Kriging

Chris Sobczak

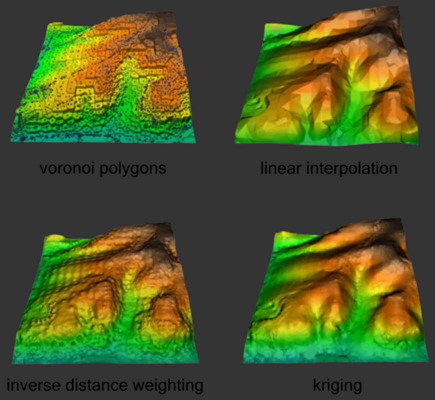
Motivation

- The First Law of Geography:
  - "Everything is related to everything else, but near things are more related than distant things." -Waldo Tobler

Figure: Waldo Tobler

Geostatistics

- A statistical technique based on the theory of random *spatial processes*
- The key trait of a geostatistical problem is *spatial correlation*
- Similar to time series problems where observations close in time are more correlated, observations in geostatistics problems that are close in geography are more correlated
- Examples:
  - Elevation
  - Ore grade
- This Springer textbook was very helpful to understand the concepts covered below (Oliver, Webster, et al. 2015)

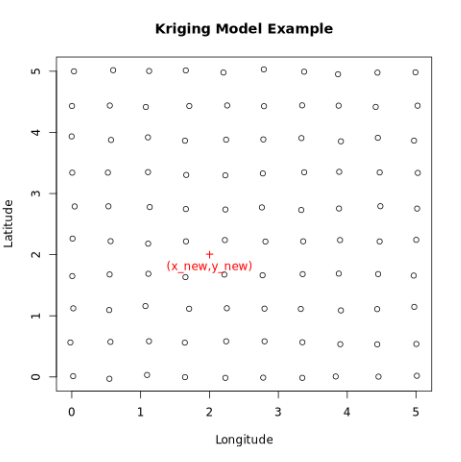


Danie G. Krige

- Danie G. Krige was a South African engineer
- He worked in the mining industry surveying, sampling and doing ore valuation
- The problem set up:
    - I want to know where the gold is
    - We can take cores of the area at a variety of depths at a variety of locations
    - From these cores/probes, can we know if it's worth mining in this geography?
- His master's thesis (Krige 1951) introduced a statistical approach that George Matheron (Matheron 1960) formalized into the method of *Kriging*

Definitions

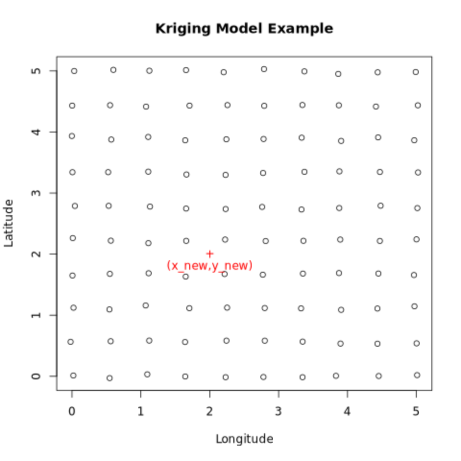
- Let X be the matrix of all *locations* of observations (longitude, latitude, altitude, depth, ...)
- Define $z(X)$ as the vector of observations we recorded at all locations X (the response variable)
- $z(x_i)$ is the value of interest at location x_i
  - The $z(x_i)$ and $z(x_i + h)$ are correlated as a function of h, the distance between x_is
- Kriging generally requires z to be a stationary spatial process (like in time series, transformations can be used)



Model

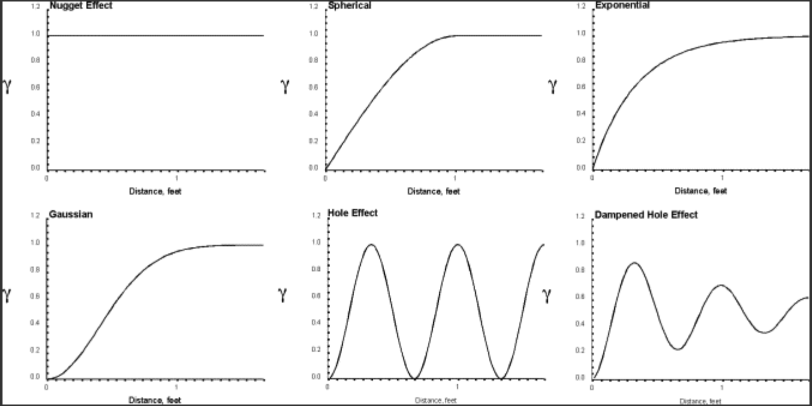
$$\hat{z}(x_{new}) = \lambda^T z(X) \quad (1)$$
$$z(x_{new}) = \lambda^T z(X) + \epsilon \quad (2)$$
$$z(x_{new}) = \lambda_1 z(x_1) + \cdots + \lambda_n z(x_n) + \epsilon \quad (3)$$

- $\hat{z}(x_i)$ is the BLUE (or BLUP) of z
- $\hat{z}(x_i)$ is a linear combination of weights and the observed $z(x_i)$s: $\lambda^T z(X)$
- Since $z(x_i)$s close to each other are correlated, $z(x_{new})$ should take more weight/information from the closest points
- **How do we assign the weights?**



The Variogram

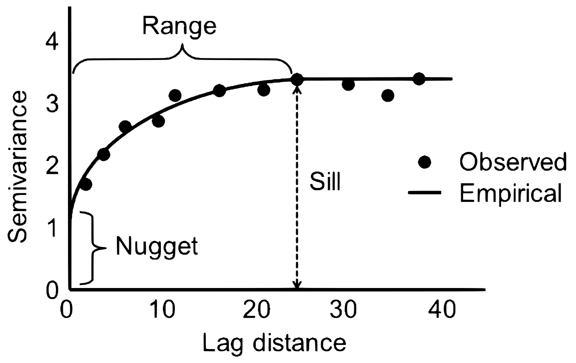
- The variogram is a function that quantifies the spatial autocorrelation of a spatial process
- Recall the *autocorrelation function* in time series



The Variogram

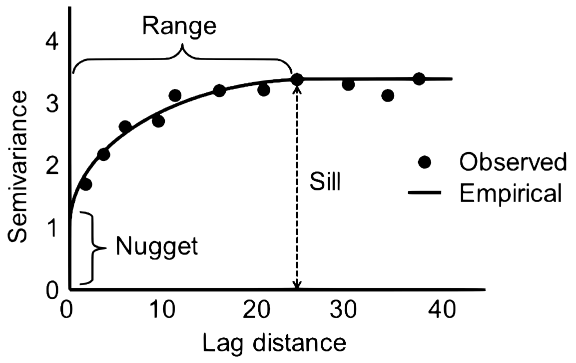
$$\hat{\gamma}(h) = \frac{1}{2m(h)} \sum_{i=1}^{m(h)} [z(x_i) - z(x_i + h)]^2 \qquad (4)$$

- This is Matheron's method of moments estimator for the sample semivariance (Matheron 1960) for one dimension
- h is a distance between x_is
- $m(h)$ is the number of paired comparisons at lag h (they are bins)



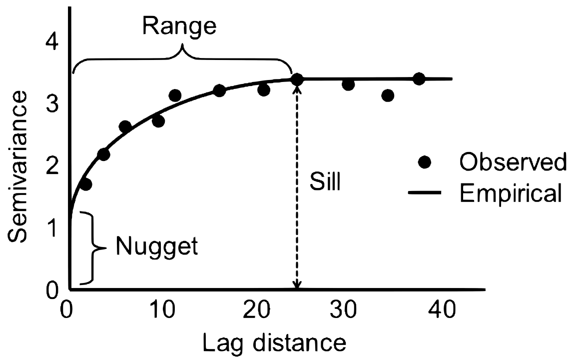
Variogram Terms

- **Nugget**: The variogram value at $h = 0$ (h is the lag distance)
- **Range**: The h value where the variogram levels off, interpreted as the maximum distance where points are correlated
- **Sill**: The value of the variogram at the **range**, representing the total variance of the data
- The relationship between these terms have significant influence over the weights $\hat{\lambda}$ (described later)



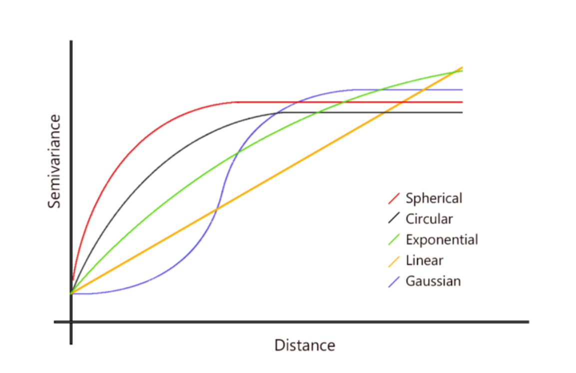
Variogram Terms

- The nugget is related to the variance of the estimates
  - In other words, the nugget is the difference between y at the same location (you would expect this to be zero)
  - *Punctual kriging* sets $\hat{y}_i = y_i$ where x_i is part of the sampled data X
- The ratio between the nugget and the sill is discussed later as they influence $\hat{\lambda}$



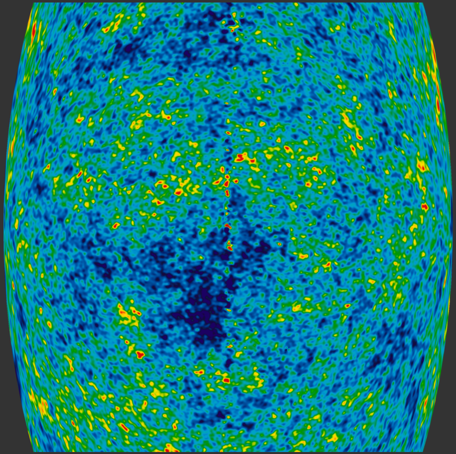
Model Choices for $\gamma(h)$

- Different models for the theoretical variogram:
  - **Spherical**: Common in geosciences, gradual transition in spatial correlation
  - **Exponential**: Spatial correlation decreases exponentially in h
  - **Gaussian**: Common in environmental studies, strong local spatial correlation (semivariance stays low for longer than the other models)
  - **Linear**: Direct proportional increase in semivariance over distance



Anisotropy

- Anisotropy is the property of a *surface* identified by differences in variation as you travel in different directions
- As opposed to *isotropy*
- Think of the globe, and temperature:
  - As you walk north it gets cold, variation increases
  - But as you walk east, it doesn't as much, or at all
  - As you walk south, it gets warmer
- If your surface is anisotropic, you might construct separate variograms for different directions



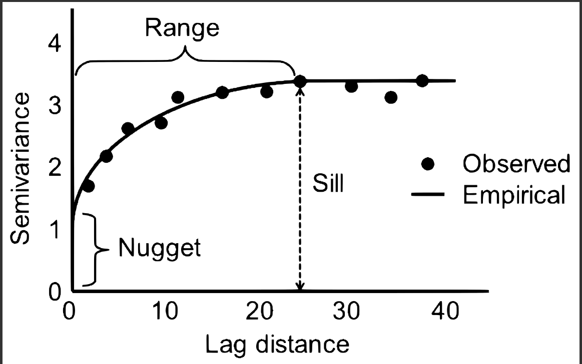
Using The Variogram to Find $\hat{\lambda}$

$$\hat{\lambda} = A^{-1}b \tag{5}$$

$$A = \hat{\gamma}(X) \tag{6}$$

$$b = \hat{\gamma}(x_{new}, X) \tag{7}$$

- $\hat{\lambda}$ is the vector of estimated weights
- A is a matrix of all values of the variogram $\hat{\gamma}$, given the sample dataset
- b is a vector of values of $\hat{\gamma}$ between x_{new} and all X from the sample
- This is very computationally heavy as the sample increases



Effects of The Variogram on $\hat{\lambda}$

- Figure from Oliver, Webster, et al. 2015 shows different nugget:sill ratios and influence on $\hat{\lambda}$
- Numbers on the dots are the $\hat{\lambda}_i$ at those x_i, circle in the middle is x_{new}
- If the variogram is flat (**figure d**), the weights are evenly spread out
- If the variogram has a wide differential from the nugget to sill (figure a), then weights are more concentrated around x_{new}

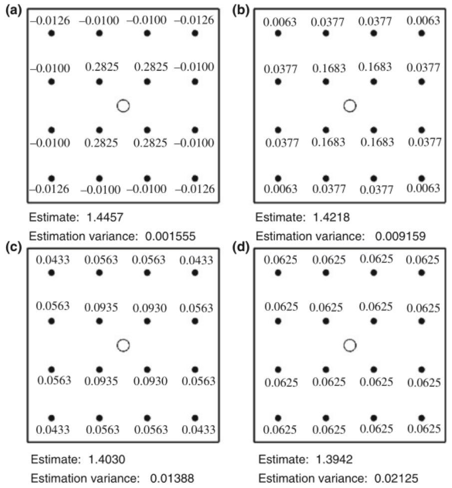


Fig. 4.1 Punctual kriging weights for 16 (4 × 4) $\log_{10}$ K^+ data from Broom's Barn Farm (Suffolk, England) based on a spherical function with a range of 426 m. The nugget:sill ratio was changed as follows: **a** $c_0 = 0{:}c = 0.02$, **b** $c_0 = 0.004{:}c = 0.016$, **c** $c_0 = 0.012{:}c = 0.008$ and **d** $c_0 = 02{:}c = 0$

Notes on the Model

$$\sum_{i=1}^{n} \lambda_i = 1 \Rightarrow \mathbb{E}\left[\hat{z}(x_{new}) - z(x_{new})\right] = 0 \tag{8}$$

$$\text{Var}(\hat{z}(x_{new})) = \mathbb{E}\left[\left(\hat{z}(x_{new}) - z(x_{new})\right)^2\right] \tag{9}$$

$$= 2\sum_{i=1}^{n} \lambda_i \gamma(x_i - x_{new}) - \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \gamma(x_i - x_j) \tag{10}$$

- The weights sum to 1 (making the model unbiased)
- This estimate of the error of interpolation distinguish kriging from the several other methods
- The variance of predictions are dependent on the prediction itself and the location (distance from a known point is important)

Notes on the Model

- Kriging is a local predictor
- One can choose to include only a subset of the closest points in X
- Then the system can operate within a moving window when mapping a surface
- This can make computation more realistic, especially if a complete surface is to be mapped

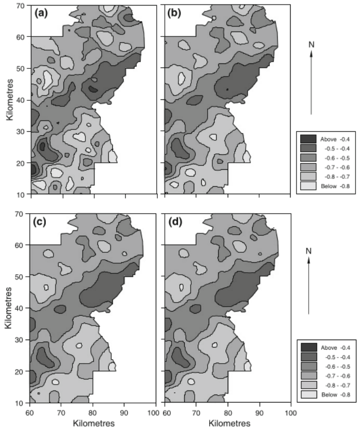
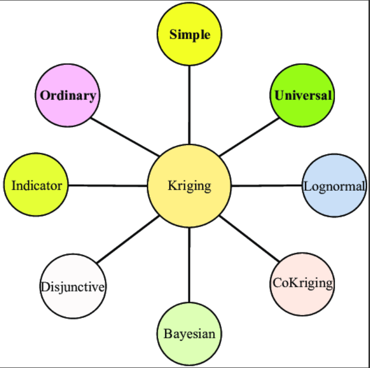


Fig. 4.6 Maps of ordinary punctually kriged predictions of $\log_{10}$Co in the Borders Region of Scotland with the kriging neighbourhood set as follows: **a** 20 maximum points, 17 minimum points, **b** 40 maximum points, 37 minimum points, **c** 80 maximum points, 77 minimum points and **d** 140 maximum points, 137 minimum points

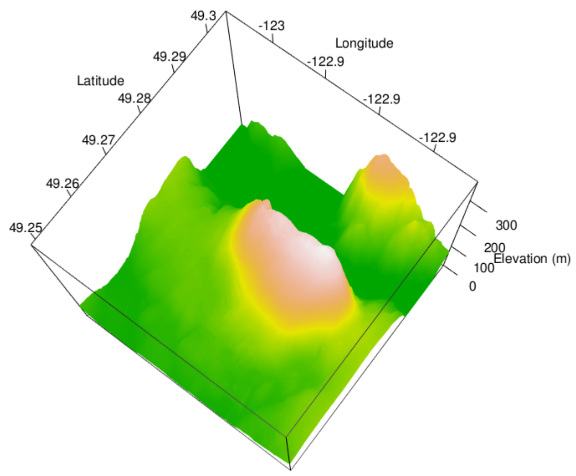
Varieties of Kriging

- Method described above is known as *ordinary kriging*
- This is different from *simple kriging* where a mean term is introduced where stronger stationarity is identified
- Other varieties are designed to manage:
  - Anisotropy
  - Lognormal distributed environmental data
  - Binary response data
  - Additional covariates
  - Others



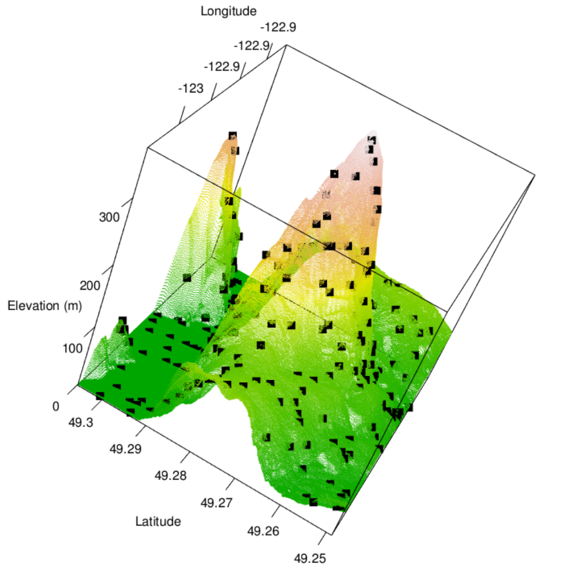
Burnaby Mountain

- Used R packages `sf`, `elevatr` and `terra` to download topography data for a window around Burnaby Mountain, BC
- This is our *true surface* from which I have a grid of true elevation values



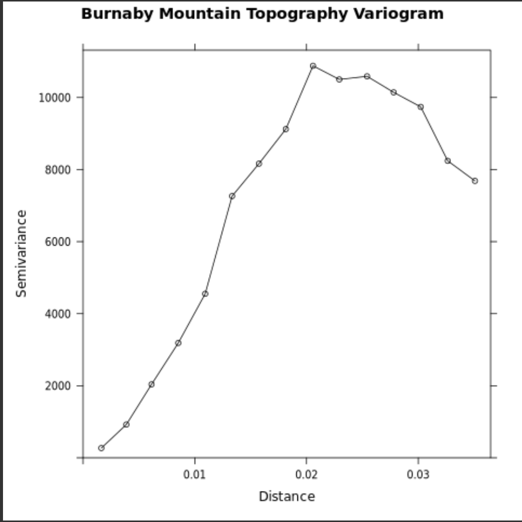
Sample X

- Took a sample of 200 from the original grid
- Selected points are shown on the right as the black squares
- Lets say we want to predict the elevation of Blusson Hall and Burnaby Mountain Park:
  - x_{bh} = Blusson Hall: -122.91264, 49.27898
  - x_{bmp} = Burnaby Mountain Park: -122.934742, 49.281914



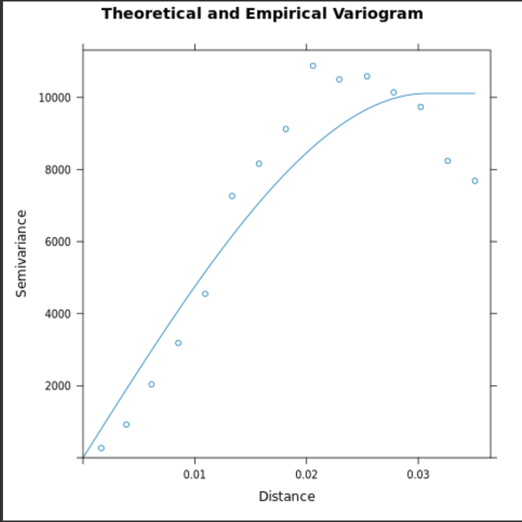
Empirical Variogram

- Looks like the sill is just over 10,000
- The range is near 0.02 (degrees)
- And there is no nugget
- The initial plot helps us pick start values to model the theoretical variogram, using `fit.variogram` (Pebesma 2004) to find the best values:
  - Nugget: 0
  - Sill: 10,112.63
  - Range: 0.03080463



Fitted Theoretical Variogram

- Looks like the sill is just over 10,000
- The range is near 0.02 (degrees)
- And there is no nugget
- The initial plot helps us pick start values to model the theoretical variogram, using `fit.variogram` (Pebesma 2004) to find the best values:
  - Nugget: 0
  - Sill: 10,112.63
  - Range: 0.03080463



Predictions

After fitting the model with `gstat`, these are the results:

Blusson Hall

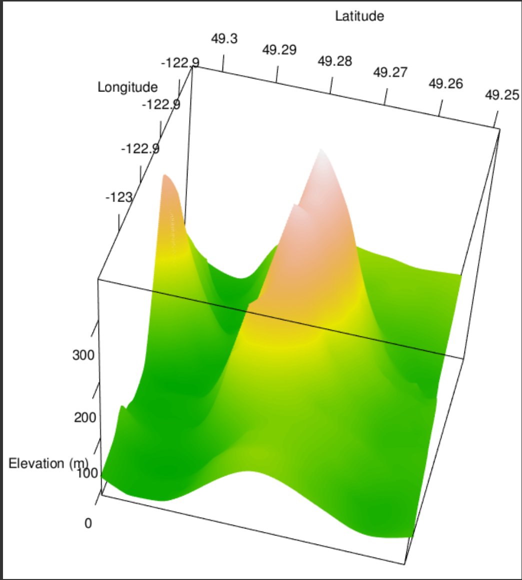
- $\hat{y} = 338.95$
- $\text{Var}(\hat{y}) = 840.0$, this is quite large due to the variance over the entire grid, relative to the size of the grid
- From the true grid from topography data, $y = 357$ (meters), so our actual error is about 20 meters

Burnaby Mountain Park (BMP)

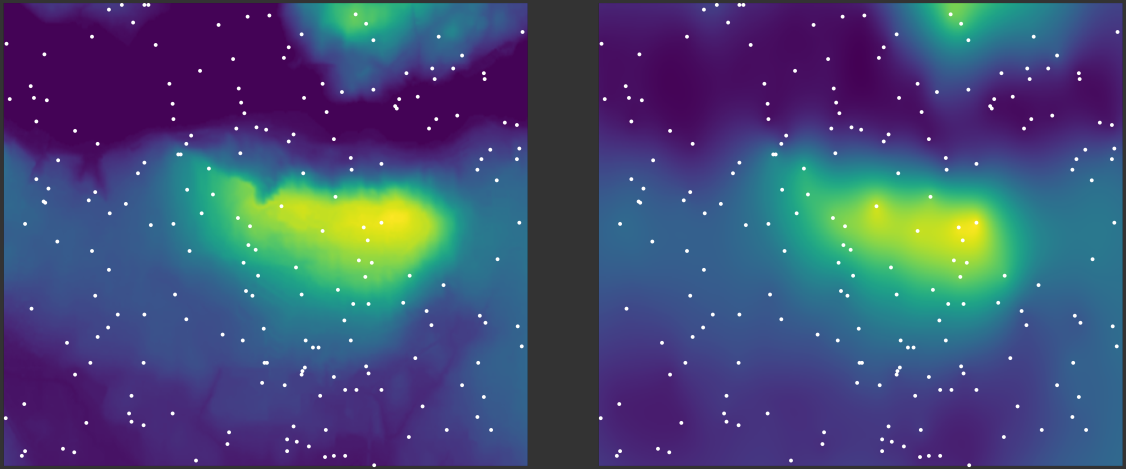
- $\hat{y} = 245.74$
- $\text{Var}(\hat{y}) > 1500$, this is even larger because BMP is even futher from any of our sample points
- From the true grid from topography data, $y = 296$ (meters), so our actual error is about 50 meters

Interpolating the Whole Grid

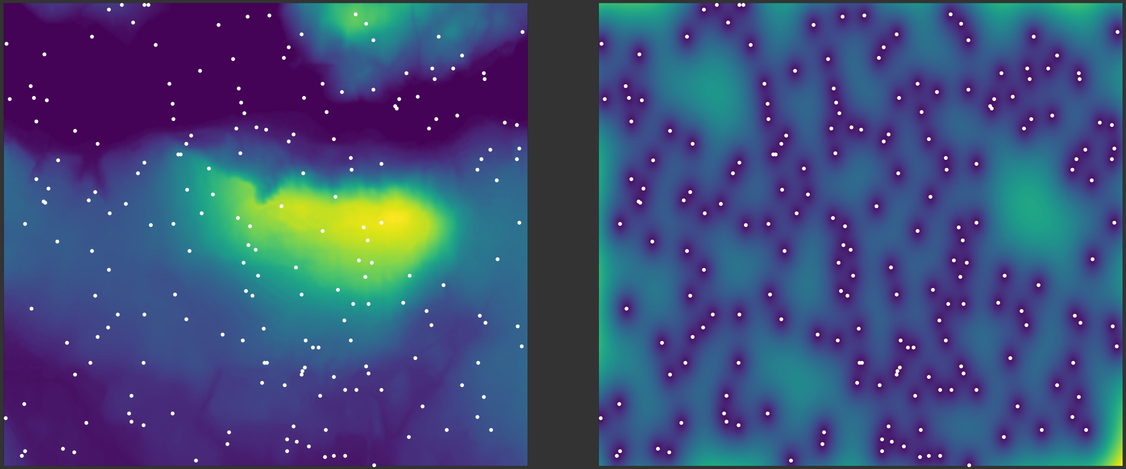
- The figure on the right shows the interpolated surface with the basic model fitted above
- Following slides show the variance and errors at different locations
- Variance increases where you are further from sampled points



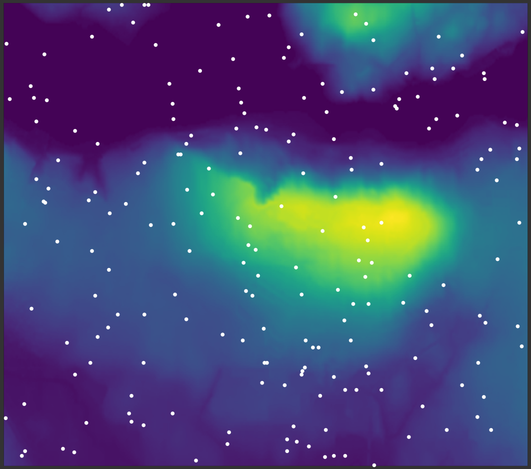
Interpolating the Whole Grid

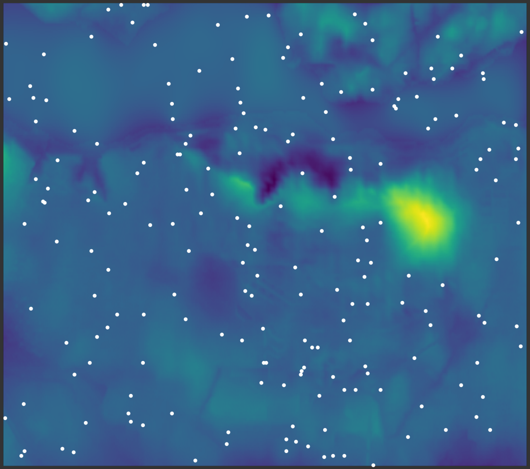


Prediction Variance



Prediction Errors





Health Example

- More than geostatistics, spatial processes can appear elsewhere
- Figure to the right shows a response surface plotted against two contaminants
- Similar response surfaces might exist for other continuous health factors

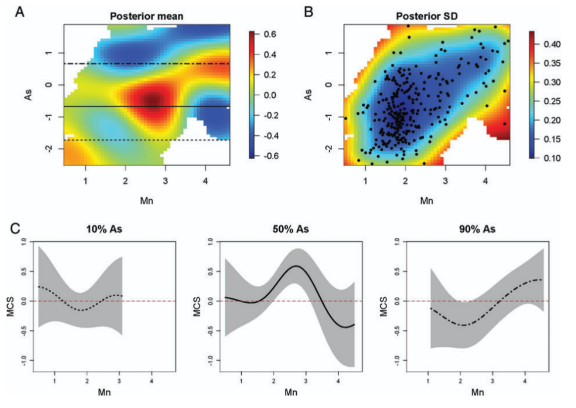


Figure: Bobb et al. 2014

Kriging Workflow Summary

- Data exploration
  - Assess stationarity
  - Assess anisotropy
  - Apply transformations
- Plot the experimental variogram
- Fit a theoretical variogram
- Fit the model using chosen variogram
- Predit values
- Helpful minimal example can be found here (Deshmukh n.d.)

Key Points

- Kriging is the benchmark spatial process prediction method
- The kriging predictor is a weighted average of the sample with more weight on the closest points
- The variogram defines the weights
- There are optimizations to be iteratively made to fit your data to a variogram, adjusting how the weights are chosen
- The variance of the predictor is dependent on the distance from the surrounding sample points

Questions?

Bibliography I

- Krige, D.G. (1951). "A statistical approach to some basic mine valuation problems on the Witwatersrand". In: Journal of the Southern African Institute of Mining and Metallurgy 52.6, pp. 119–139. DOI: 10.10520/AJA0038223X_4792. eprint: https://journals.co.za/doi/pdf/10.10520/AJA0038223X_4792. URL: https://journals.co.za/doi/abs/10.10520/AJA0038223X_4792.
- Matheron, Georges (1960). "Krigeage d'un panneau rectangulaire par sa périphérie". In: Note géostatistique 28.
- Pebesma, Edzer J. (2004). "Multivariable geostatistics in S: the gstat package". In: Computers and Geosciences 30, pp. 683–691. URL: https://doi.org/10.1016/j.cageo.2004.03.012.

Bibliography II

- Bobb, Jennifer F. et al. (Dec. 2014). "Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures". In: Biostatistics 16.3, pp. 493–508. ISSN: 1465-4644. DOI: 10.1093/biostatistics/kxu058. eprint: https://academic.oup.com/biostatistics/article-pdf/16/3/493/25417184/kxu058.pdf. URL: https://doi.org/10.1093/biostatistics/kxu058.
- Oliver, Margaret A, Richard Webster, et al. (2015). Basic steps in geostatistics: the variogram and kriging. Vol. 106. Springer.
- Deshmukh, Dr. Ankit (n.d.). Understanding Variogram in Geospatial Analysis — ankitdeshmukh.com. https://ankitdeshmukh.com/post/2025-01-17-variogram/. [Accessed 24-11-2025].