Geostatistics with Global Mapper

Contents

1 Introduction

Geostatistics is a branch of statistics focused on quantities that vary with geographic location [1]. The central tool of geostatistics is the variogram which provides information about the variability in a quantity as a function of spatial separation. Kriging is an estimation method that uses the variogram to provide optimal estimates at points of interest. Fig. 1 shows variography and kriging tools applied to ozone measurements in Global Mapper.

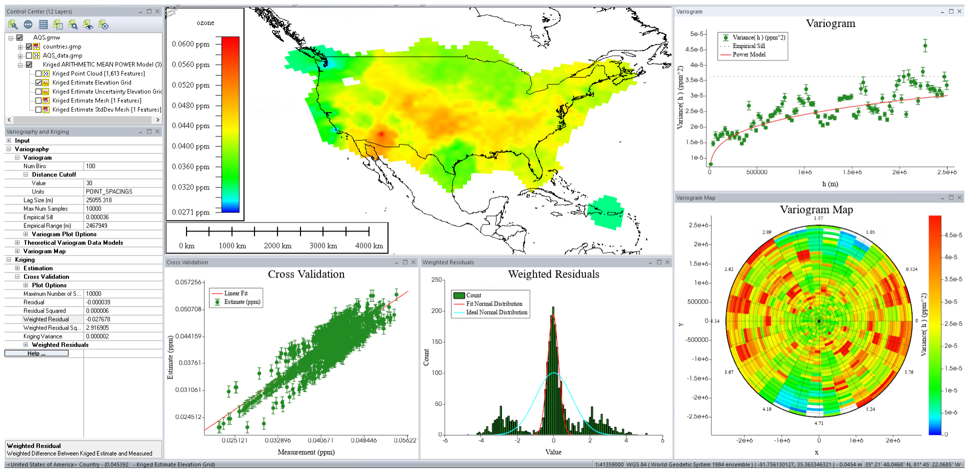

Figure 1: Variography and Kriging Tools Applied to Ozone Measurements in Global Mapper

2 Variogram and Semivariogram

The semivariogram, $\gamma(h)$, can be interpreted as the variance at separation distance h. For example, the variance in ozone concentration for all pairs of samples separated by a distance $h = 10$ km. A set of variance estimates at different separation distances is known as an empirical variogram. A theoretical variogram is found by fitting an empirical variogram to a data model as shown in Fig. 2.

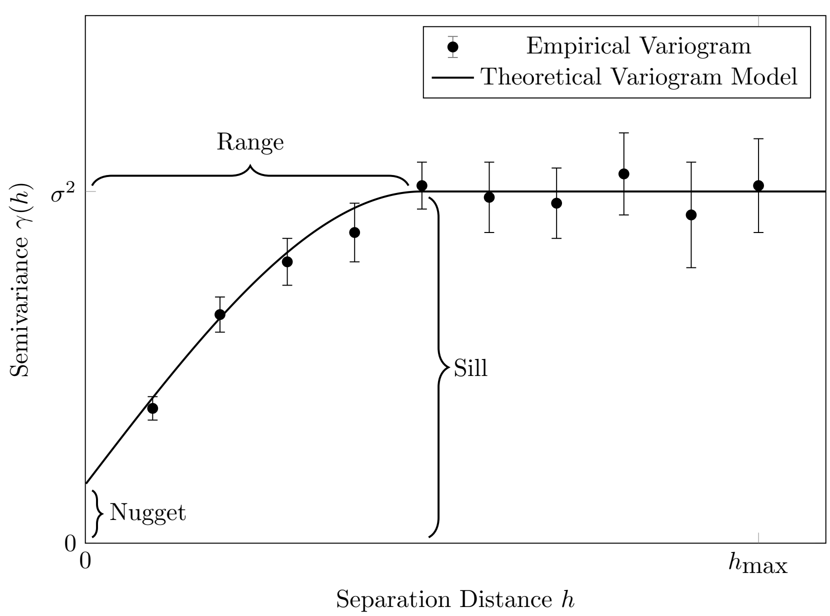

Figure 2: Notional Plot Relating the Empirical Variogram to a Theoretical Model

The range is the distance at which the data is no longer correlated. The sill is the value of the semivariance at distances greater than the range and corresponds to the variance of the data as a whole. The nugget represents the variance in the data at zero separation distance.

The semivariogram is defined as follows. Consider a set of measurements, $\{\mathcal{Z}(\boldsymbol{s}_i)\}$, taken at specified locations, $\{\boldsymbol{s}_i\}$. Let $\Delta\mathcal{Z}(\boldsymbol{s}_i, \boldsymbol{s}_j)$ be the measurement difference between measurement points $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$.

$$\Delta\mathcal{Z}(\boldsymbol{s}_i, \boldsymbol{s}_j) = \mathcal{Z}(\boldsymbol{s}_i) - \mathcal{Z}(\boldsymbol{s}_j) \tag{1}$$

The variogram, $2\gamma(\boldsymbol{s}_1, \boldsymbol{s}_2)$, is defined as the variance of the difference between field values at two locations.

$$2\gamma(\boldsymbol{s}_i, \boldsymbol{s}_j) = \mathrm{Var}\left(\Delta\mathcal{Z}(\boldsymbol{s}_i, \boldsymbol{s}_j)\right) = E\left[\left((\mathcal{Z}(\boldsymbol{s}_i) - \mu(\boldsymbol{s}_i)) - (\mathcal{Z}(\boldsymbol{s}_j) - \mu(\boldsymbol{s}_j))\right)^2\right] \tag{2}$$

If the mean is constant then $\mu(\boldsymbol{s}_i) = \mu(\boldsymbol{s}_j)$ and this reduces to

$$2\gamma(\boldsymbol{s}_i, \boldsymbol{s}_j) = E\left[\Delta\mathcal{Z}(\boldsymbol{s}_i, \boldsymbol{s}_j)^2\right] \tag{3}$$

Under assumptions of isotropy and stationarity the variogram can be written as a function of the separation distance $h = ||\boldsymbol{s_2} - \boldsymbol{s_1}||$.

$$\gamma(h) = \gamma(\boldsymbol{s_1}, \boldsymbol{s_2}) \tag{4}$$

$\gamma(h)$ is known as the semivariogram and is closely related to the measurement variance which corresponds to the value of the semivariogram at infinite distance.

$$\lim_{h \to \infty} \gamma(h) = \mathrm{var}\left(\mathcal{Z}\left(\boldsymbol{s}\right)\right) = \sigma^2 \tag{5}$$

2.1 Empirical Variogram

An estimate for the semivariogram at h can be constructed as an average over the set of measurement pairs, $N(h)$, separated by distance h.

$$N(h) = \{(\boldsymbol{s}_i, \boldsymbol{s}_j : |\boldsymbol{s}_i - \boldsymbol{s}_j| = h)\} \tag{6}$$

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{i,j \in N} \Delta\mathcal{Z}^2_{\boldsymbol{s}_i, \boldsymbol{s}_j} \tag{7}$$

The confidence of the estimate is reported as the standard deviation of the mean $\hat{\gamma}(h)$.

$$\sigma_{\hat{\gamma}(h)} = \frac{2}{\sqrt{|N(h)|}} \sum_{i,j \in N} \left(\frac{\Delta\mathcal{Z}^2_{\boldsymbol{s}_i, \boldsymbol{s}_j}}{2} - \hat{\gamma}(h)\right)^2 \tag{8}$$

For practical evaluation, h must be discretized and point pairs binned over a range of h values. Fig. 3 shows a typical variogram.

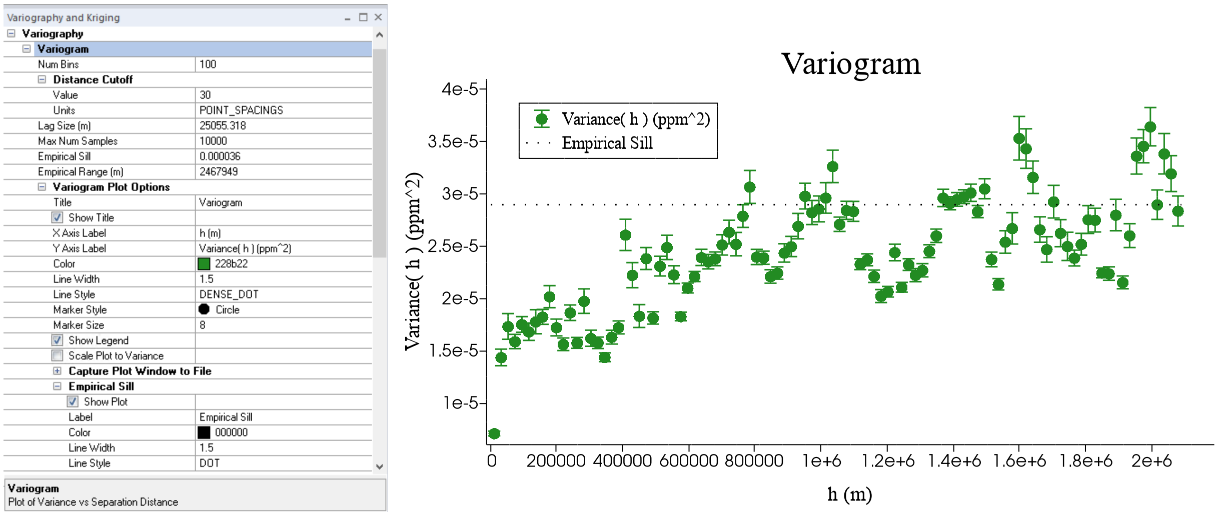

Figure 3: Example Empirical Variogram

2.1.1 Empirical Variogram Properties

- Number of Bins
  Number of distances at which the variogram is estimated
- Distance Cutoff
  The maximum value for the separation distance, h
- Lag Size
  The width of a range bin
- Maximum Number of Samples
  The maximum number of measurement pairs used to construct a variance estimate for a given separation distance
- Empirical Sill
  Value of the variance for distances greater than the correlation range, the maxumum variance
- Empirical Range
  The range at which the variance estimated stops changing signaling that the data is no longer correlated, the distance at which the sill is reached

2.2 Theoretical Variogram

Once the experimental variogram is estimated, a continuous expression is found from fitting the experimental values to a mathematical model. Models include

2.2.1 Data Models

- Linear (valid id $\mathbb{R}^d, d \geq 1$)

$$\gamma(\boldsymbol{h};\boldsymbol{\theta}) = \begin{cases} 0, & \boldsymbol{h} = \boldsymbol{0}, \\ c_0 + b_l||\boldsymbol{h}||, & \boldsymbol{h} \neq \boldsymbol{0}, \end{cases} \tag{9}$$

  where $\boldsymbol{\theta} = (c_0, b_l)$, $c_0 \geq 0$ and $b_l \geq 0$.

- Spherical (valid in $\mathbb{R}^1, \mathbb{R}^2$, and $\mathbb{R}^3$)

$$\boldsymbol{\gamma}(\boldsymbol{h};\boldsymbol{\theta}) = \begin{cases} 0, & \boldsymbol{h} = \boldsymbol{0}, \\ c_0 + c_s\left(\frac{3||\boldsymbol{h}||}{2a_s} - \frac{||\boldsymbol{h}||^3}{2a_s^3}\right), & 0 < ||\boldsymbol{h}|| \le a_s, \\ c_0 + c_s, & ||\boldsymbol{h}|| \ge a_s, \end{cases} \tag{10}$$

where $\boldsymbol{\theta} = (c_0, c_s, a_s)$, $c_0 \ge 0, c_s \ge 0$ and $a_s \ge 0$.

- Exponential (valid id $\mathbb{R}^d, d \ge 1$)

$$\boldsymbol{\gamma}(\boldsymbol{h};\boldsymbol{\theta}) = \begin{cases} 0, & \boldsymbol{h} = \boldsymbol{0}, \\ c_0 + c_e\left\{1 - \exp\left(\frac{-||\boldsymbol{h}||}{a_e}\right)\right\}, & \boldsymbol{h} \ne \boldsymbol{0}, \end{cases} \tag{11}$$

where $\boldsymbol{\theta} = (c_0, c_e, a_e)$, $c_0 \ge 0, c_e \ge 0$ and $a_e \ge 0$.

- Gaussian

$$\boldsymbol{\gamma}(\boldsymbol{h};\boldsymbol{\theta}) = \begin{cases} 0, & \boldsymbol{h} = \boldsymbol{0}, \\ c_0 + c_g\left\{1 - \exp\left(\frac{-||\boldsymbol{h}^2||}{a_g^2}\right)\right\}, & \boldsymbol{h} \ne \boldsymbol{0}, \end{cases} \tag{12}$$

where $\boldsymbol{\theta} = (c_0, c_g, a_g)$, $c_0 \ge 0, c_g \ge 0$ and $a_g \ge 0$.

- Power Function

$$\boldsymbol{\gamma}(\boldsymbol{h};\boldsymbol{\theta}) = \begin{cases} 0, & \boldsymbol{h} = \boldsymbol{0}, \\ c_0 + b_p||\boldsymbol{h}||^\lambda, & \boldsymbol{h} \ne \boldsymbol{0}, \end{cases} \tag{13}$$

where $\boldsymbol{\theta} = (c_0, b_p, \lambda)$, $c_0 \ge 0, b_p \ge 0$ and $0 \le \lambda \le 2$.

Fig. 4 shows data models and Fig. 5 shows a typical theoretical variogram.

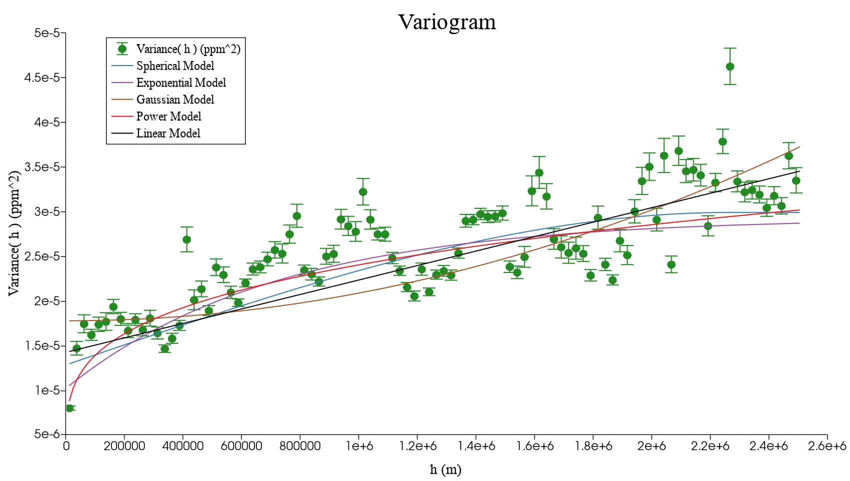

Figure 4: Data Models

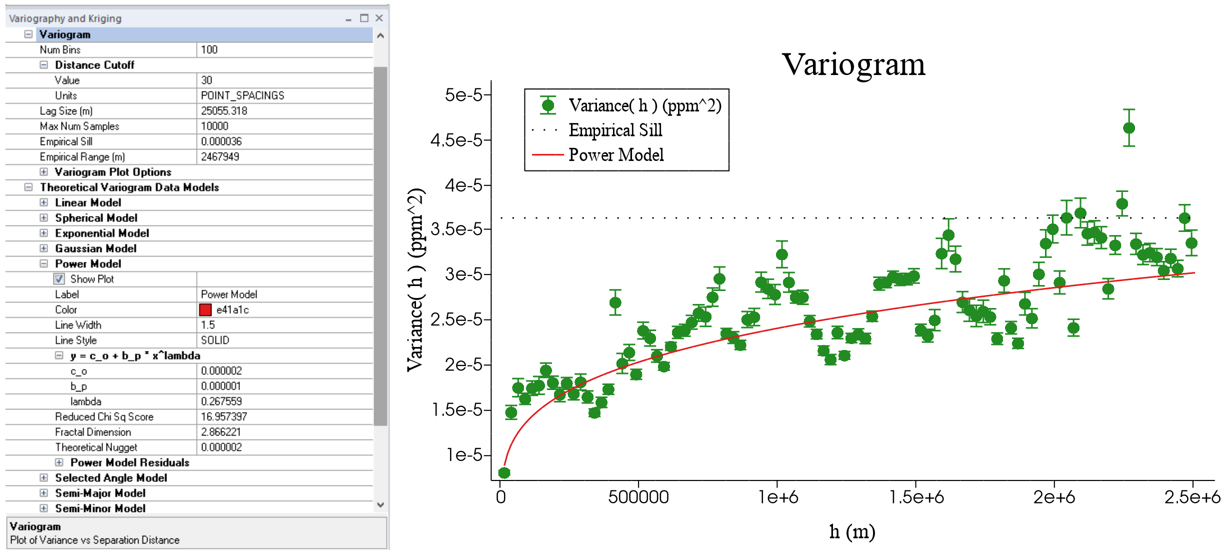

Figure 5: Example Theoretical Variogram

2.2.2 Data Model Properties

- Model Parameters
  Values for model parameters found by least squares fit
- Reduced Chi Sq Score
  Goodness of fit reported as a reduced chi-squared statistic
- Theoretical Sill
  Value of data model in the limit of large separation distance. Only for models that converge.
- Theoretical Range
  Range at which data model effectively becomes constant. Only for models that converge.
- Theoretical Nugget
  Value of data model at zero distance. Typically associated with measurement noise.

2.3 Variogram Map

Directional dependence can be explored with the variogram map which plots the variogram as a function of angle. Fig. 6 shows a variogram map.

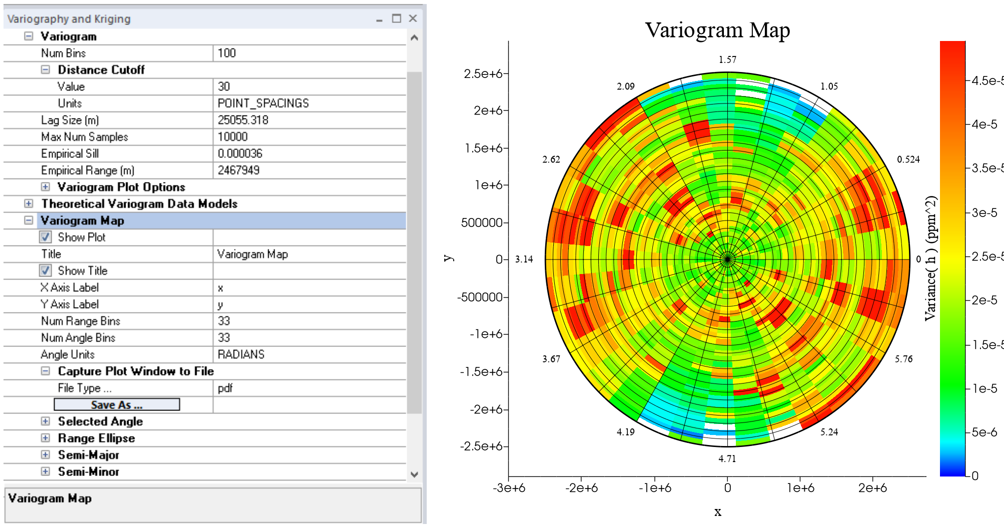

Figure 6: Example Variogram Map

2.3.1 Variogram Map Properties

- Number of Range Bins
  Number of distances at which the variogram is estimated in any direction

- Number of Angle Bins
  Number of angle values at which the variogram is estimated

- Angle Units
  Radians or degrees

2.3.2 Selected Angle

The variance in a specific direction can be picked out and plotted alongside the global (averaged over all directions) variogram. Fig. 7 shows a variogram map.

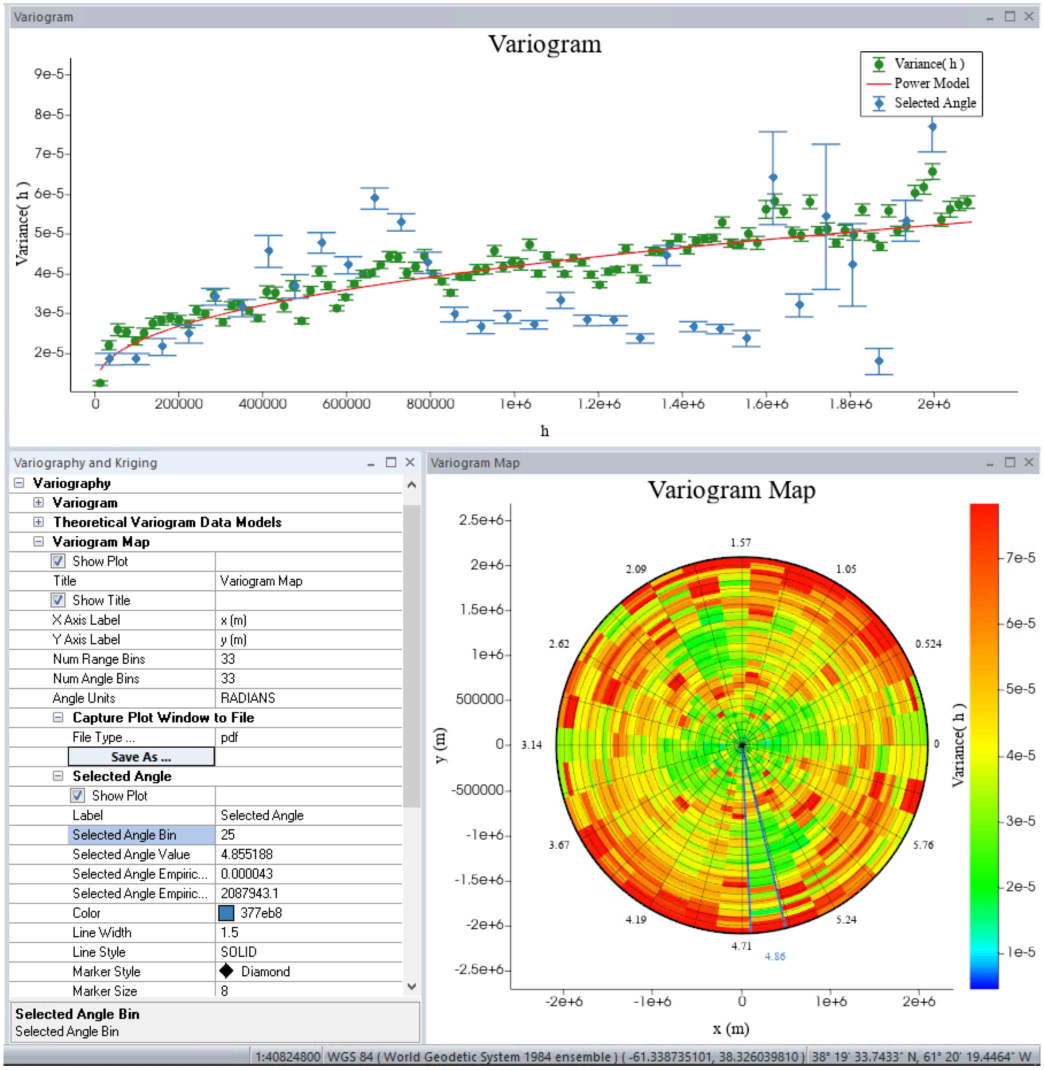

Figure 7: Directional Variogram Along Selected Angle

2.3.3 Selected Angle Properties

- Selected Angle Bin
  Index of angle bin used for directional variogram

- Selected Angle Value
  Value of angle corresponding to selected bin

- Selected Angle Empirical Sill
  Empirical sill for directional variogram at given angle

- Selected Angle Empirical Range
  Empirical range for directional variogram at given angle

3 Kriging

Kriging is a statistical estimation method that produces an optimal estimate of a random process under assumptions of stationarity and Gaussian statistics. Relative to deterministic methods, Kriging has the advantage that it naturally provides an estimate of uncertainty and is amenable to cross validation.

Consider a set of measurements, $\mathcal{Z}(s_i)$, where s_i are the observed locations. Estimates at arbitrary locations can be produced as weighted averages over nearby measurements.

$$\hat{\mathcal{Z}}(s_0) = \sum_{i=1}^{N} \lambda_i \mathcal{Z}(s_i) \tag{14}$$

Where $\mathcal{Z}(s_i)$ is the measured value at point s_i, λ_i is a statistical weight, s_0 is the prediction location and N is the number of measured values. The weights, λ_i, are chosen s.t. variance is a minimum and there is no bias.

The process by which Kriging is used to estimate values is

1. Build empirical variogram from input data to characterize spatial correlation
2. Construct theoretical variogram via fitting empirical variogram to an appropriate model
3. Produce estimates (with variance) at desired interpolation points by solving the Kriging system of equations.

3.1 Ordinary Kriging

This section describes the mathematics of ordinary Kriging in sufficient detail to support high level understanding and implementation. The estimation model assumption for ordinary Kriging is

$$\mathcal{Z}(\boldsymbol{s}) = \mu + \delta(\boldsymbol{s}), \quad \boldsymbol{s} \in D, \mu \in \mathcal{R} \text{ and } \mu \text{ unknown} \tag{15}$$

Optimal weights for 14 are determined by minimization of the variance, σ^2, subject to the constraints imposed by stationarity and lack of bias.

$$\sigma^2 \equiv E\left(\mathcal{Z}(\boldsymbol{s}) - \hat{\mathcal{Z}}(\boldsymbol{s})\right)^2 \tag{16}$$

This results in the Kriging system of equations [1] from which the optimal weights can be determined

$$\underbrace{\begin{pmatrix} \gamma(\boldsymbol{s}_1-\boldsymbol{s}_1) & \dots & \gamma(\boldsymbol{s}_1-\boldsymbol{s}_N) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(\boldsymbol{s}_N-\boldsymbol{s}_1) & \dots & \gamma(\boldsymbol{s}_N-\boldsymbol{s}_N) & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix}}_{\Gamma_0} \underbrace{\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_N \\ \mu \end{pmatrix}}_{\boldsymbol{\lambda}_0} = \underbrace{\begin{pmatrix} \gamma(\boldsymbol{s}_1-\boldsymbol{s}_0) \\ \vdots \\ \gamma(\boldsymbol{s}_N-\boldsymbol{s}_0) \\ 1 \end{pmatrix}}_{\boldsymbol{\gamma}_0} \tag{17}$$

$$\boldsymbol{\lambda}_0 = \Gamma_0^{-1}\boldsymbol{\gamma}_0 \tag{18}$$

Where γ is a valid theoretical variogram, μ is a Lagrange multiplier and $\{\lambda_i\}$ are the Kriging weights. Solving for the weights gives

$$m = -\left(\frac{1 - \mathbf{1}^T\Gamma^{-1}\boldsymbol{\gamma}}{\mathbf{1}^T\Gamma^{-1}\mathbf{1}}\right) \tag{19}$$

and

$$\boldsymbol{\lambda}^T = (\boldsymbol{\gamma} - \mathbf{1}m)^T\,\Gamma^{-1} \tag{20}$$

where $\boldsymbol{\gamma}, \boldsymbol{\lambda}$ and Γ (without the zero subscript) are the $1, N$ elements of the variables in Eq. 18 and $\mathbf{1}$ is an $N \times 1$ matrix containing a row of ones. Substituting back into the variance gives the Kriging variance.

$$\begin{aligned} \sigma^2(\boldsymbol{s}_0) = \boldsymbol{\lambda}_0^T\boldsymbol{\gamma}_0 &= \sum_{i=1}^{N} \lambda_i \gamma(\boldsymbol{s}_o - \boldsymbol{s}_i) + m \\ &= \boldsymbol{\gamma}^T\Gamma^{-1}\boldsymbol{\gamma} - \tfrac{\left(\mathbf{1}^T\Gamma^{-1}\boldsymbol{\gamma}-1\right)^2}{\mathbf{1}^T\Gamma^{-1}\mathbf{1}} \\ &= 2\sum_{i=1}^{n} \lambda_i\gamma(\boldsymbol{s}_0 - \boldsymbol{s}_i) - \sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j\gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) \end{aligned} \tag{21}$$

Fig. 8 shows typical Kriging results.

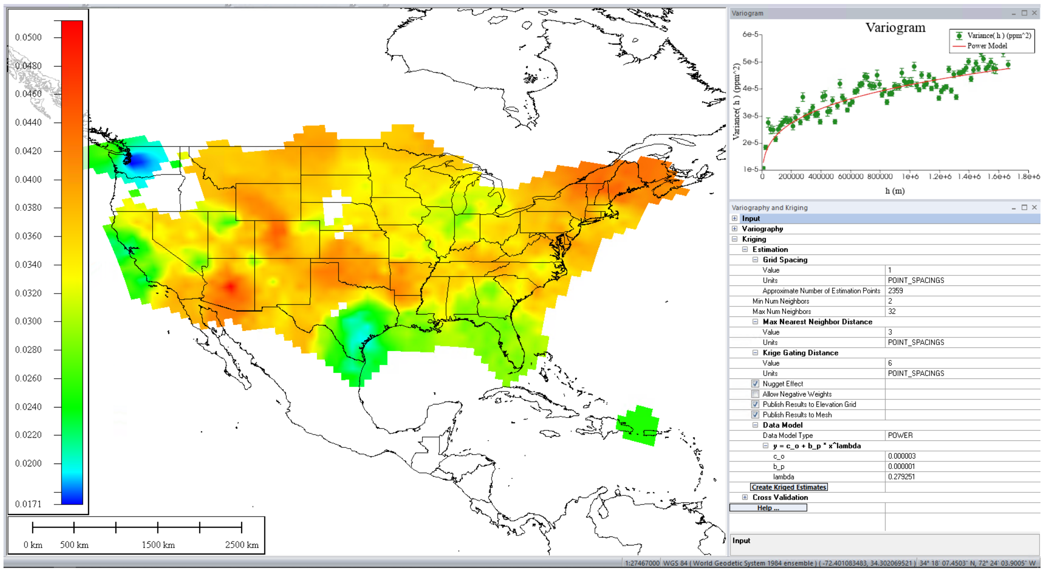

Figure 8: Kriged Estimate Grid

3.2 Kriging Estimation Properties

- Grid Spacing
  Spacing between kriged estimates

- Minimum Number of Neighbors
  Minimum number of neighboring measurements required for construction of a kriged estimate

- Maximum Number of Neighbors
  Maximum number of neighboring measurements considered in construction of a kriged estimate

- Maximum Nearest Neighbor Distance
  Maximum distance to nearest neighboring measurement allowed for construction of a kriged estimate

- Krige Gating Distance
  Distance within neighboring measurements are collected

- Nugget Effect
  Under the assumption that the variogram, $\gamma(d)$, is zero at $d = 0$, the Kriging method is an "exact interpolator". Exact interpolation refers to the property that the estimation procedure does not change the values of the measurements at the measured locations. The limit, $\lim_{h \to 0} \gamma(h)$, is called the "nugget" and is not zero in general. Enforcing the exact interpolator property forcing the data model at zero results in a discontinuity in the variogram and introduces numerical instability. Relaxing this property by allowing $\gamma(0)$ to be equal to the nugget can give smoother results but does not retain the exact interpolator property.

- Allow Negative Weights
  Allow negative kriging weights which can sometimes result in numerical instability

- Data Model
  Theoretical variogram data model used for construction of kriged estimates

3.3 Cross Validation

Validation of Kriging model is typically done by cross validation which is performed as follows. Given a measurement set $\{\mathcal{Z}(s_i)\}$, an estimate, $\hat{\mathcal{Z}}(s_i)$, at each measurement location , s_i, is produced based on all points except the measurement at that location. Fig. 9 shows cross validation parameters and a plot of estimate, measurement pairs.

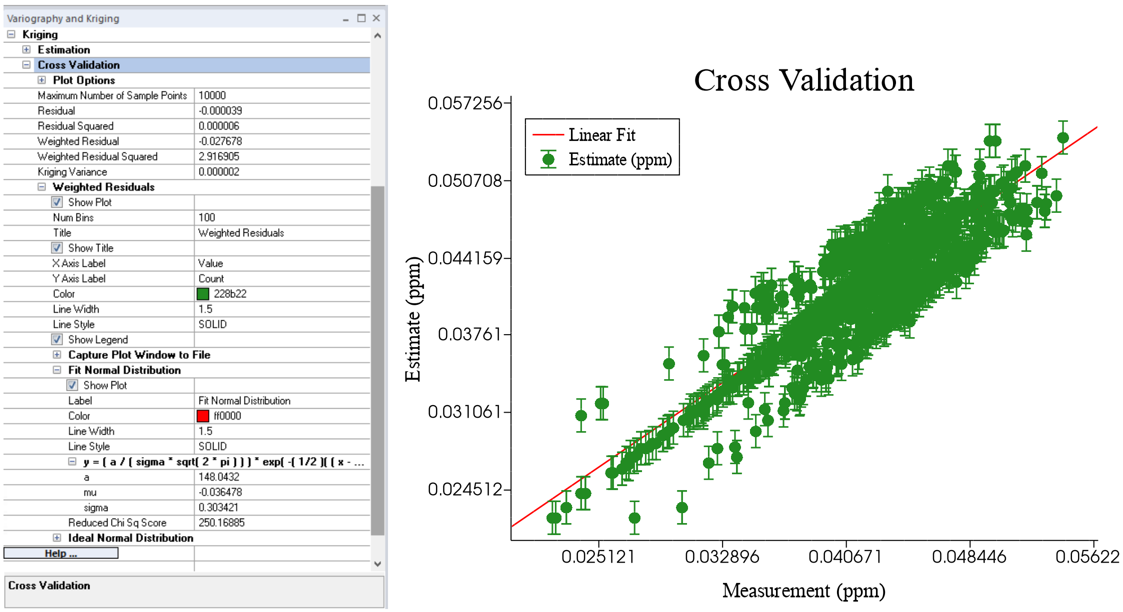

Figure 9: Cross Validation

For each estimate, measurement pair, $\left\{\hat{\mathcal{Z}}(s_i), \mathcal{Z}(s_i)\right\}$, the weighted residual , $\Delta\hat{\mathcal{Z}}(s_i)$, is defined as

$$\Delta\hat{\mathcal{Z}}(s_i) = \frac{\hat{\mathcal{Z}}(s_i) - \mathcal{Z}(s_i)}{\sigma(s_i)} \tag{22}$$

If data model assumptions hold, the weighted residuals will be normally distributed with zero mean and unit variance. Fig. 10 shows a residual distribution plot.

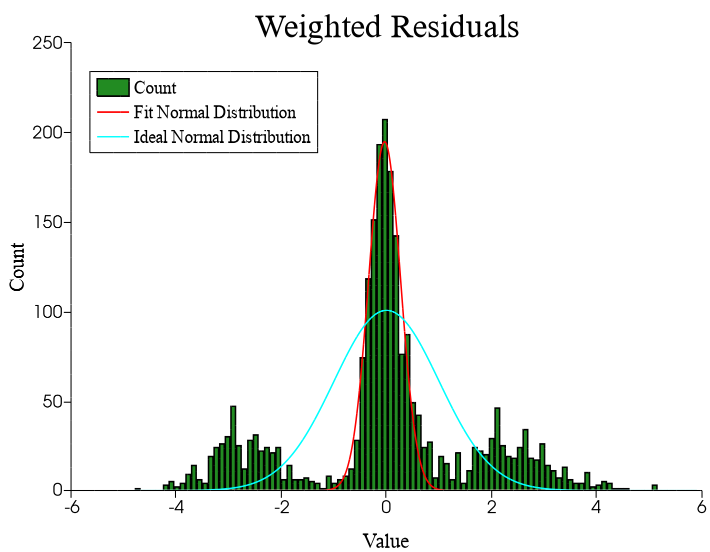

Figure 10: Distribution of Weighted Residuals

3.3.1 Cross Validation Properties

- Minimum Number of Sample Points
  Number of estimate, measurement pairs evaluated in cross validation
- Averages reported
  - Residual
  $$\frac{1}{n}\sum_{i=1}^{n}\hat{\mathcal{Z}}(s_i) - \mathcal{Z}(s_i) \tag{23}$$
  - Residual Squared
  $$\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\mathcal{Z}}(s_i) - \mathcal{Z}(s_i)\right)^2 \tag{24}$$
  - Weighted Residual
  $$\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\mathcal{Z}}(s_i) - \mathcal{Z}(s_i)}{\sigma(s_i)} \tag{25}$$
  - Weighted Residual Squared
  $$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{\mathcal{Z}}(s_i) - \mathcal{Z}(s_i)}{\sigma(s_i)}\right)^2 \tag{26}$$
  - Kriging Variance
  $$\frac{1}{n}\sum_{i=1}^{n}\sigma^2(s_i) \tag{27}$$

References

[1] N.A.C. Cressie. *Statistics for spatial data*. Wiley series in probability and mathematical statistics: Applied probability and statistics. J. Wiley, 1993.