 90. 9999. -Dir 03: azm,atol,bandh,dip,dtol,bandv
10 5.0 3.0        - nlag,xlag,xtol
1                        -standardize sills? (0=no, 1=yes)
2                        -number of variograms
1  1  1              -tail var., head var., variogram type
1  1 -3              -tail var., head var., variogram type

type 1 = traditional semivariogram
      2 = traditional cross semivariogram
      3 = covariance (-3 calculates variance-covariance)
      4 = correlogram (-4 calculates 1-correlation)
      5 = general relative semivariogram
      6 = pairwise relative semivariogram
      7 = semivariogram of logarithms
      8 = semimadogram
      9 = indicator semivariogram - continuous - requires a cutoff
     10= indicator semivariogram - categorical - requires a category
     11= indicator cross semivariogram - cont. - requires two cutoffs
     12= indicator cross semivariogram - cat. - requires two categories

```

Note that the number of lags, the lag spacing, and the lag tolerance must now be specified for each direction. This permits the simultaneous calculation of horizontal and vertical directional variograms. It also permits the more reliable specification of directional variograms when the domain is anisotropic. The basic calculation scheme within the program remains the same.

The other major change is to the output format. Input data parameters are saved with the calculated variogram points. They are saved on four input lines after each title line. Each line begins with a four character index.

- **HDIR** – horizontal direction parameters: the azimuth, the azimuth tolerance and the bandwidth perpendicular to the horizontal direction.
- **VDIR** – vertical direction parameters: the dip, dip tolerance, and the bandwidth perpendicular to the vertical direction.
- **LAGS** – lag parameters: number of lags, the lag distance, and the lag tolerance. Note that these are potentially different in each direction.
- **VARI** – variogram parameters: the variogram type, the variable number for the tail, the variable number for the head and the categories or cutoffs (for indicator and indicator variograms). For indicator cross variograms two categories or cutoffs are required.

```

-HDIR  -30  22.5  25.0  -azm, azm tol, azm bandw
-VDIR  -90  22.5  25.0  -dip, dip tol, dip bandw
-LAGS   10  15.0   7.5  -# lags, lag dis, lag tol
-VARI    9   1   1  0.73 -var type, tail var, head var, ind cat/cut

```

The `vargplt2004` program recognizes this format and removes any lines with a recognized four character index. The new `varfit` programs require this format to get information about the experimental variograms.

Updates to `varfit`

The `varfit` program has received several updates over the past year. Two major changes were made to `varfit`. The first change was to split `varfit` into three specialized programs: (1) `varfit` for single variable variograms, (2) `varfit_lmc` for fitting linear models of coregionalization, and (3) `varfit_ind` for fitting indicator variograms. The second change was to require the new `gamv2004` format for all of the `varfit` programs. Each specialized program has received specific enhancements discussed below.

Single Variable Variogram (`varfit`)

There have been two important changes made to `varfit` in addition to the changes mentioned above. Any number of variograms in multiple directions can be used for fitting the variogram model and the angles can be frozen or allowed to vary for each nested structure as the optimization sees fit.

The variogram specification portion of the parameter file has also been simplified. The user does not need to specify the variogram parameters since `varfit` reads the settings for each experimental variogram from the `gamv2004` output file. This will help to reduce the number of errors that occur. The old and new parameter file lines are below:

OLD VARIOGRAM SPECIFICATION:

```
2 2 1 2          -ivar1, ivar2, idir, ivarn
gamv-hori.out    -   variogram file
```

NEW VARIOGRAM SPECIFICATION:

```
gamv-hori.out    -variogram #1 file
2                -   variogram number in file
```

Linear Model of Coregionalization (`varfit_lmc`)

Only one significant change was made to `varfit_lmc`. This is the need to specify the coordinate system rotation angles at the start of the parameter file. The azimuth and dip of the input experimental variograms are used to determine which principal direction the variogram was calculated along. The additional parameter file line is:

```
START OF MAIN PARAMETERS:
0.0 0.0 0.0          -ang1, ang2, and ang3 (GSLIB definition)
```

Unlike `varfit`, `varfit_lmc` can only use experimental variograms calculated along the three principal directions. If an input experimental variogram does not correspond to one of the three principal directions an error is reported.

Indicator Variogram (`varfit_ind`)

Indicator variograms require additional constraints than traditional variograms. There are a number of parameters “ p ” that should be consistent from one cutoff to the next. For example, some parameters are (1) relative nugget effect, (2) sill contributions for the number of structures, and (3) one or more ranges. A measure of smoothness for each parameter has been added to the objective function.

Consider the parameters $p_i, i=1, \dots, n_c$ where n_c is the number of cutoffs or thresholds. The measure of consistency implemented that still allows reasonable discontinuities is:

$$smoothness = \sum_{j=2}^{n_c-1} \left[\frac{p_{i-1} + p_{i+1}}{2} - p_i \right]^2$$

This measure of consistency requires a weighting during the optimization process. A fixed weight did not work well. For this reason, the user is required to input an appropriate weighting for the components of the objective function. The three components are: (1) the fit to the experimental variogram points, (3) the transition of the structure sill contributions across the thresholds, and (3) the transition of the structure ranges across the different thresholds. The weights are specified as a scaling factor with 1 being the standard. The additional parameter file line is:

```
1 1 1 -fitting preferences: points, cc and range transitions
```

All of the variogram parameters are written to an output file that is formatted for checking the transitions between thresholds. This file can be easily imported into Excel, or similar program, for plotting. Adjustments can be made to the above weights based on the smoothness of these plots and the fit of the modelled variogram to the experimental points. The additional output file name is an additional parameter in the input file:

```
varfit_ind_thresholds.out -file for threshold information
```

In addition, the number and type of structures are forced to be the same for all indicator direct and cross variograms to make the models easier to interpret.

Examples

Two short examples are presented here to show the enhancements made to the `varfit` programs. The first example is a horizontal anisotropic variogram and the second is a multiple threshold indicator variogram.

Horizontal Anisotropic Variogram

This example was done to test how `varfit` handles a large number of experimental variograms in any direction. The data set is two dimensional and has been normal score transformed. The experimental variograms were calculated horizontally in 10° azimuth increments from 0° to 360° . This resulted in 37 variograms as input to `varfit`.

The optimization process took less than 30 seconds using all of the experimental variograms. **Figure 1** and **Figure 2** show 6 of the experimental variograms and the resulting variogram model. The fit for the remaining 31 variograms was as good as the fit in Figures 1 and 2. The final variogram model was:

$$\gamma(\mathbf{h}) = 0.15 + 0.01 \cdot \Gamma^1(\mathbf{h}) + 0.75 \cdot \Gamma^2(\mathbf{h})$$

where $\Gamma^1(\mathbf{h})$ is an exponential structure with a maximum range of 11.4m, and a minimum range of 7.3m, and $\Gamma^2(\mathbf{h})$ is a spherical structure with a maximum range of 228.0m, and a minimum range of 113.4m.

Multiple Indicator Threshold Variogram

This example was done to ensure that `varfit_ind` produces good fitting variogram models that have consistent parameters between thresholds. We used a 3D variogram with 10 thresholds to test the variogram fitting and parameter consistency in `varfit_ind`.

All 10 thresholds were fit simultaneously. **Figure 3** and **Figure 4** show the experimental and modelled variograms for thresholds #2 and #8 respectively. **Figure 5** and **Figure 6** show how the parameters change between thresholds. The parameter transitions between thresholds are smooth. This should help alleviate order relations problems. The variogram model for all of the thresholds has the following format:

$$\gamma(\mathbf{h}) = C0 + C1 \cdot \text{sph}_{\substack{hmax1 \\ hmin1 \\ vert1}}(\mathbf{h}) + C2 \cdot \text{sph}_{\substack{hmax2 \\ hmin2 \\ vert2}}(\mathbf{h}) + C3 \cdot \text{exp}_{\substack{hmax3 \\ hmin3 \\ vert3}}(\mathbf{h})$$

where $C0$ is the nugget effect, $C[1-3]$ are the variance contributions for three nested structures, and $hmax[1-3]$, $hmin[1-3]$, and $vert[1-3]$ are the ranges for the nested structures. **Table 1**, **Table 2**, and **Table 3** list the variogram parameters.

Table 1: Modelled indicator variogram variance contributions.

Threshold	$C0$	$C1$	$C2$	$C3$
1	0.23	0.26	0.24	0.28
2	0.30	0.24	0.28	0.17
3	0.35	0.21	0.33	0.12
4	0.38	0.15	0.37	0.10
5	0.38	0.11	0.43	0.08
6	0.38	0.09	0.48	0.05
7	0.37	0.10	0.51	0.02
8	0.37	0.13	0.49	0.00
9	0.38	0.18	0.44	0.00
10	0.40	0.22	0.38	0.00

Table 2: Ranges for the first and second nested structures.

Threshold	$hmax1$	$hmin1$	$vert1$	$hmax2$	$hmin2$	$vert2$
1	26.1	121.1	0.0	220.8	187.7	22.5
2	22.1	97.6	12.2	201.1	155.6	35.3
3	18.3	77.4	24.5	180.1	127.1	48.9
4	14.2	61.2	37.3	155.1	102.8	62.9
5	11.3	46.6	48.7	126.3	82.5	76.9
6	10.1	34.0	55.6	98.4	66.6	88.5
7	11.6	23.8	58.2	74.1	54.8	99.1
8	16.5	16.5	57.3	54.9	46.5	108.1
9	22.5	12.2	53.6	39.7	39.7	117.0
10	29.0	8.8	49.2	29.0	32.1	99999

Table 3: Ranges for the third nested structure.

Threshold	<i>hmax3</i>	<i>hmin3</i>	<i>vert3</i>
1	99999	199.1	99999
2	99999	175.7	99999
3	244.7	154.3	99999
4	199.4	135.9	99999
5	155.5	122.8	94.5
6	116.4	116.4	110.0
7	88.2	107.2	125.8
8	65.5	98.9	141.8
9	46.5	90.0	157.9
10	29.1	81.1	99999

Conclusions

The original `varfit` program was widely accepted because it simplified the variogram modeling process. The enhancements made to `varfit` make it more powerful and easier to use. All of the information for the experimental variograms is passed to `varfit` by using the new `gamv` output format. This helps reduce the number of errors that occur. Any number of directions can now be used to model single variable variograms, not just the three principal directions. Multiple threshold indicator variograms can now be fit with logical constraints on the parameter consistency between thresholds.

Acknowledgements

The authors would like to thank the Centre for Computational Geostatistics and the Natural Sciences and Engineering Research Council of Canada for their support.

References

- [1] Deutsch, C.V. and Journel, A.G., 1998. *GSLIB: Geostatistical Software Library and User's Guide*, 2nd edn. Oxford University Press, New York, 369 pp.
- [2] Ellis, T.M.R., Philips, I.R. and Lahey T.M., 1994. *Fortran 90 Programming*. Addison-Wesley, New York, 825 pp.
- [3] Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York, 496 pp.
- [4] Gringarten, E. and Deutsch, C.V., 2001. Teacher's aide: Variogram interpretation and modeling. *Mathematical Geology*, 33(4):507–534.
- [5] Harris, J.W. and Stocker, H., 1998. *Handbook of Mathematics and Computational Science*. Springer-Verlag New York, Inc., New York, 1056 pp.
- [6] Isaaks, E.H. and Srivastava, R.M., 1989. *An Introduction to Applied Geostatistics*. Oxford University Press, New York, 561 pp.
- [7] Johnson, R.A. and Wichern, D.W., 1998. *Applied Multivariate Statistical Analysis*. Prentice-Hall, Inc., New Jersey, 750 pp.

- [8] Journel, A.G. and Huijbregts, Ch.J., 1978. *Mining Geostatistics*. Academic Press, New York, 600 pp.
- [9] Neufeld, C., Larrondo, P., and Deutsch, C.V., 2004. VARFIT: A Program for Semi-Automatic Variogram Modelling. *Submitted to Computers & Geosciences*.
- [10] Reklaitis, G.V., Ravindran, A. and Ragsdell, K.M., 1983. *Engineering Optimization: Methods and Applications*. John Wiley & Sons, Inc., New York, 704 pp.
- [11] Isaaks, E.H., 2001. *Sage 2001 – The Most Powerful Windows Variography Software on the Planet*. <http://www.isaaks.com>

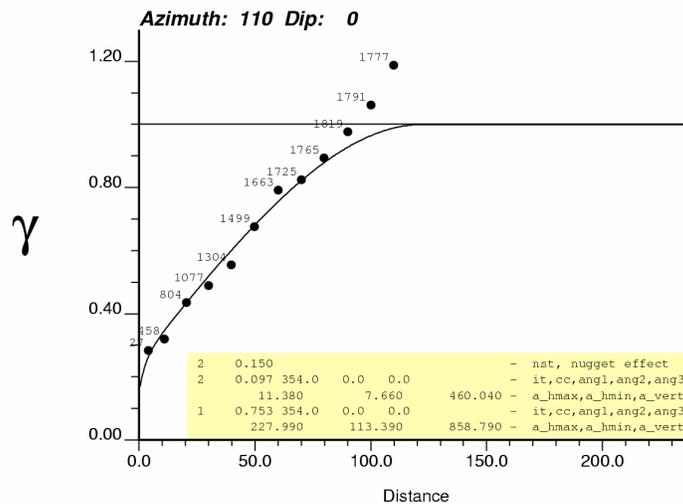
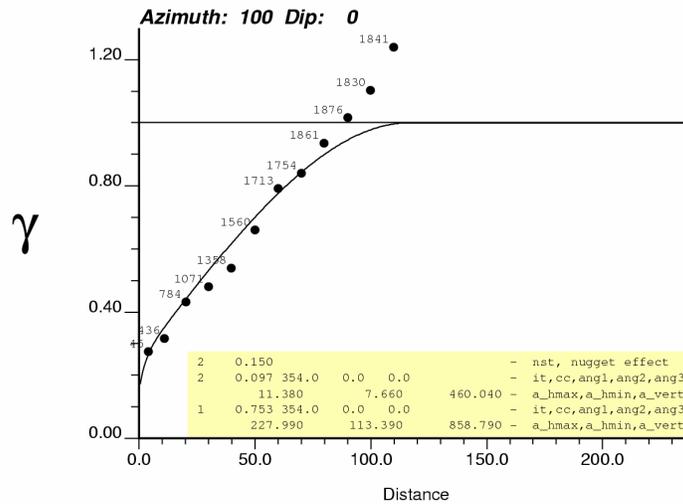
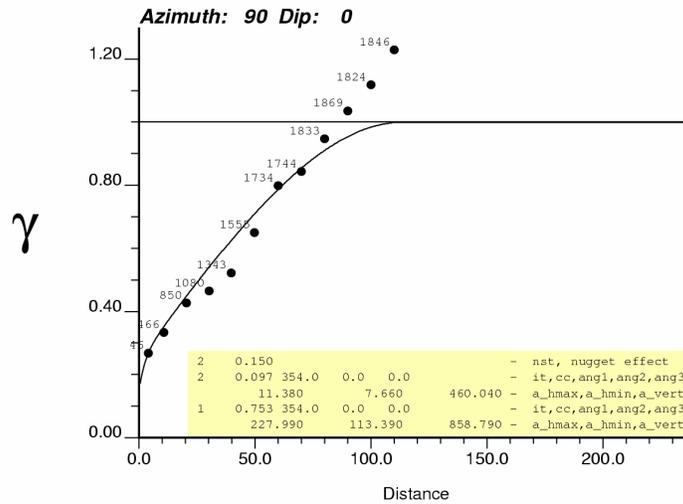


Figure 1: 3 of the 37 experimental variograms used for example 1.

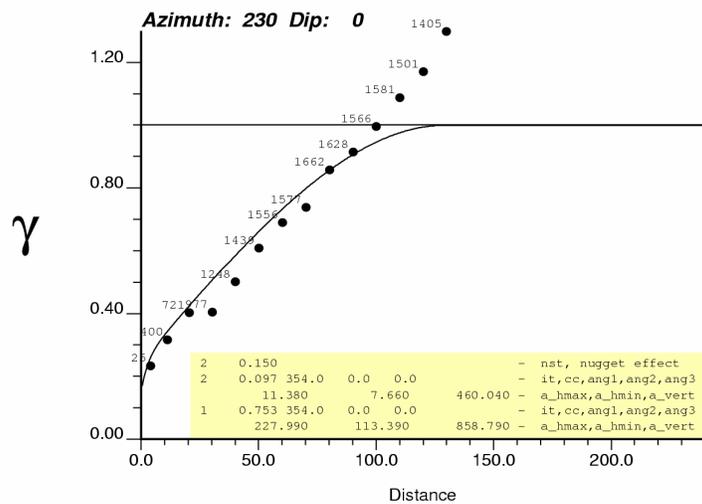
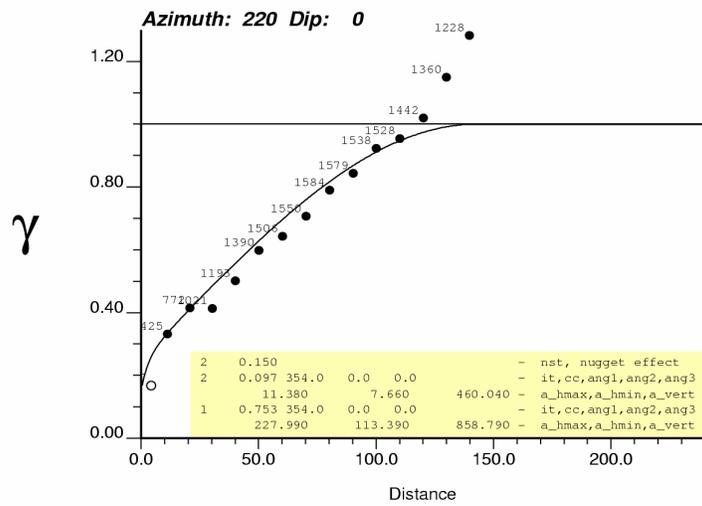
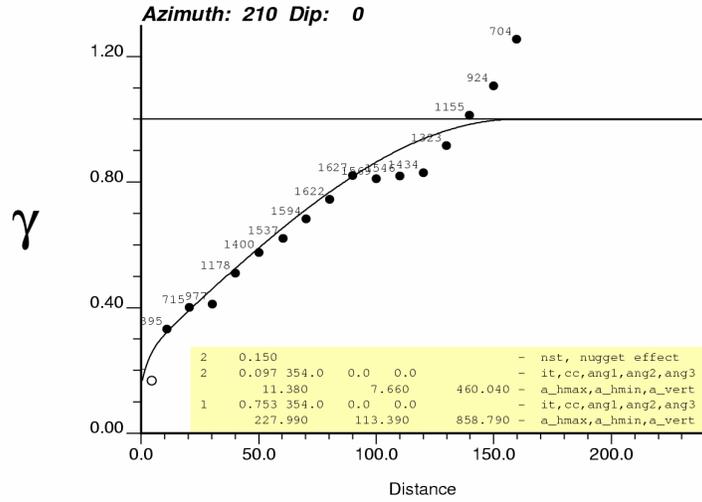


Figure 2: 3 additional variograms used for example 1.

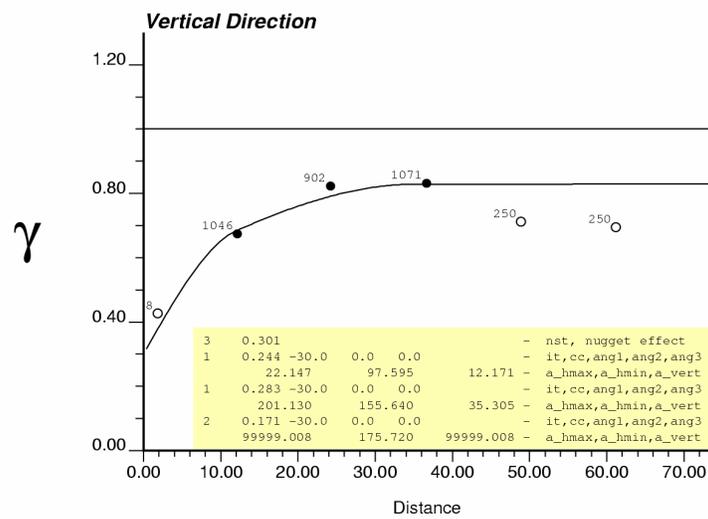
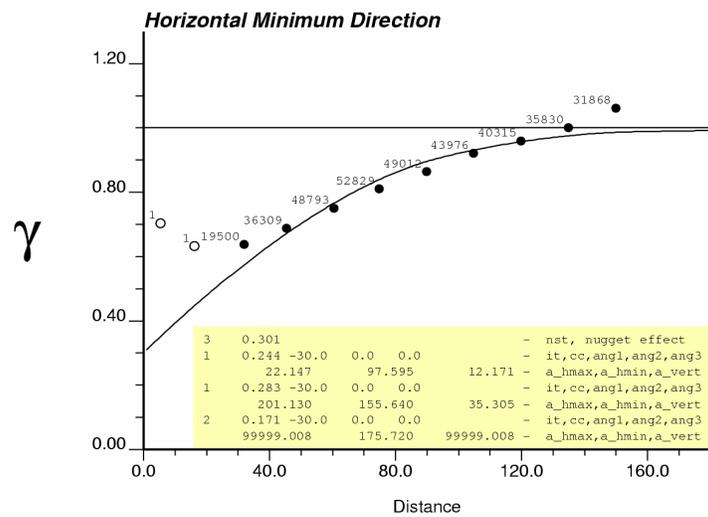
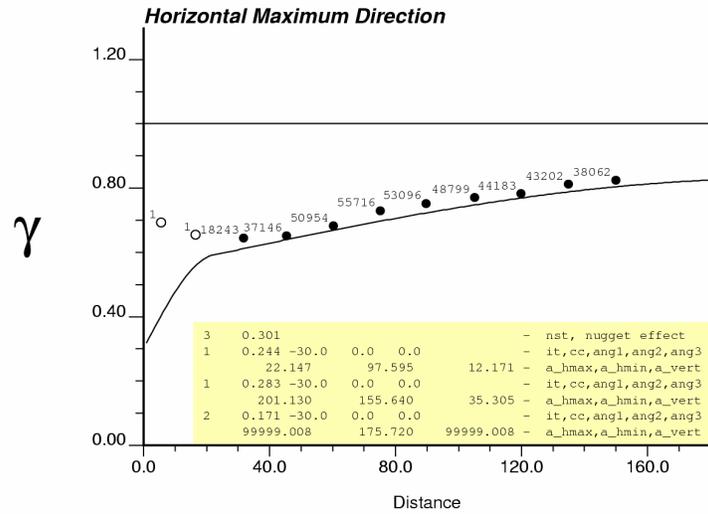


Figure 3: Modelled variogram for indicator threshold #2.

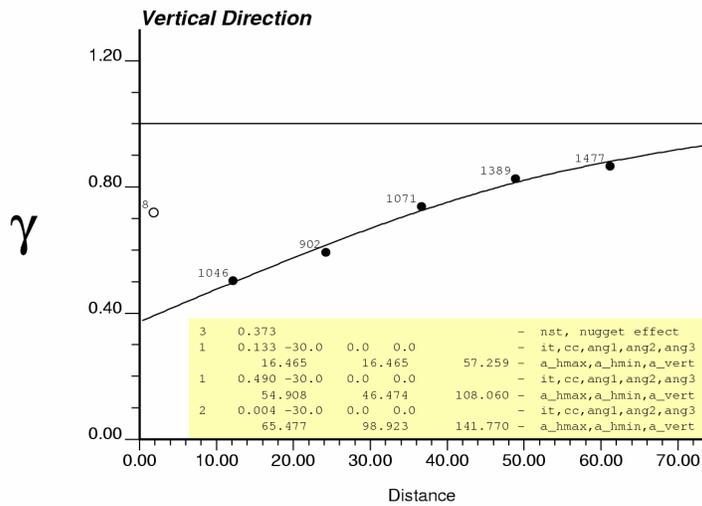
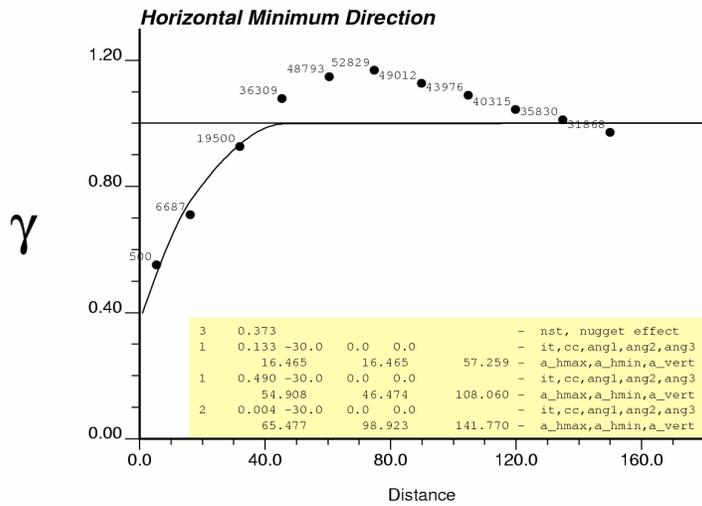
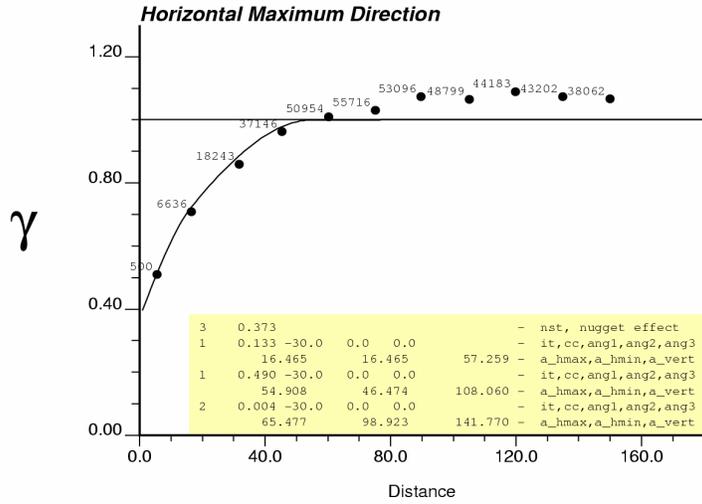


Figure 4: Modelled variogram for indicator threshold #8.

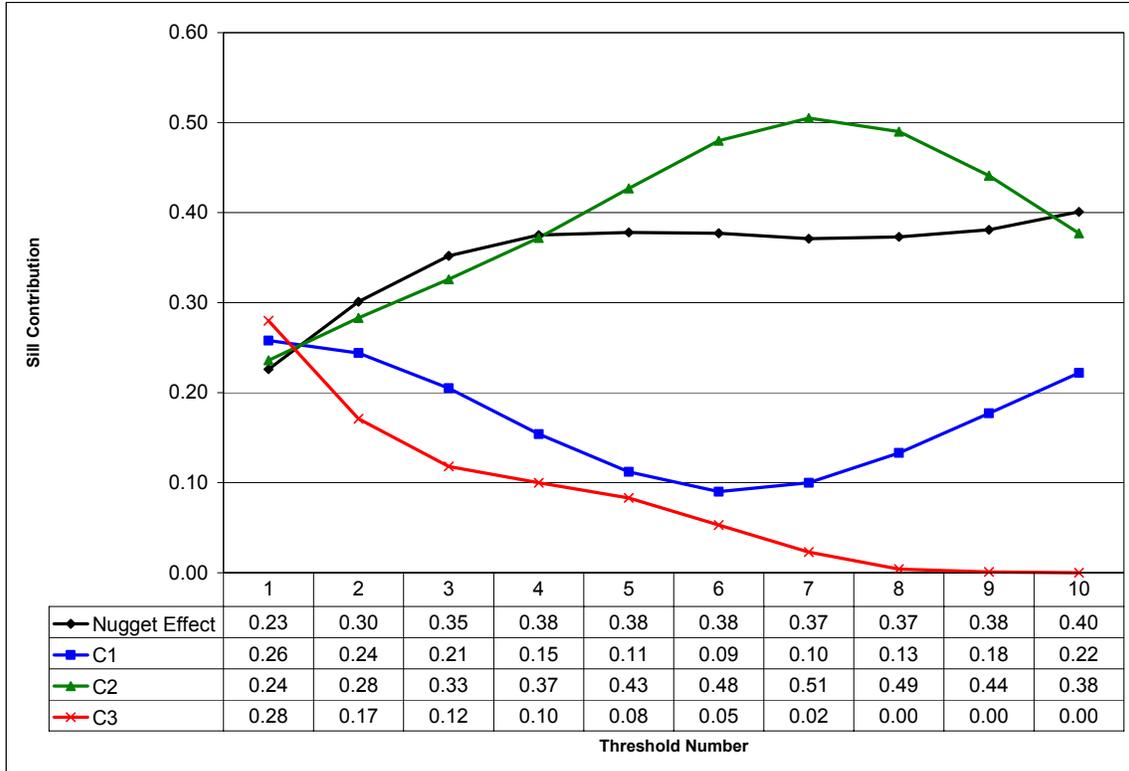


Figure 5: Change in nugget effect and sill contributions between thresholds.

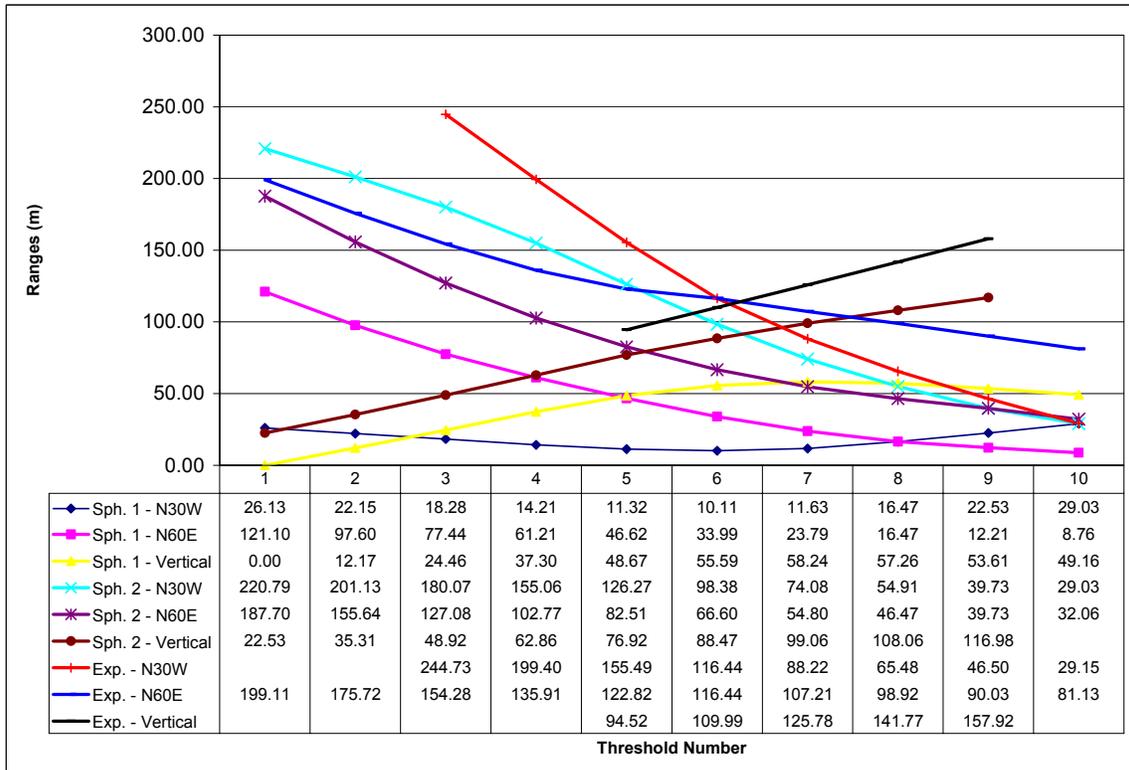


Figure 6: Change in ranges between thresholds.