VARFIT: A Program for Semi-Automatic Variogram Modelling

Paula F. Larrondo (larrondo@ualberta.ca), Chad T. Neufeld (neufeld@ualberta.ca),
and Clayton V. Deutsch (cdeutsch@ualberta.ca)
Department of Civil and Environmental Engineering,
University of Alberta

Abstract

The variogram is an important input in geostatistical techniques. Calculated experimental variograms must be modelled with a legitimate analytical model. This program obtains a legitimate model of spatial variability that closely fits directional experimental variograms. The program mimics the iterative procedure performed by the experienced geostatistician. In addition this program can fit linear models of coregionalization that are licit and positive semi-definite. The program varfit *is GSLIB compatible, but it can be run with any variogram calculation output that has been properly reformatted.*

The nugget effect, sill contributions, structure types and ranges, are fit to experimental variogram points in up to three directions simultaneously. The user can fix components of the variogram as needed. The fitting can be weighted by lag distance and/or number of pairs at each point, as an option for a better fit at shorter distances or at lags with more pairs, respectively.

Introduction

The variogram is one of the most extensively used statistical measures in geostatistics. Most geostatistical estimation and simulation algorithms require a variogram model.

Prior to variogram specification, a model must be fit to the experimental points. This requires an initial step of identifying directions of continuity. Useful and straightforward tools like a variogram map can be useful for this purpose. Once the directions are determined, experimental points can then be calculated. These are the points that must be fit using known, licit variogram models.

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 time required to model variograms increases substantially when multiple variables are considered. In these cases a linear model of coregionalization (LMC) may be required. An LMC imposes additional constraints and increases the

number of variograms to be modelled simultaneously to:

$$\text{number of variograms} = \frac{n_{variables} \cdot (n_{variables} + 1)}{2} \tag{1}$$

For example, when modelling a deposit that has 4 variables there are 10 variograms that must be fit using the same structures, the same ranges, and the corresponding contributions must be positive semi-definite.

A prototype program, `varfit`, was developed to decrease the time required for variogram modelling. This program is intended to be a helpful tool for fitting variogram models. The program does not calculate experimental variogram points or directions of continuity. It is the responsibility of the user to determine the correct spatial correlation of the data. The program automates the model fitting process given the experimental variogram points. Additional geological knowledge such as anisotropy directions and ranges can be incorporated in the fit by fixing selected variogram components.

Variogram and LMC Definition

Experimental variograms must be provided as input to `varfit`. Up to three directions for each variogram can be considered: the principal direction, *Hmax*; the horizontal direction orthogonal to the principal direction, *Hmin*; or the direction perpendicular to the plane defined by the first two directions, *Vert*. The program can optimize up to these three directions for each variogram.

Some automatic variogram fitting software consider many directions simultaneously when calculating the experimental variogram points. The apparent advantage of this approach is that the geostatistician does not need to know the principal directions of continuity; they are determined and fit automatically. The `varfit` program cannot be used in this mode, nor would we recommend it. There is considerable danger in fitting structures that may be artifacts of the data spacing and/or orientation. We recommend the geostatistician hand contour their data and calculate variogram maps to identify the main directions of continuity. Then, carefully choose variogram distance and angle parameters to get the best possible directional variograms. The experimental points can then be modelled semi-automatically by `varfit`. The rotation angles are aligned with the geological interpretation and any artifacts due to the data spacing are avoided.

Traditional semi-variogram. The traditional experimental semi-variogram is calculated as the squared difference of the variable Z at a lag distance $\mathbf{h}$ in one of the directions *Hmax*, *Hmin*, or *Vert*:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{N(\mathbf{h})} [Z(\mathbf{u}) - Z(\mathbf{u} + \mathbf{h})]^2 \tag{2}$$

where $N(\mathbf{h})$ is the number of data pairs at a lag separation distance $\mathbf{h}$.

For conventional estimation and simulation, a variogram function is required for all distances and all directions. Yet, we are only able to calculate the experimental $\hat{\gamma}(\mathbf{h})$ for a few specific lag distances $\mathbf{h}$. The need to interpolate and extrapolate the variogram function, and the fact that the variogram model must be "positive semi-definite" necessitates fitting a legitimate model to the variogram points. A variogram model is constructed as a sum of licit variogram functions:

$$\begin{aligned}\gamma(\mathbf{h}) &= \sum_{i=0}^{nst} b^i \cdot \Gamma^i(\mathbf{h}) \\ &= b^0 + b^1 \cdot \Gamma^1(\mathbf{h}) + b^2 \cdot \Gamma^2(\mathbf{h}) \dots\end{aligned} \quad (3)$$

where nst is the number of "nested structures," b^i is the variance contribution for the ith nested structure, and $\Gamma^i(\mathbf{h})$ is the elementary licit variogram function for the ith structure. The nugget effect is $\Gamma^0(\mathbf{h}) = 1$ and b^0 is the variance contribution of this structure. Licit variogram models used in `varfit` are spherical, exponential, Gaussian, and hole effect.

Cross semi-variogram. The experimental cross semi-variogram is calculated as the squared difference between the variable Z and the variable Y separated by a lag distance $\mathbf{h}$ in one of the directions *Hmax*, *Hmin*, or *Vert*:

$$\hat{\gamma}_{Z,Y}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{N(\mathbf{h})} [Z(\mathbf{u}) - Z(\mathbf{u} + \mathbf{h})] [Y(\mathbf{u}) - Y(\mathbf{u} + \mathbf{h})] \quad (4)$$

where $N(\mathbf{h})$ is the number of $Z - Y$ data pairs at a lag distance $\mathbf{h}$.

Cross variograms are modelled for the same reason that traditional variograms are modelled: the need to interpolate/extrapolate the cross variogram function, and the fact that the cross variogram model must be "positive semi-definite".

A cross variogram model is constructed as a sum of licit variogram functions:

$$\begin{aligned}\gamma_{z,y}(\mathbf{h}) &= \sum_{i=0}^{nst} b_{z,y}^i \cdot \Gamma^i(\mathbf{h}) \\ &= b_{z,y}^0 + b_{z,y}^1 \cdot \Gamma^1(\mathbf{h}) + b_{z,y}^2 \cdot \Gamma^2(\mathbf{h}) \dots\end{aligned} \quad (5)$$

where the subscripts denote the variables that the cross variograms are associated with.

Linear Model of Coregionalization. A linear model of coregionalization combines the semi-variograms and cross semi-variograms so that they are mathematically consistent to constitute a set of models that taken together are positive semi-definite. A two variable

case, z and y, is shown below:

$$\gamma_{z,z}(\mathbf{h}) = b_{z,z}^0 + b_{z,z}^1 \cdot \Gamma^1(\mathbf{h}) + b_{z,z}^2 \cdot \Gamma^2(\mathbf{h}) \dots \tag{6}$$

$$\gamma_{z,y}(\mathbf{h}) = b_{z,y}^0 + b_{z,y}^1 \cdot \Gamma^1(\mathbf{h}) + b_{z,y}^2 \cdot \Gamma^2(\mathbf{h}) \dots \tag{7}$$

$$\gamma_{y,y}(\mathbf{h}) = b_{y,y}^0 + b_{y,y}^1 \cdot \Gamma^1(\mathbf{h}) + b_{y,y}^2 \cdot \Gamma^2(\mathbf{h}) \dots \tag{8}$$

where $\Gamma^i(\mathbf{h}), i = 1, \dots, nst$ are the nested structures made up of common variogram models.

The LMC is fit in a similar manner to the traditional semi-variogram. However there are some additional considerations: (1) each nested structure that is present in the cross variogram must also be present in the corresponding direct variograms (with the same structure type and range) and (2) the set of b^i coefficients must be positive semi-definite (the check used is detailed later in the paper). Any number of variables may be considered in practice and in `varfit`.

Initial Variogram Parameters

After the experimental points have been determined the first step in `varfit` is to calculate the initial components of the variogram that have not been specified by the user. The variogram components that can be specified by the user are outlined in the appendix where the program parameters are explained.

Sill. When the sill is to be optimized during the modelling of the experimental variogram, the initial value is calculated as the average of all the experimental points:

$$sill = \frac{1}{N(\mathbf{h})} \sum_{N(\mathbf{h})} \hat{\gamma}(\mathbf{h}) \tag{9}$$

If the sill is fixed, it will not be changed by `varfit` and the sum of the nugget effect and variance contributions remains constant:

$$\text{Total Sill} = \text{Nugget Effect} + \text{Variance Contributions from Structures} \tag{10}$$

Nugget effect. When the nugget effect is not fixed by the user, the initial value is set to 0. If the nugget effect is fixed it will not be changed by `varfit`. If the sill is also fixed, the sum of the nugget effect and variance contributions remains constant.

Nested structure variance contributions. After the sill and nugget effect have been determined the remaining variance, the sill minus the nugget effect, is divided equally among the nested variogram structures.

Nested structure types. All of the initial structures are set to spherical. The only exception is when the user specifies that cyclicity is present. The type of structure added by `varfit` for cyclicity is the hole effect model.

Hmax, Hmin, and Vert ranges. The initial range for each direction is calculated as a function of the maximum distance in that direction.

$$a_j^i \quad = \quad \frac{i}{nst} \cdot (\text{maximum lag distance in direction } j) \qquad (11)$$

where $i = 1, \ldots, nst$ and $j = $ *Hmax, Hmin,* and *Vert.* For example, if there were three nested structures with a maximum range of 1.2, the range of the first structure would be 1/3 of the maximum range, the second structure would be 2/3 of the maximum range, and the range of the third structure would be equal to the maximum range.

Experimental Point Weighting

During the optimization process in `varfit` an objective function is used to measure the "goodness" of the model. When calculating the objective function in `varfit` the default is to give all of the experimental points an equal weight. For added flexability the points may be assigned different weights based on inverse distance, the number of pairs for that point, or a user specified preference. The different weighting options may be used independently or all together.

When an inverse distance or number of pairs weighting has been chosen, each direction for each variogram is weighted independently. This is done to ensure that the modelled variogram is a best fit considering all directions simultaneously. The user preference adjusts the weighting of the experimental points to improve the fit for a specific variogram in a specific direction. The user preference should only be used when necessary.

Inverse distance. The inverse distance weighting should be used to get a better fit at shorter lag spacing. The points for each variogram in each direction are weighted by the following weight:

$$\lambda_i(j, k, l) \quad = \quad \frac{\frac{1}{d(\mathbf{u}_i)}}{\sum\limits_{i=1}^{n} \frac{1}{d(\mathbf{u}_i)}} \qquad (12)$$

where i is the experimental variogram point, j is the variogram direction, k is the head variable of the variogram, l is the tail variable of the variogram, and $d(\mathbf{u}_i)$ is the lag separation distance for the ith point.

Number of pairs. The number of pairs weighting should be used to get a better fit where more information was used to calculate the experimental points. The points for each variogram in each direction are weighted by the following weight:

$$\lambda_i(j,k,l) \;\; = \;\; \frac{np(\mathbf{u}_i)}{\sum\limits_{i=1}^{n} np(\mathbf{u}_i)} \tag{13}$$

where i is the experimental variogram point, j is the variogram direction, k is the head variable of the variogram, l is the tail variable of the variogram, and $np(\mathbf{u}_i)$ is the number of pairs for the experimental point i.

User preference. The user preference weight is a scaling weight that is applied to the weights that a specific variogram in a specific direction has already received. It adjusts equal weighted points, inverse distance weighted points, or number of pairs weighted points. It should be used to provide a better fit for a specific variogram in the presence of sparse data or complex cases. Note that as the fit for one particular variogram is forced to be better, the fit of a different variogram will become worse.

Optimization Methodology

The method `varfit` uses to attain a final model is very similar to the process that an experienced user would undertake. An initial model is proposed and then small changes are made to the initial model until a "best fit" is achieved. The program `varfit` does this by iterating the initial model and checking the change in the objective function until a model that minimizes the objective function is reached.

For each iteration, one parameter of the variogram is chosen at random and then changed randomly. When the nugget effect is changed, its value will be randomly increased or decreased up to ±7.5% of the total sill. The variance contributions of each structure are changed identically to the nugget effect. When the sill of the variogram is fixed the variance contributions of each structure are scaled to maintain the correct sill value. When the structure type is changed, `varfit` randomly chooses between spherical, exponential or Gaussian structures. The ranges are iterated by choosing a random value in the range of ±7.5% of the maximum lag distance in the corresponding direction. Each range, major, minor and vertical, is optimized independently. Any one, or all of these, parameters may be fixed by the user, and they will not be changed during the iterative process.

During the optimization process two checks are done to measure the validity, or "goodness," of the fit. The first check is to minimize the objective function. Any change that increases the objective function is rejected, and any change that decreases the objective function is accepted. The second check is to make sure that the LMC model is positive semi-definite. Any change that causes the LMC to not be positive semi-definite is rejected.

Objective function.   The objective function is the squared difference between the experimental variogram points and the model variogram at each lag separation distance. The squared difference is calculated as:

$$\text{Objective Function} = \sum_{k=1}^{n_{var}} \sum_{l=k}^{n_{var}} \sum_{idir=1}^{n_{dir}} \sum_{ilag=1}^{n_{lag}} \lambda_{ilag}^{idir} \cdot \left[\gamma_{k,l}(\mathbf{h}_{ilag}^{idir}) - \hat{\gamma}_{k,l}(\mathbf{h}_{ilag}^{idir}) \right]^2 \qquad (14)$$

where n_{var} is the number of variables, n_{dir} is the number of directions, n_{lag} is the number of lags distances, λ_{ilag}^{idir} is the weight for the experimental point, $\hat{\gamma}_{k,l}(\mathbf{h}_{ilag}^{idir})$ is the experimental variogram value for each lag distance $ilag$ and direction $idir$, and $\gamma_{k,l}(\mathbf{h}_{ilag}^{idir})$ is the corresponding variogram value from the fitted model. If there is only one variable, that is $nvar = 1$, then one variogram is used to calculate the objective function. If there is more than one variable, $nvar > 1$, then the number of variograms used for the objective function is consistent with the number of variograms from the LMC relation in Equation (1).

After each iteration a new objective function is calculated and compared with the previous one. If the new objective function is less than the previous one the change is accepted. However, if the change increases the objective function, the change is rejected and the variogram parameter that was changed is restored to its previous value.

Positive semi-definite check.   As mentioned before, the set of $b_{i,j}^l$ coefficients must be positive semi-definite. Consider an LMC with P variables. The $b_{i,j}^l$ coefficients can be organized into a $P \times P$ symmetric covariance matrix where $b_{i,j}^l = b_{j,i}^l$ with $l = 0, \ldots, L$ structures, and $i, j = 1, \ldots, P$ variables:

$$B^l(\mathbf{h}) = \begin{bmatrix} b_{1,1}^l(\mathbf{h}) & \cdots & b_{1,P}^l(\mathbf{h}) \\ \vdots & \ddots & \vdots \\ b_{P,1}^l(\mathbf{h}) & \cdots & b_{P,P}^l(\mathbf{h}) \end{bmatrix}, \; \forall \mathbf{h}, l \qquad (15)$$

To ensure that the LMC is positive semi-definite two checks are done by `varfit`: (1) all of the diagonal values in the matrix are greater than or equal to zero, $b_{i,i}^l > 0, i = 1, \ldots, P$, and (2) all of the leading principal determinants of the matrix are greater than or equal to zero:

$$\det B^l(\mathbf{h}) = \begin{vmatrix} b_{1,1}^l(\mathbf{h}) & \cdots & b_{1,p}^l(\mathbf{h}) \\ \vdots & \ddots & \vdots \\ b_{p,1}^l(\mathbf{h}) & \cdots & b_{p,p}^l(\mathbf{h}) \end{vmatrix} \geq 0, \; p = 1, \ldots, P, \; \forall \mathbf{h}, l \qquad (16)$$

After each iteration the diagonal values and principle leading determinants are checked. If any one of them fails the positive semi-definiteness check, the iteration is rejected and the variogram parameter that was changed is restored to its previous value.

Examples

Three examples are provided to illustrate the application of the `varfit` program. The first example is a single variable variogram with a zonal anisotropy, the second is a test using a specified input variogram model, and the last example is a four variable LMC. The parameters for the program, `varfit`, are explained in the appendix.

1: Single Variable with Zonal Anisotropy

The first test is to use real data for a single variable. The experimental points showed a zonal anisotropy so this was specified in the parameter file. Figure 1 shows how `varfit` modelled the experimental points using a zonal anisotropy in the vertical direction. The final variogram model is also shown in Figure 1.
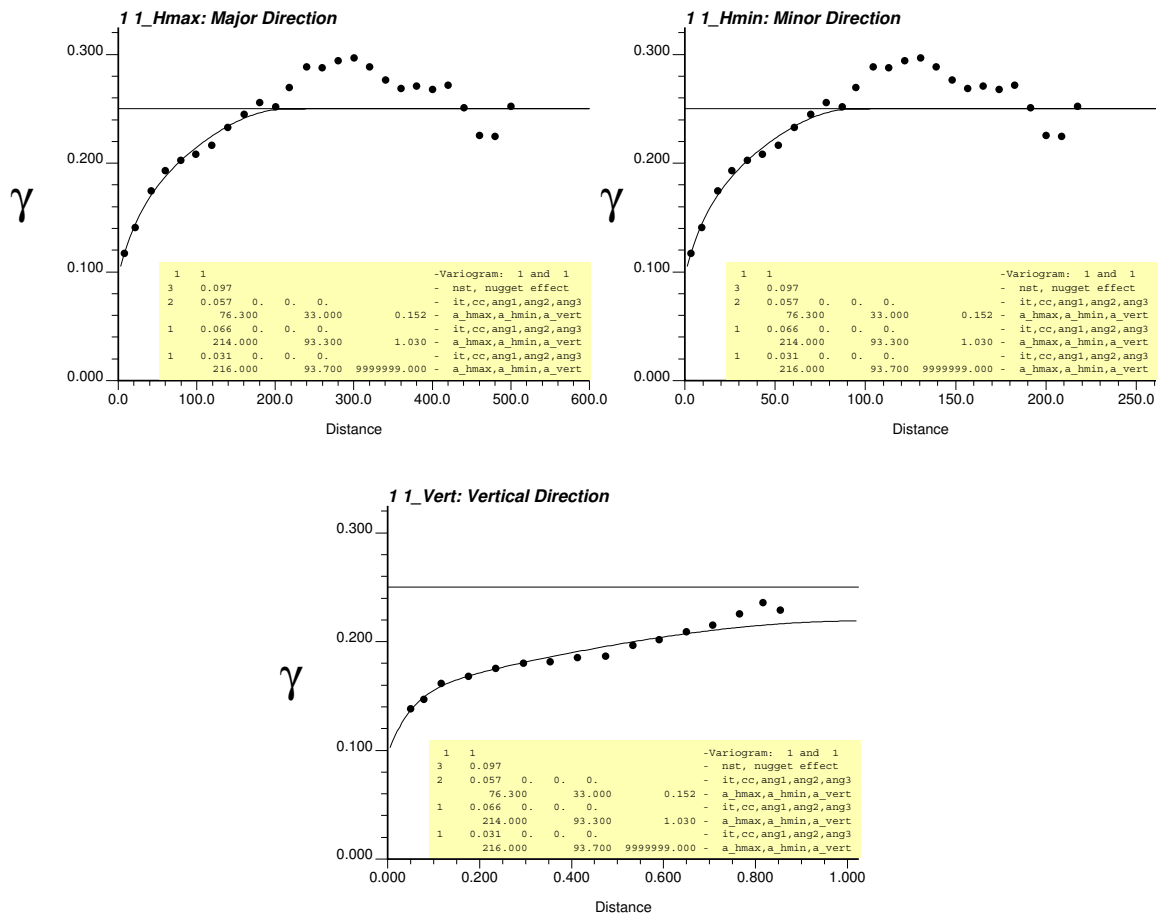

Figure 1: Single variable variogram with a zonal anisotropy.

2: Two Variables — Intrinsic Model of Coregionalization

For this example a two variable intrinsic LMC model is chosen to see if `varfit` could reproduce a known input model, and to see if its objective function and positive semi-definite checks are working properly. The following direct variogram is chosen:

$$\gamma(\mathbf{h}) \quad = \quad 0.0 + 0.5 \cdot \Gamma^1(\mathbf{h}) + 0.5 \cdot \Gamma^2(\mathbf{h}) \tag{17}$$

where $\Gamma^1(\mathbf{h})$ is a spherical structure with a range of 3m, and $\Gamma^2(\mathbf{h})$ is a spherical structure with a range of 15m, and to test the LMC portion of `varfit` three different cross variograms are tested:

$$\gamma_{(z,y)}(\mathbf{h}) \quad = \quad 0.0 + 0.35 \cdot \Gamma^1(\mathbf{h}) + 0.35 \cdot \Gamma^2(\mathbf{h}) \tag{18}$$
$$\gamma_{(z,y)}(\mathbf{h}) \quad = \quad 0.0 + 0.5 \cdot \Gamma^1(\mathbf{h}) + 0.2 \cdot \Gamma^2(\mathbf{h}) \tag{19}$$
$$\gamma_{(z,y)}(\mathbf{h}) \quad = \quad 0.0 + 0.2 \cdot \Gamma^1(\mathbf{h}) + 0.5 \cdot \Gamma^2(\mathbf{h}) \tag{20}$$

All of the cross variograms have a correlation coefficient of 0.7. The first cross variogram is the intrinsic model, the second variogram is the short-range model where the short range structure receives as much weight as the positive semi-definite check will allow, and the last variogram is the long-range model. Figure 2 shows the three input cross variogram models.

The test is performed three times, once for each cross variogram, and `varfit` modelled the input variograms exactly. Figure 3 the variogram models from `varfit` and the points from the input models.
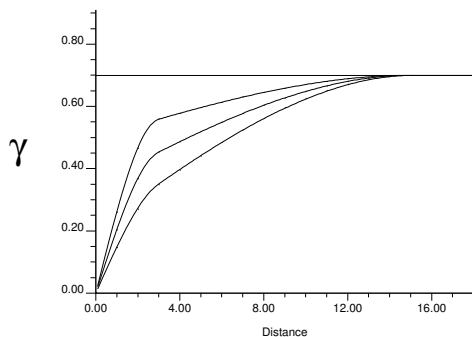
Figure 2: Intrinsic LMC case input cross variogram models. The upper variogram is the short-range model, the middle variogram is the intrinsic model, and the lower variogram is the long range model.
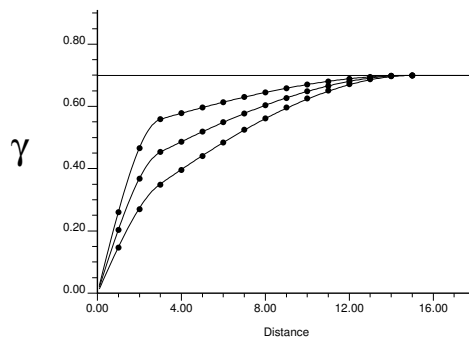
Figure 3: Intrinsic LMC case output cross variogram models. The points are from the input models and the lines are the output from `varfit`. In all three cases `varfit` fit the input exactly.

9

3: Four Variable LMC

Recall that for an LMC, the number of variograms that need to be simultaneously modelled is:

$$\text{number of variograms} = \frac{n_{variables} \cdot (n_{variables} + 1)}{2}$$

Therefore, for four variables, 10 variograms need to be modelled. Modelling 10 variograms with the same nested structures and ranges, while making sure the group of models is positive semi-definite could take hours. Including the time to set up the parameter file and to run the program this could be completed in much less time by using `varfit`.

After running `varfit` the variogram models were checked externally to ensure positive semi-definiteness. The modelled variograms from `varfit` are:

$$
\begin{aligned}
\gamma_{1,1}(\mathbf{h}) &= 0.100 + 0.510 \cdot \Gamma^1(\mathbf{h}) + 0.355 \cdot \Gamma^2(\mathbf{h}) + 0.035 \cdot \Gamma^3(\mathbf{h}) \\
\gamma_{1,2}(\mathbf{h}) &= 0.030 + 0.029 \cdot \Gamma^1(\mathbf{h}) + 0.100 \cdot \Gamma^2(\mathbf{h}) + 0.030 \cdot \Gamma^3(\mathbf{h}) \\
\gamma_{1,3}(\mathbf{h}) &= -0.002 + -0.024 \cdot \Gamma^1(\mathbf{h}) + -0.025 \cdot \Gamma^2(\mathbf{h}) + -0.075 \cdot \Gamma^3(\mathbf{h}) \\
\gamma_{1,4}(\mathbf{h}) &= -0.010 + -0.000 \cdot \Gamma^1(\mathbf{h}) + -0.105 \cdot \Gamma^2(\mathbf{h}) + -0.015 \cdot \Gamma^3(\mathbf{h}) \\
\gamma_{2,2}(\mathbf{h}) &= 0.100 + 0.510 \cdot \Gamma^1(\mathbf{h}) + 0.355 \cdot \Gamma^2(\mathbf{h}) + 0.035 \cdot \Gamma^3(\mathbf{h}) \\
\gamma_{2,3}(\mathbf{h}) &= -0.050 + -0.118 \cdot \Gamma^1(\mathbf{h}) + -0.259 \cdot \Gamma^2(\mathbf{h}) + -0.114 \cdot \Gamma^3(\mathbf{h}) \\
\gamma_{2,4}(\mathbf{h}) &= 0.050 + -0.272 \cdot \Gamma^1(\mathbf{h}) + -0.209 \cdot \Gamma^2(\mathbf{h}) + -0.197 \cdot \Gamma^3(\mathbf{h}) \\
\gamma_{3,3}(\mathbf{h}) &= 0.100 + 0.369 \cdot \Gamma^1(\mathbf{h}) + 0.284 \cdot \Gamma^2(\mathbf{h}) + 0.247 \cdot \Gamma^3(\mathbf{h}) \\
\gamma_{3,4}(\mathbf{h}) &= 0.030 + 0.131 \cdot \Gamma^1(\mathbf{h}) + 0.194 \cdot \Gamma^2(\mathbf{h}) + 0.002 \cdot \Gamma^3(\mathbf{h}) \\
\gamma_{4,4}(\mathbf{h}) &= 0.100 + 0.362 \cdot \Gamma^1(\mathbf{h}) + 0.298 \cdot \Gamma^2(\mathbf{h}) + 0.240 \cdot \Gamma^3(\mathbf{h})
\end{aligned}
$$

where $\Gamma^1(\mathbf{h})$ is a spherical structure with a range of 13.1m horizontally and 11.4m vertically, $\Gamma^2(\mathbf{h})$ is an exponential structure with a range of 79.4m horizontally and 11.4m vertically, and $\Gamma^3(\mathbf{h})$ is a spherical structure with a range of 9999m horizontally and 16.6m vertically.

Discussion

As shown in the previous examples, `varfit` is capable of fitting a model to a simple single variable variogram that has zonal anisotropy, reproducing an input model, and providing a legitimate LMC model.

The program runs very fast. On a fairly new desktop computer, 2.0 GHz Pentium 4, it takes less than a second for a single variable variogram. As the number of variables increase, the checks become more complex and more variograms need to be modelled simultaneously. The four variable LMC example took approximately 20 seconds to run.

While `varfit` does a good job in the cases we tested, there are common situations where fixing some of the variogram parameters makes a significant difference in the processing time

and final results. The sill is the most important variogram parameter that should be fixed. When trends, zonal anisotropies, and/or cyclicity are present, automatically determining the sill does not work very well and is inefficeint. However, in most cases the practicioner will know what the sill is supposed to be before any attempts are made to model the experimental variograms. Since there is a drastic difference between certain structure types `varfit` may get "stuck" with one particular type and not be able to change to a better structure type. Choosing the type of structure is a simple task for the practicioner and will give better results. If other geological information is available that allows additional variogram parameters to be fixed then they can, and should be fixed. However, trying to impose an incorrect model onto the experimental points will not work.

Conclusions

The program `varfit` is designed to provide geostatisticians with a legitimate variogram model. It mimics the iterative process that an experienced practitioner would undertake when modelling a variogram. The program can fit single variable and multiple variable (LMC) variograms.

Experimental variograms must be calculated by the user and provided as input to `varfit`. We do not recommend automated methods for calculating the experimental variograms. There is considerable danger in fitting structures that may be artifacts of the data spacing and/or orientation.

The nugget effect, variance contributions, structure types, and ranges, are iteratively perturbed and fit to the experimental variogram points with n variables in up to three directions simultaneously. Any subset of parameters can be fixed by the user, that can account for expert geological interpretation.

The "goodness" of the fit is measured by the sum of the squared differences between the experimental variogram points and the modelled variogram called the objective function. During the modelling process any changes that decrease the objective function are accepted while those that increase it are rejected. The points used to calculate the objective function can be weighted by the number of pairs or the inverse of the lag distance. This is done to get a better fit at shorter distances and/or where the points have more pairs. When fitting an LMC an additional positive semi-definite check is done. Whenever an iteration causes the group of variograms to be non-positive semi-definite the change is rejected.

Future work.

The next step we see for semi-automatic variogram fitting is indicator variograms. The approach will be very similar to the current version of `varfit` but will allow multiple indicators to be fit simultaneously with the option of an indicator LMC.

References

[1] Cressie N. Fitting variogram models by weighted least squares. *Mathematical Geology*, 17(5):563–586, 1985.

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

[3] Ellis T. M. R., Philips I. R., and Lahey T. M. *Fortran 90 Programming*. Addison-Wesley, 1994.

[4] Genton M. G. Variogram fitting by generalized least squares using an explicit formula for the covariance structure. *Mathematical Geology*, 30(4):323–345, 1998.

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

[6] Gringarten E. and Deutsch C. V. Teacher's Aide: Variogram Interpretation and Modeling. *Mathematical Geology*, 33(4):507–534, 2001.

[7] Harris J. W. and Stocker H. *Handbook of Mathematics and Computational Science*. Springer-Verlag New York, Inc., New York, 1998.

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

[9] Johnson R. A. and Wichern D. W. *Applied Multivariate Statistical Analysis*. Prentice-Hall, Inc., New Jersey, 1998.

[10] Journel A. G. and Huijbregts Ch. J. *Mining Geostatistics*. Academic Press, 1978.

[11] Reklaitis G. V., Ravindran A., and Ragsdell K. M. *Engineering Optimization: Methods and Applications*. John Wiley & Sons, Inc., New York, 1983.

Appendix: Parameter Definition

The semi-automatic variogram modelling program, `varfit`, takes a set of experimental variogram points and fits a model to those points. It can fit multiple variograms and provide positive semi-definite linear models of coregionalization. The output from the program consists of two files: (1) the final variogram models and (2) a multiple page PostScript plot of all the variograms. For ease of use, the parameter file is broken into four sections. The required blocks are:

- **MAIN:** output files, number of iterations, number of variables

- **VARIOGRAM SPECIFICATION:** variogram head and tail variables, direction, input file, titles

None, one of, or both of the optional blocks may be present in the parameter file. The optional blocks include:

- **PAGE TITLE:** titles for each page in the output PostScript file

- **ADVANCED OPTIONS:** fix variogram components, anisotropy, cyclicity

The parameters are listed below and shown in Figure 4:

MAIN block of the parameter file: The main block of the parameter file is where the basics of the program operation are set. It is required by the program. Included in the main block are:

- **Direction angles 1, 2,** and **3:** The rotation angles as used in GSLIB: **ang1** is the azimuth correction, **ang2** is the dip correction, and **ang3** is the plunge correction.

- **Number of variables:** The number of different variables that were used to calculate the experimental variogram points.

- **Number of nested structures:** The user must specify the number of nested structures that are used in the variogram modelling. A better fit is achieved with more structures, but it may not be warranted by limited experimental data. Note that choosing zonal anisotropy or cyclicity adds a nested structure (these options are in the ADVANCED OPTIONS).

- **Inverse distance weighting (0=no, 1=yes):** This parameter is set to 0 or 1 depending on whether or not an inverse distance weighting will be used. Points that are closer to the origin will receive a higher weight than points further away.

- **Number of pairs weighting (0=no, 1=yes):** This parameter is set to 0 or 1 depending on whether or not a number of pairs weighting will be used. Points with a higher number of pairs will receive a higher weighting than points with a lower number of pairs.

  No weighting, one type of weighting, or both types of weighting may be chosen. To equally weight the points, choose no weighting options.

- **Minimum number of pairs to use:** The minimum number of pairs to use for calculating the objective function. Any point with less pairs is plotted as an empty circle in the variogram plots.

- **Number of iterations:** The number of iterations to perform while fitting the variogram(s). If this number is set too low the objective function will still be decreasing when the program stops.

- **File for PostScript output:** This file contains the multiple page PostScript output file.

- **File for variogram model(s):** This file contains the fitted variogram models. They are formatted according to GSLIB convention.

VARIOGRAM SPECIFICATION block of the parameter file: The variogram specification block of the parameter file is where the variograms are defined and the input file names are given. For each variogram direction the following three lines must be repeated. It is required by the program. Included in the variogram specification block are:

- **ivar1, ivar2, idir, ivarn:** The variogram head variable, **ivar1**, the variogram tail variable, **ivar2**, the direction of the variogram, **idir**, where 1 is Hmax direction, 2 is the Hmin direction, and 3 is the Vert direction, and the variogram number in the input file, **ivarn**.

- **Variogram file:** File with the experimental variogram points for the variogram.

- **Variogram title:** Optional title to be plotted with the variogram in the PostScript file.

PAGE TITLES block of the parameter file: The page title block of the parameter file is where titles for each page of the output file are specified. For each title the following two lines must be repeated. It is optional and may be completely or partially left out of the parameter file. Included in the page titles block are:

- **ivar1, ivar2:** **ivar1** and **ivar2** variogram for the page title.

- **title:** Title of the page that contains the **ivar1** and **ivar2** variogram.

ADVANCED OPTIONS block of the parameter file. The advanced options block of the parameter file is where specific variogram components can be fixed. It is optional but if it is used it must be complete. Included in the advanced options block are:

- **Zonal Anisotropy: Hmax, Hmin, Vert (0=no, 1=yes):** These indicator variables specify if there is a zonal anisotropy in any of the directional variograms. If "yes" is set for **Hmax** or **Hmin**, the program will consider that a horizontal zonal anisotropy exists regardless of the direction. If a zonal anisotropy is specified for both horizontal and vertical directions, the program will choose the vertical direction and assumes that no zonal anisotropy exists for the horizontal direction. The program will automatically add one structure for the zonal anisotropy.

- **Cyclicity: Hmax, Hmin, Vert (0=no, 1=yes):** These indicator variables specify if there is cyclicity in any of the directional variograms. The program will assume that cyclicity can be in only one direction, so if more that one direction is set to 1, all directions will be reset to 0. If cyclicity is set in any direction zonal anisotropies will be removed. The program will automatically add one structure to account for cyclicity modelling.

- **Number of sills to fix:** The number of variogram sills to fix during the optimization. For each sill that is fixed a line must be added beneath the number of sill to be fixed.

- **Number of nugget effects to fix:** The number of variogram nugget effects to fix during the optimization. For each nugget effect that is fixed a line must be added beneath the number of nugget effects to be fixed.

- **Number of structure types to fix:** The number of structure types to fix during the optimization. For each structure type that is fixed a line must be added beneath the number of structure types to be fixed.

- **Number of Hmax ranges to fix:** The number of Hmax ranges to fix during the optimization. For each Hmax range that is fixed a line must be added beneath the number of Hmax ranges to be fixed.

- **Number of Hmin ranges to fix:** The number of Hmin ranges to fix during the optimization. For each Hmin range that is fixed a line must be added beneath the number of Hmin ranges to be fixed.

- **Number of Vert ranges to fix:** The number of Vert ranges to fix during the optimization. For each Vert range that is fixed a line must be added beneath the number of Vert ranges to be fixed.

- **Number of variogram preferences:** The number of variogram preferences to set during the optimization. For each preference that is set a line must be added beneath the number of preferences that are set. When optimizing the variograms the fit in one direction may be compromised to improve the fit in another direction. Setting a variogram preference adjusts the weighting that a specific variogram receives to force an improved fit in that direction at the expense of another direction.

- **Hmax/Vert anisotropy (0=optimize, 1=fix):** The horizontal to vertical anisotropy can be fixed if there is geological knowledge to consider or if the experimental data is inadequate to fit automatically.

- **Hmin/Hmax anisotropy (0=optimize, 1=fix):** The horizontal **Hmin/Hmax** anisotropy is defined the same as the **Hmax/Vert** anisotropy.

- **ivar1, ivar2, sill:** The definition of the sill that is fixed. It is similar to the variogram definition but the direction does not need to be specified. This line is located at the end of the parameter file and needs to be inserted beneath the number of sills that are fixed.

- **ivar1, ivar2, nugget effect:** The definition of the nugget effect that is fixed. It is similar to the variogram definition but the direction does not need to be specified. This line is located at the end of the parameter file and needs to be inserted beneath the number of nugget effects that are fixed.

- **structure number and structure type:** The definition of the structure type that is fixed. Structure types codes are: (1) Spherical, (2) Exponential, (3) Gaussian, and (5) Hole Effect.

- **structure number and Hmax range:** The definition of the Hmax range that is fixed.

- **structure number and Hmin range:** The definition of the Hmin range that is fixed.

- **structure number and Vert range:** The definition of the Vert range that is fixed.

- **ivar1, ivar2, idir, number for preference:** The definition of the variogram preference that is set. It is similar to the variogram definition. The variogram preference is a real number that is multiplied by the weight the variogram points receive.

```
                     Parameters for VARFIT
                     *********************

START OF PARAMETERS:


START OF MAIN:
0.0  0.0  0.0                 -directions: ang1,2,3 (GSLIB definition)
2                             -number of variables
2                             -number of nested structures
0                             -inverse distance weighting (0=no, 1=yes)
0                             -number of pairs weighting (0=no, 1=yes)
10                            -minimum number of pairs to use
5000                          -number of iterations
varfit.ps                     -file for PostScript output
varfit.var                    -file for variogram model(s)


START OF VARIOGRAMS SPECIFICATION:
1  1  1  1                    -ivar1, ivar2, idir, ivarn
gamv_11.out                   -    variogram file
Cu Semivariogram              -    variogram title
1  2  1  1                    -ivar1, ivar2, idir, ivarn
gamv_12.out                   -    variogram file
Cu-Ag Cross-semivariogram     -    variogram title
2  2  1  1                    -ivar1, ivar2, idir, ivarn
gamv_22.out                   -    variogram file
Ag Semivariogram              -    variogram title


START OF PAGE TITLES:
1 1                           -ivar1, ivar2
Copper Variogram              -    title for ivar1, ivar2 page
1 2                           -ivar1, ivar2
Copper-Silver Crossvariogram  -    title for ivar1, ivar2 page
2 2                           -ivar1, ivar2
Silver Variogram              -    title for ivar1, ivar2 page


START OF ADVANCED OPTIONS:
0    0    0                   -zonal Anis: Hmax, Hmin, Vert (0=no, 1=yes)
0    0    0                   -cyclicity:  Hmax, Hmin, Vert (0=no, 1=yes)
0                             -number of sills to fix
0                             -number of nugget effects to fix
0                             -number of structure types to fix
0                             -number of Hmax ranges to fix
0                             -number of Hmin ranges to fix
0                             -number of Vert ranges to fix
0                             -number of variogram preferences
0    10.0                     -fix Hmax/Vert anis. (0=no, 1=yes)
0     1.0                     -fix Hmin/Hmax anis. (0=no, 1=yes)

Insert the required lines (from below) for the options chosen
above (must be placed below the option flag).  Some parameters
can be fixed for each variogram (sill or nugget effect) while
others (hmax, hmin, hvert, anisotropy) are fixed for all variograms
1  2  0.65                    -    ivar1, ivar2, sill
1  2  0.05                    -    ivar1, ivar2, nugget effect
1  2                          -    structure number and structure type
1  1000                       -    structure number and Hmax range
1   200                       -    structure number and Hmin range
1    10                       -    structure number and Vert range
1    10                       -    ivar1, ivar2, idir, number for preference
```

Figure 4: An example parameter file for `varfit`.