Kriging

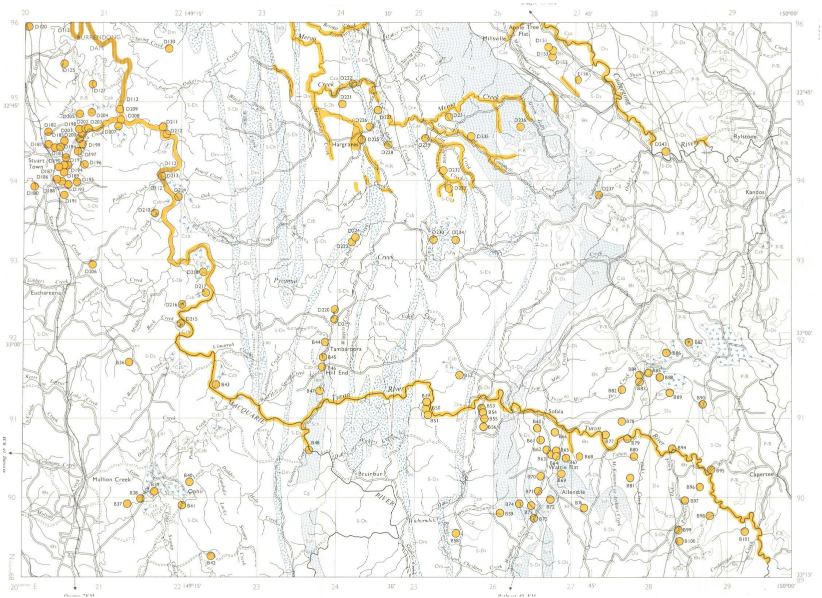
Statistical Learning Reading Group

Udo Boehm

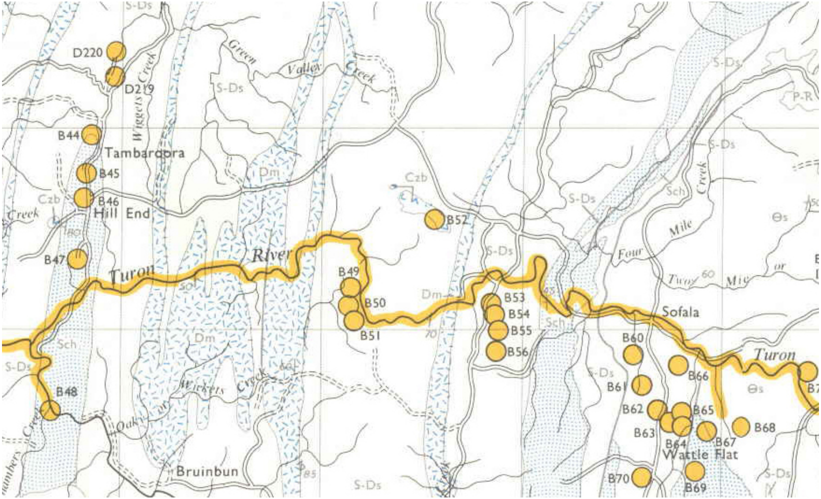
14 March 2018

Gold Digging in Sofala, NSW

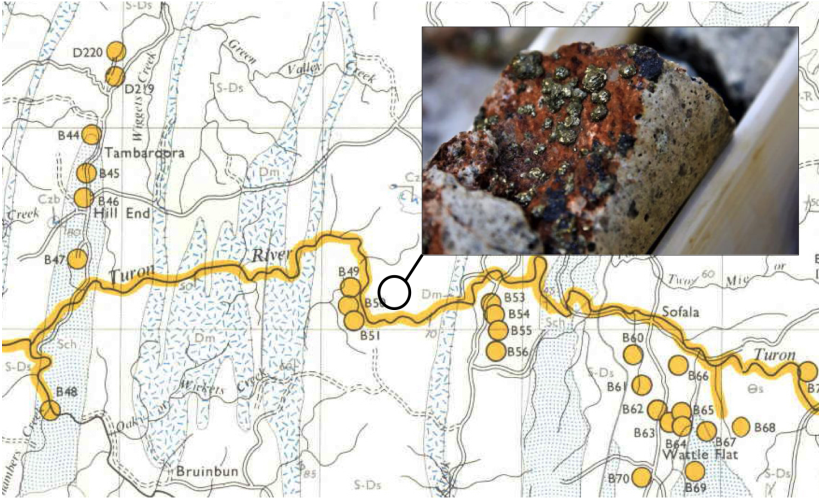
Gold Digging in Sofala, NSW



Gold Digging in Sofala, NSW



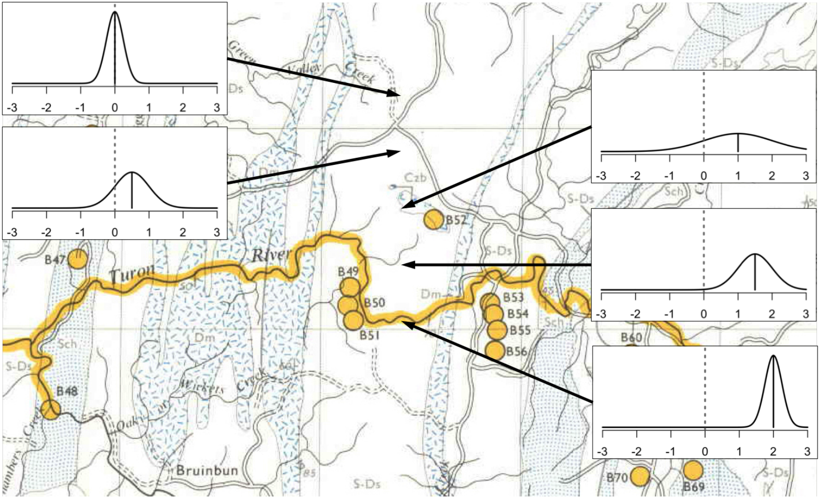
Gold Digging in Sofala, NSW



History of Kriging

- Gold mining in South Africa in the 1940s
- Mines established in locations with surface exposure of ore ⇒ biased sampling
- Sample mean of assays times estimated ore-body volume used to predict recoverable ore
- Sample standard deviation used to estimate variability in ore quality throughout the ore body
- D. G. Krige in the 1950s noted three flaws:
  - Gold-assay data are log-normal
  - Local variability (block grade) is lower than global variability (core sample grade)
  - Block grade and core sample grade are correlated
- Similar techniques and models developed in meteorology, forestry, physics, geodesy

Gold Digging in Sofala, NSW



Gaussian Process

- Observations $Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_N)$ at locations $\mathbf{s}_1, \dots, \mathbf{s}_N$ of random variable Z
- Here, each $\mathbf{s}_i$ is a 2-dimensional vector
- Data are a partial realisation of a stochastic process:

$$\{Z(\mathbf{s}) : \mathbf{s} \in D\} \tag{1}$$

  with $D \subset \mathbb{R}^2$
- Replications are not independent but are spatially correlated
- **Goal:** estimating $Z(\mathbf{s}_0)$ for location(s) $\mathbf{s}_0$
- Rewrite the model: $Z(\mathbf{s}) = \mu(\mathbf{s}) + \epsilon(\mathbf{s}) =$ drift + zero-mean residual
- **Estimation requires:**
  - Description of *drift* $\mathbb{E}[Z(\mathbf{s})] = \mu(\mathbf{s})$
  - Description of *spatial covariance* $\mathrm{Cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j))$

Gaussian Process

- Simplifying assumption 1 - *Stationarity*:

$$\mathbb{E}[Z(\mathbf{s})] = \mu \qquad \forall \mathbf{s} \in D$$

- Simplifying assumption 2 - *Second-order stationarity*:

$$\text{Cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j)) = C(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)) \qquad \forall \mathbf{s}_i, \mathbf{s}_j \in D \quad (2)$$

- Covariance between observations depends on distance and direction but not location
- C is called the *covariance function*
- Even stronger assumption - *Isotropy*:

$$C(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)) = C(Z(\mathbf{s}_k) - Z(\mathbf{s}_l)) \qquad \forall \mathbf{s}_i, \mathbf{s}_j, \mathbf{s}_k, \mathbf{s}_l \in D : \|\mathbf{s}_i - \mathbf{s}_j\| = \|\mathbf{s}_k - \mathbf{s}_l\|$$

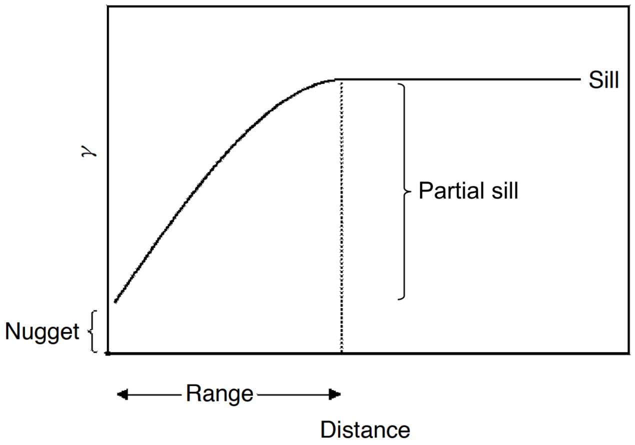
- Covariance between observations depends on distance but not on direction or location

Estimating Spatial Covariance - Semivariogram

- Alternative characterisation of spatial autocorrelation for isotropic processes (with some theoretical advantages)
- If:
$$\mathbb{V}[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)] = 2\gamma(\mathbf{s}_i - \mathbf{s}_j)$$
$Z(\cdot)$ is called *intrinsically stationary* and $2\gamma(\cdot)$ is called the *variogram*
- $\gamma(\cdot)$ is the *semivariogram* and only depends on the spatial lag $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$
- Properties:
  - $\gamma(-\mathbf{h}) = \gamma(\mathbf{h})$
  - $\gamma(\mathbf{0}) = 0$
  - $\gamma(\mathbf{h})/\|\mathbf{h}\|^2 \to 0$ as $\|\mathbf{h}\|^2 \to \infty$
  - $\gamma(\cdot)$ is conditionally negative definite for $\sum_{i=1}^{m} a_i = 0$:

$$\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j \gamma(\mathbf{s}_i - \mathbf{s}_j) \leq 0$$

Semivariogram



Relationship with Covariance function

- For a second-order stationary process $Z(\cdot)$:

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h})$$

- If $C(\mathbf{h}) \to 0$ as $\|\mathbf{h}\| \to \infty$, then $\gamma(\mathbf{h}) \to C(\mathbf{0})$, so $C(\mathbf{0})$ is the variance of the process
- Range of the semivariogram in direction $\mathbf{r}_0/\|\mathbf{r}_0\|$ is the smallest value $\|r_0\|$ so that $\gamma(\mathbf{r}_0) = C(\mathbf{0})$
- Spatial correlogram: $\rho(\mathbf{h}) = C(\mathbf{h})/C(\mathbf{0})$
- So why do we need all three functions? Because the class of intrinsically stationary processes is larger than the class of second-order stationary processes (Brownian motion: $\gamma(\cdot)$ exists, $C(\cdot)$ does not)

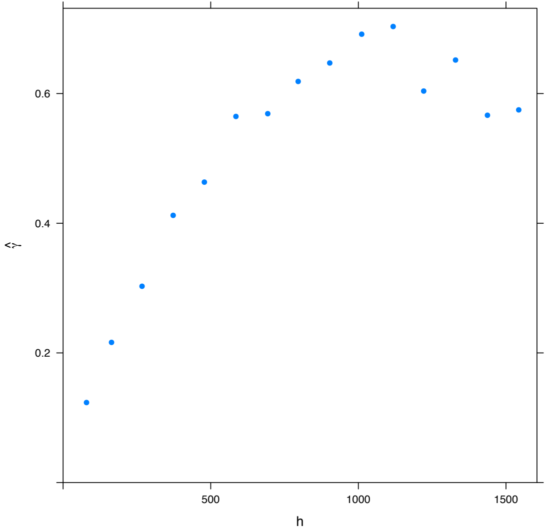
Estimating the Semivariogram

- Classic semivariogram estimator:
- Requires sufficient observations for each spatial lag
- With irregularly spaced data, pool data over tolerance regions
- Under intrinsic stationarity ($\mathbb{E}[Z(\mathbf{s})]$ constant):

$$
\begin{aligned}
2\gamma(\mathbf{h}) &= \mathbb{V}[Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{h})] \\
&= \mathbb{E}[(Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{h}))^2] - (\mathbb{E}[Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{h})])^2 \\
&= \mathbb{E}[(Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{h}))^2]
\end{aligned}
$$

- Estimate $\mathbb{E}[(Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{h}))^2]$ by averaging squared differences across pairs of observations

Estimating the Semivariogram



Parametric Isotropic Semivariogram Models

- Several parametric families can be fitted to estimated semivariogram
- Models typically assume isotropy but can be modified to fit well-defined anisotropy types:
  - Geometric anisotropy: range changes with direction, shape and sill remain constant
  - Zonal anisotropy: sill changes with direction, range remains constant
- Several fitting methods (NLS, ML)
- Methods for model comparison (single-parameter tests, chi-squared tests, AIC)
- Sum of models that are valid in $\mathbb{R}^d$ is again a valid model; model addition used for complex processes

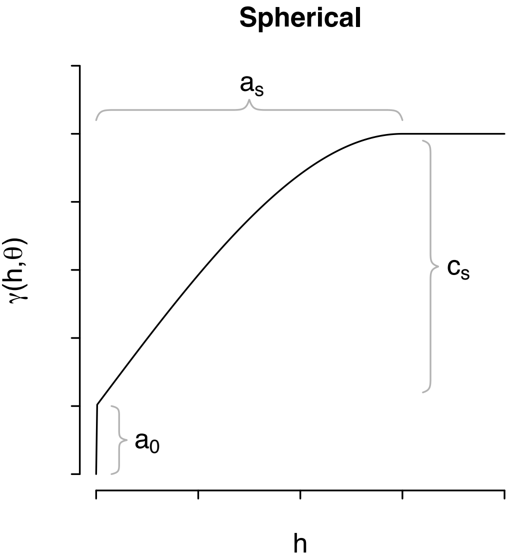
Spherical Semivariogram Model

- A semivariogram model is valid if it satisfies negative-definiteness
- *Spherical model*:

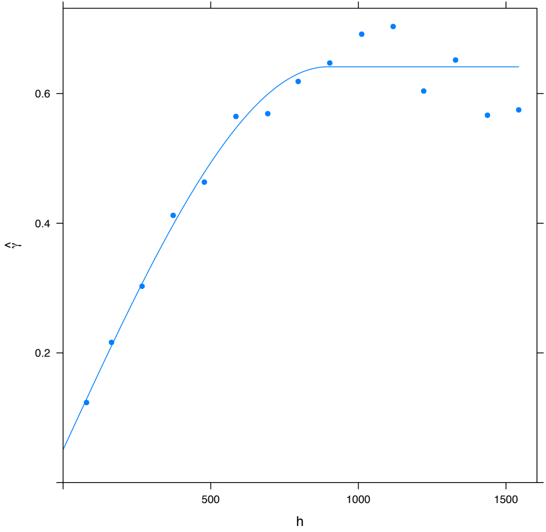
$$\gamma(h,\theta) = \begin{cases} 0, & h = 0 \\ c_0 + c_s[\frac{3}{2}\frac{h}{a_s} - \frac{1}{2}\left(\frac{h}{a_s}\right)^3], & 0 < h \leq a_s \\ c_0 + c_s, & h > a_s \end{cases}$$

  where $\theta = (c_0, c_s, a_s)$; $c_0, c_s \geq 0, a_s > 0$.
- c_0: nugget effect, c_s: partial sill, a_s: range
- Valid in 1, 2, and 3 dimensions

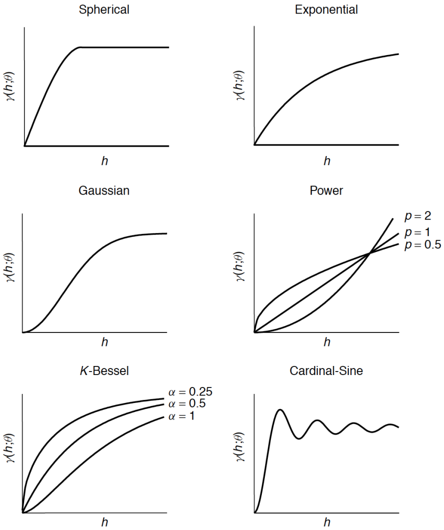
Spherical Semivariogram Model



Example Spherical Semivariogram Fit



Other Models



Estimating the Mean - Kriging

- Once we know how observations are related across space we can use these to:
  - Estimate $\mathbb{E}[Z(\mathbf{s}_0)]$ at some location $\mathbf{s}_0$
  - Obtain uncertainty bounds for our estimate
- Types of kriging:
  - *Simple*: mean known
  - *Ordinary*: mean constant but unknown
  - *Universal*: mean nonstationary and unknown
  - *Filterd*: smoothing and prediction for noisy data
  - *Lognormal*: optimal spatial estimation for lognormal data

Simple Kriging

- Let $S = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ be a set of locations at which we have observed realisations $z(\mathbf{s}_1), \ldots, z(\mathbf{s}_n)$ of the random variable $Z(\mathbf{s}) = \mu(\mathbf{s}) + \epsilon(\mathbf{s})$
- Assume $\mu(\mathbf{s})$ is known and $\gamma(\mathbf{s}_i - \mathbf{s}_j)$ is independent of the location
- We use the shorthand notation $Z_i = Z(\mathbf{s}_i)$ and $\mu_i = \mu(\mathbf{s}_i)$
- We want to estimate $Z_0 = Z(\mathbf{s}_0)$ using a linear estimator:

$$Z^* = \sum_{i=1}^{n} \lambda_i Z_i + \lambda_0$$

that minimises the *mean-squared prediction error* (MSPE):

$$\mathbb{E}[(Z^* - Z_0)^2] = \mathbb{V}[Z^* - Z_0] + (\mathbb{E}[Z^* - Z_0])^2$$

Eliminating Bias

- To minimise $\mathbb{V}[Z^* - Z_0] + (\mathbb{E}[Z^* - Z_0])^2$ we first eliminate the bias term, that is, we choose:

$$\lambda_0 = \mu_0 - \sum_{i=1}^{n} \lambda_i \mu_i$$

- This gives for the estimator:

$$Z^* = \mu_0 + \sum_{i=1}^{n} \lambda_i \left(Z_i - \mu_i\right)$$

- Without loss of generality $\mu_0 = 0$

Minimising Mean Squared Prediction Error

- We expand:

$$\mathbb{V}[Z^* - Z_0] = -\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\gamma(\mathbf{s}_i - \mathbf{s}_j) + 2\sum_{i=1}^{n}\lambda_i\gamma(\mathbf{s}_i - \mathbf{s}_0)$$

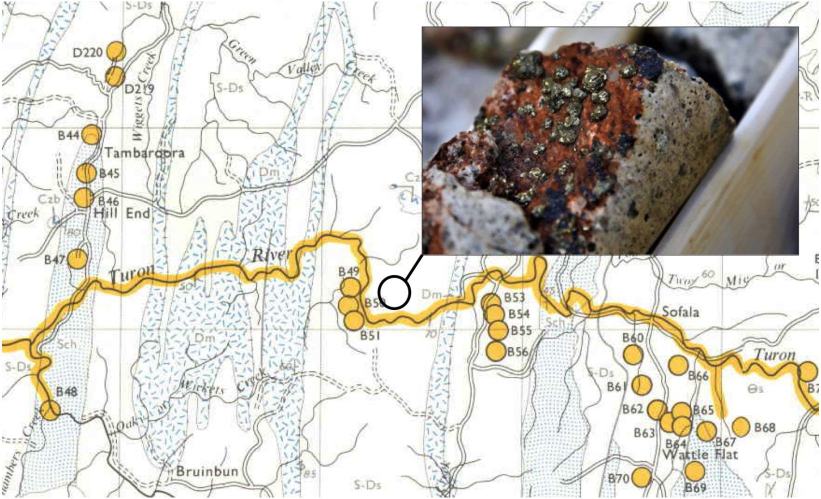
- Since we know/estimated $\gamma(\cdot)$, we can minimise the MSPE
- Taking derivatives wrt. λ_i and equating with 0 gives a system of equations, the *Simple Kriging System*:

$$\sum_{j=1}^{n}\lambda_j\gamma(\mathbf{s}_i - \mathbf{s}_j) = \gamma(\mathbf{s}_i - \mathbf{s}_0)$$

- Solution is the BLUP
- Measure of the error is given by simple kriging variance:

$$\sigma^2_{\mathrm{SK}} = \mathbb{E}[(Z^* - Z_0)^2] = C(\mathbf{0}) - \sum_{i=1}^{n}\lambda_i\gamma(\mathbf{s}_i - \mathbf{s}_0)$$

What if we don't know the mean?



Universal Kriging

- Basic idea: mean is a linear combination of 'nice' basis functions:

$$\mu(\mathbf{s}) = \sum_{\ell=0}^{L} a_\ell f^\ell(\mathbf{s}),$$

  mostly monomials in **s** of low degree
- We use the shorthand notation $f_i^\ell = f^\ell(\sim_i)$
- We again want to minimise the MSPE $\mathbb{E}[(Z^* - Z_0)^2] = \mathbb{V}[Z^* - Z_0] + (\mathbb{E}[Z^* - Z_0])^2$ using a linear predictor of the form:

$$Z^* = \sum_{i=1}^{n} \lambda_i Z_i,$$

Eliminating Bias

- Expanding the bias term we get:

$$\begin{aligned}\mathbb{E}[Z^* - Z_0] &= \sum_{i=1}^{n} \lambda_i \sum_{\ell=0}^{L} a_\ell f_i^\ell - \sum_{\ell=0}^{L} a_\ell f_0^\ell \\ &= \sum_{\ell=0}^{L} a_\ell \left(\sum_{i=1}^{n} \lambda_i f_i^\ell - f_0^\ell \right)\end{aligned}$$

- To minimise the MSPE we need to eliminate the bias, that is, we must have:

$$\sum_{i=1}^{n} \lambda_i f_i^\ell = f_0^\ell \qquad \forall \ell = 0, 1, \ldots, L,$$

called the *universal kriging conditions*

Minimising Mean Squared Prediction Error

- With these conditions in place, we expand the remaining term of the MSPE:

$$\mathbb{V}[Z^* - Z_0] = -\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j) + 2\sum_{i=1}^{n} \lambda_i \gamma(\mathbf{s}_i - \mathbf{s}_0)$$

- Our objective function now is:

$$Q = \mathbb{V}[Z^* - Z_0] + \sum_{\ell=0}^{L} a_\ell \left(\sum_{i=1}^{n} \lambda_i f_i^\ell - f_0^\ell \right)$$

where we need to estimate both, the a_ℓ and the λ_i

Minimising Mean Squared Prediction Error

- Taking partial derivatives wrt. a_ℓ and λ_i and equating to 0 gives:

$$\begin{cases} \sum_{j=1}^n \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j) + \sum_{\ell=0}^L a_\ell f_i^\ell = \gamma(\mathbf{s}_i - \mathbf{s}_0), & i = 1, \dots n \\ \sum_{i=1}^n \lambda_i f_i^\ell = f_0^\ell, & \ell = 0, \dots L, \end{cases}$$

  the *Universal Kriging System*

- Measure of the error is given by universal kriging variance:

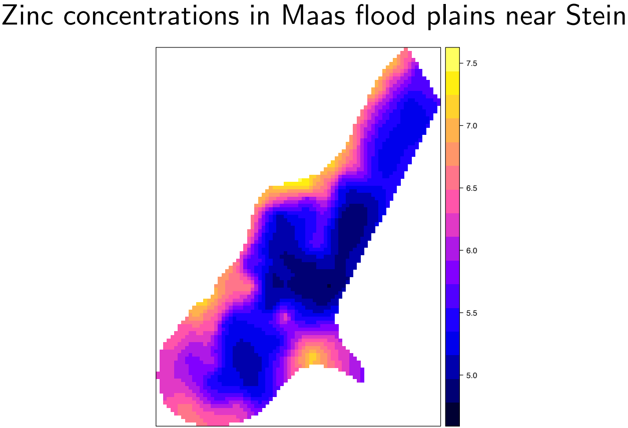
$$\sigma^2_{\mathrm{UK}} = \mathbb{E}[(Z^* - Z_0)^2] = \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_i - \mathbf{s}_0) + \sum_{\ell=0}^L a_\ell f_0^\ell$$

Misspecification of the Variogram

- Effect on kriging estimates is largely negligible as long as the behaviour near the origin is correct
- Effect on kriging variance can be substantial; sensitivity analysis is recommended

Predicting Zinc Concentrations

Predicting Zinc Concentrations



Thank You

More about kriging:

Waller, L.A. & Gotway, C. A. (2004). *Applied Spatial Statistics for Public Health Data*. Wiley.
Chiles, Jean-Paul & Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. Wiley.

Kriging tutorial:

https://rpubs.com/nabilabd/134781

Image sources:

https://www.resourcesandenergy.nsw.gov.au/__data/assets/image/0007/526246/sofala-hill-end-stuart-town-1-250-000-gold-deposits-map.jpg
http://www.equusmining.com/wp-content/uploads/2018/02/slide1-775x317.jpg
https://upload.wikimedia.org/wikipedia/commons/thumb/0/08/Sofala_Denison_Street_003.JPG/1200px-Sofala_Denison_Street_003.JPG
https://postcodebijadres.nl/gemeente/stein