Pengenalan Geostatistic

Definsi Geostatistic

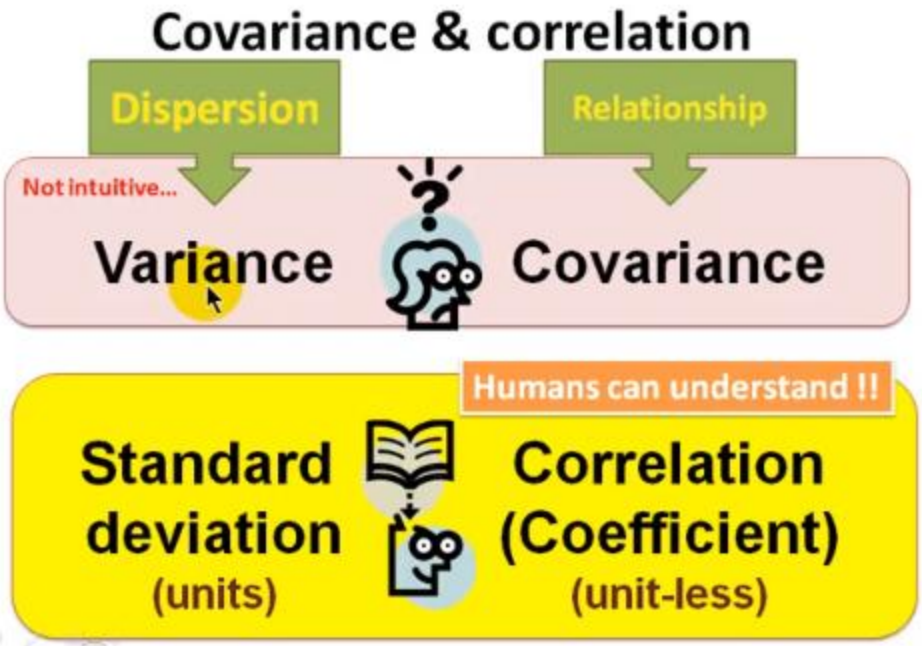
- In its broadest sense, geostatistics can be defined as the branch of statistical sciences that studies spatial/temporal phenomena and capitalizes on spatial relationships to model possible values of variable(s) at unobserved, unsampled locations (Caers, 2005)
- Geostatistics: study of phenomena that vary in space and/or time (Deutsch, 2002)
- Geostatistics can be regarded as a collection of numerical techniques that deal with the characterization of spatial attributes, employing primarily random models in a manner similar to the way in which time series analysis characterizes temporal data. (Olea, 1999)
- "Geostatistics offers a way of describing the spatial continuity of natural phenomena and provides adaptations of classical regression techniques to take advantage of this continuity." (Isaaks and Srivastava, 1989)

- Komponen dasar dari Geostatistic
  - Semivariogram analysis
  - Kriging
  - Stochastic simulation
- Metode geostatistical optimal dengan kondisi data:
  - Normally distributed
  - Stationary

  - Bagaimana menguji data terdistibusi normal?

Semivariogram

Covariance, Correlation and Semivariance



- Covariance
  - Ukuran seberapa besar 2 random variable berubah bersama.

$$\boxed{\sigma(X,Y) = E[(X-E(X))(Y-E(Y))]}$$

- Correlation
  - Nilai derajat dari 2 random variable secara linier berelasi.

$$\boxed{\rho(X,Y) = \frac{\sigma(X,Y)}{\sigma(X)\sigma(Y)}}$$

- Correlation coefficient (ρ)
  - 1 : perfect positive correlation
  - 0 : uncorrelated
  - -1 : perfect negative correlation

- Variable dalam spatial statistic
  - u: vector dari koordinat spatial (x,y atau easting, northing)
  - z(u) : fungsi dari koordinat spatial.
  - h : lag vector; pemisah antara 2 koordinat spatial.
  - z(u+h): lagged variable.

- Covariance, Correlation dan semivariance dalam spatial statistic
  - Covariance $$C(h) = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} z(u_\alpha) \cdot z(u_\alpha + h) - m_0 \cdot m_{+h}$$
  - Correlation $$\rho(h) = \frac{C(h)}{\sqrt{\sigma_0^2 \cdot \sigma_{+h}^2}}$$
  - Semivariance $$\gamma(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [z(u_\alpha + h) - z(u_\alpha)]^2$$
  - Means $$m_0 = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} z(u_\alpha) \qquad m_{+h} = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} z(u_\alpha + h)$$

- Variances

$$\boxed{\sigma_0^2 = \frac{1}{N(h)}\sum_{\alpha=1}^{N(h)}[z(u_\alpha)-m_0]^2} \quad \boxed{\sigma_{+h}^2 = \frac{1}{N(h)}\sum_{\alpha=1}^{N(h)}[z(u_\alpha+h)-m_{+h}]^2}$$

- Pada kondisi second-order stationarity (mean dan variance tetap)

$$\boxed{C(0)=Cov(z(u),z(u))=\sigma^2(z(u))}$$

$$\boxed{\rho(h)=\frac{C(h)}{C(0)}}$$

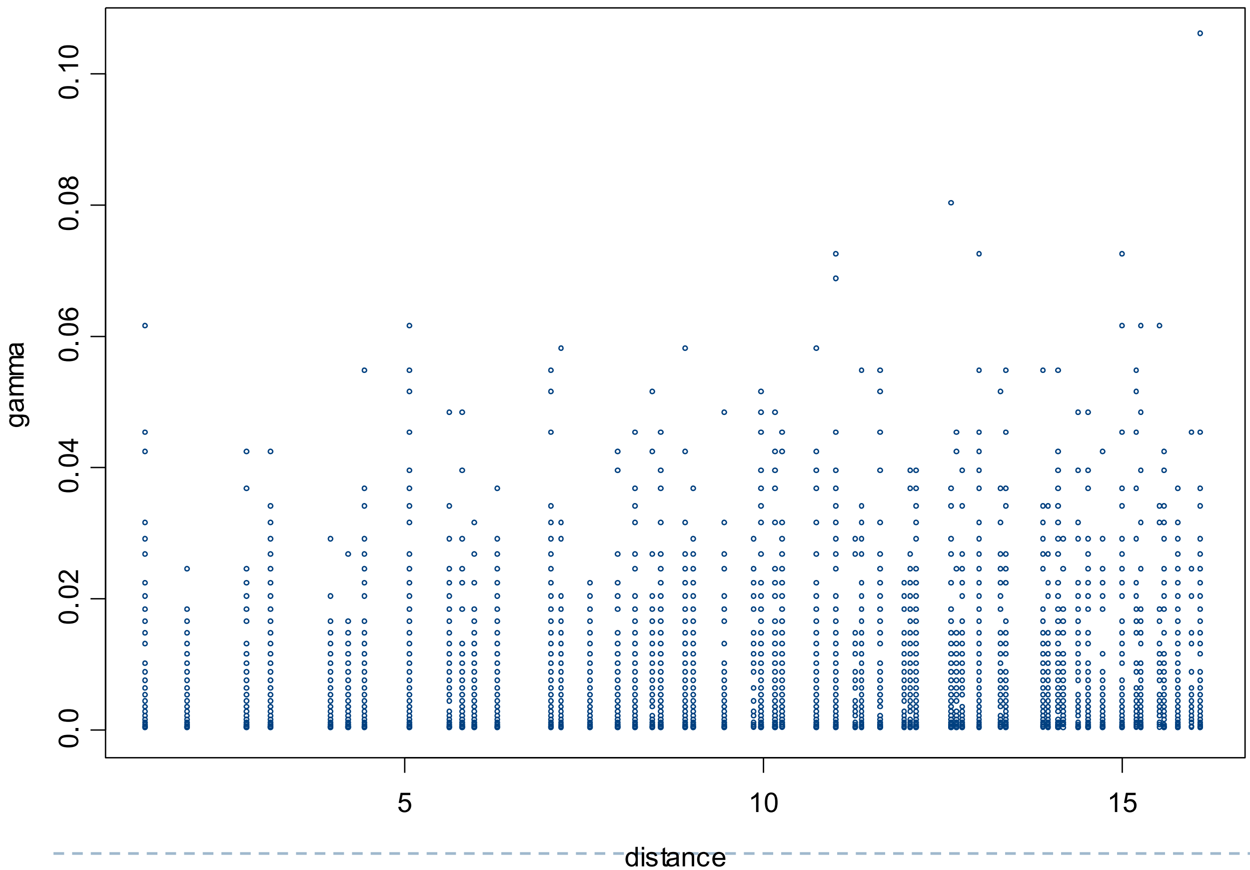
$$\boxed{\gamma(h)=C(0)-C(h)}$$

- H-scattergram adalah scatter plot dari z(u) dan z(u+h).
- Semivariance → seberapa menyebarnya data dari garis 45^{o} pada h-scattergram.
- Covariance dan correlation adalah ukuran similaritas sedangkan semivariance adalah ukuran disimilaritas.
- Nilai lag (h) biasanya menggunakna toleransi misalkan nilai lag nya 1000 dengan toleransi 500 (i.e. 500-1500)

Variogram Cloud

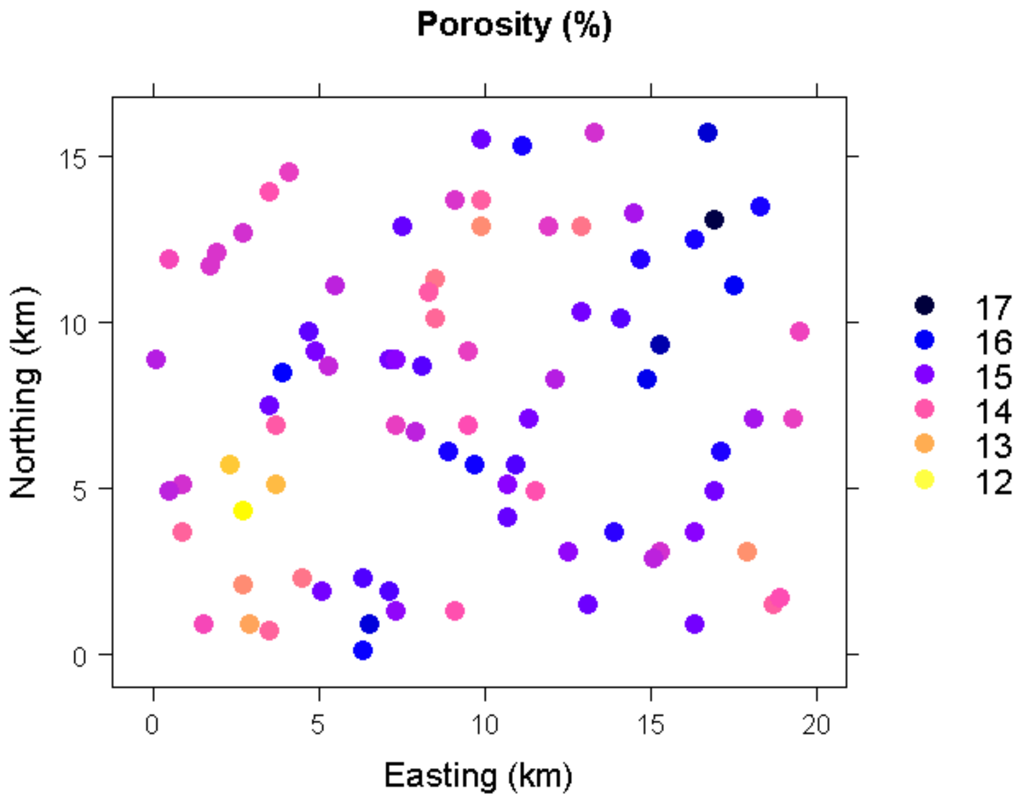
- Plot γ dan lag.

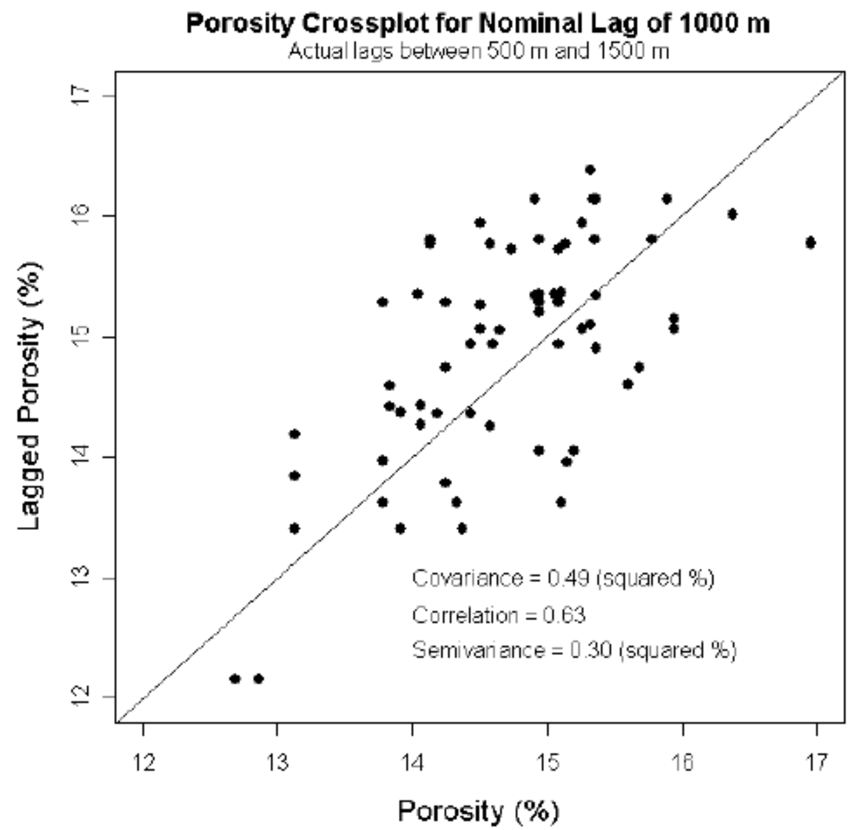
$$\gamma_{ij} = \frac{[Z(s_i) - Z(s_j)]^2}{2}$$

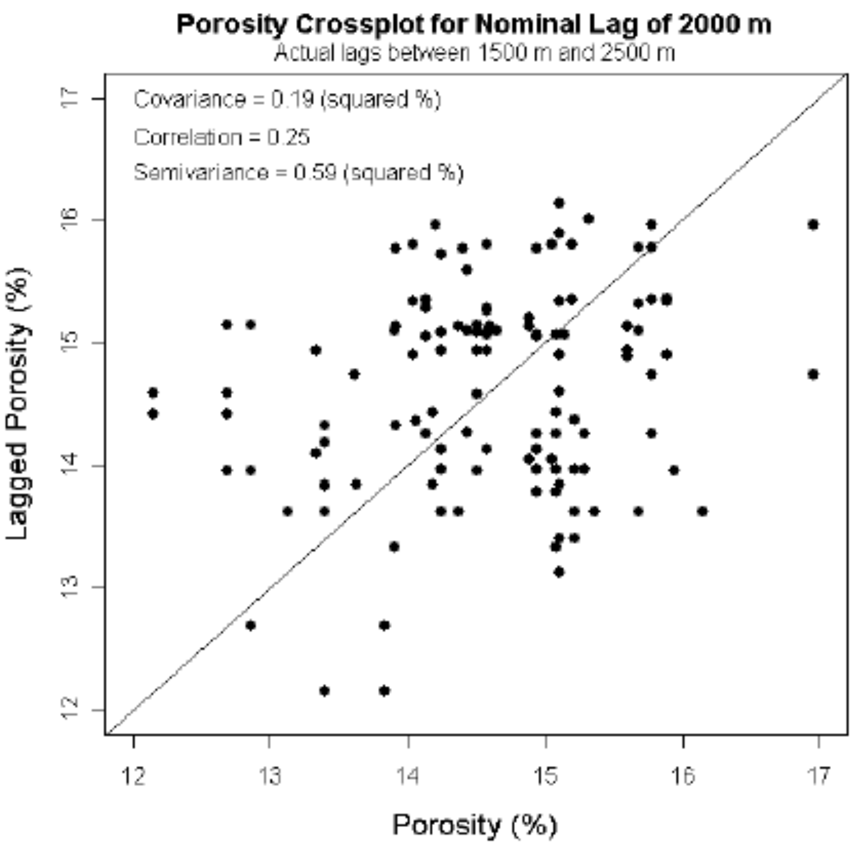


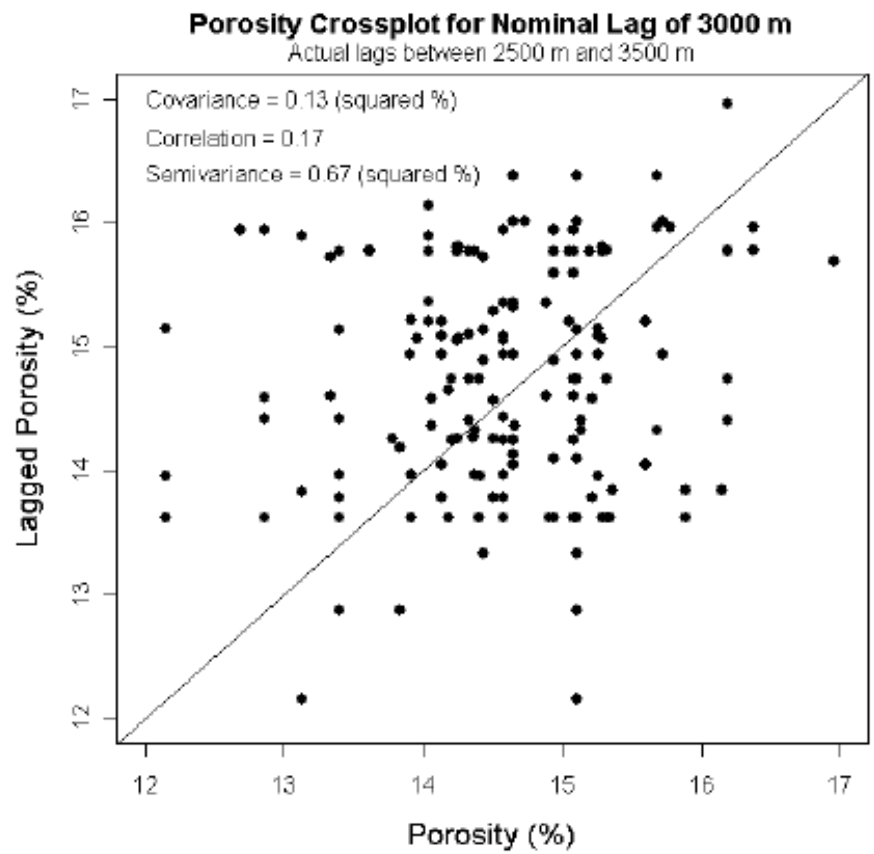
Contoh Covariance, Correlation dan Semivariance

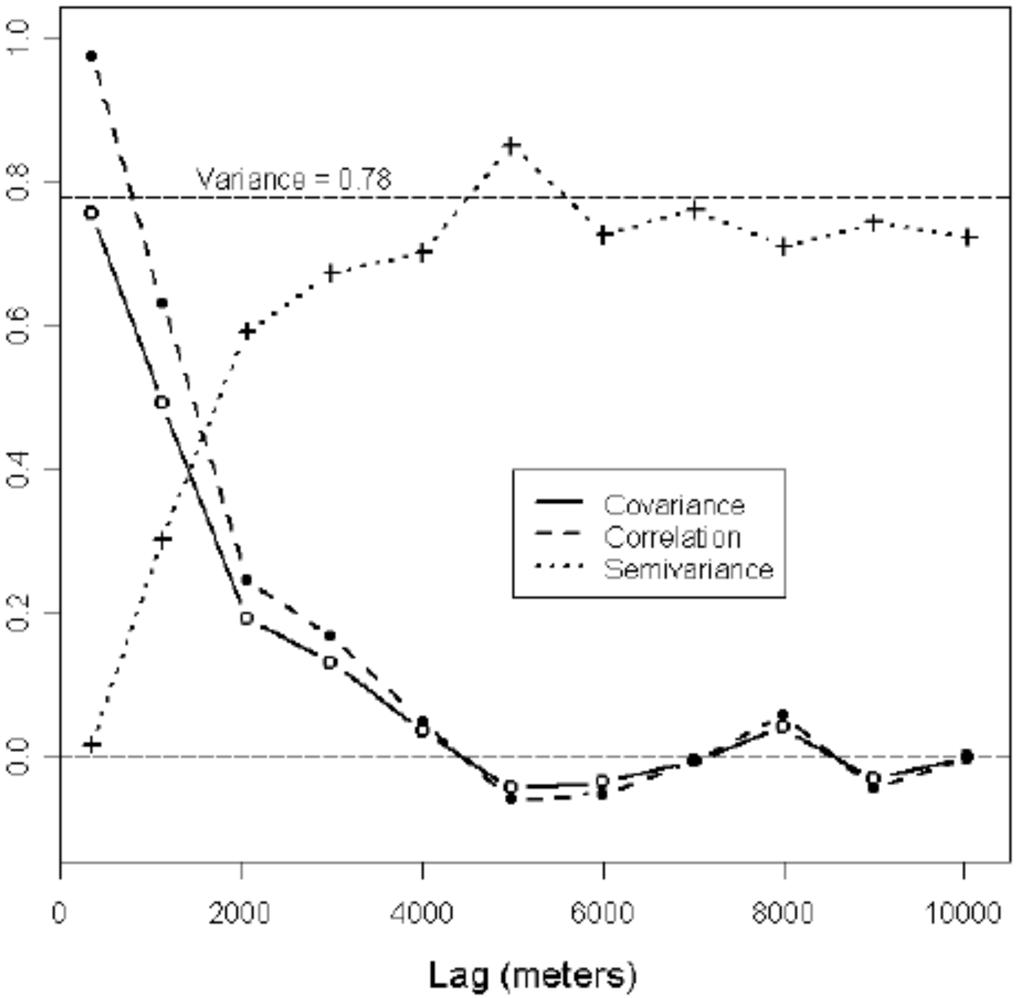
- Data : the Big Bean Oil Field, Zone A






Porosity Crossplot for Nominal Lag of 2000 m
Actual lags between 1500 m and 2500 m
Covariance = 0.19 (squared %)
Correlation = 0.25
Semivariance = 0.59 (squared %)
Lagged Porosity (%)
Porosity (%)
12
13
14
15
16
17

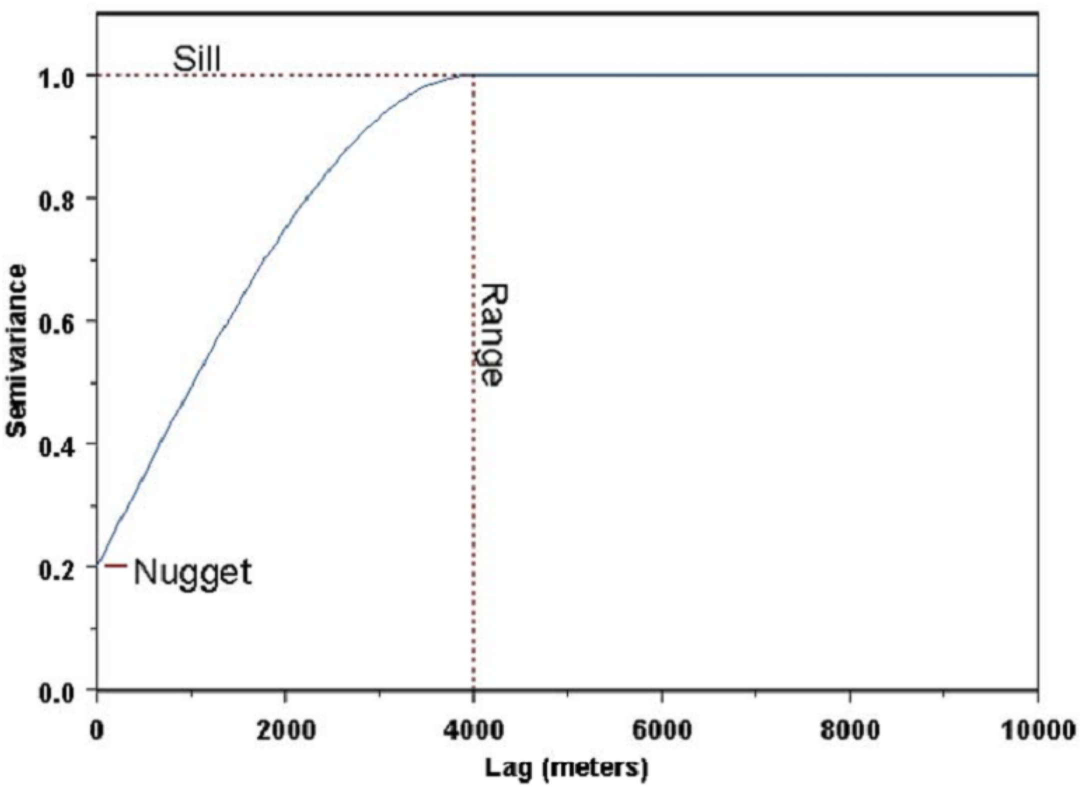

Porosity Crossplot for Nominal Lag of 3000 m
Actual lags between 2500 m and 3500 m
Covariance = 0.13 (squared %)
Correlation = 0.17
Semivariance = 0.67 (squared %)
Lagged Porosity (%)
Porosity (%)
12
13
14
15
16
17



- Correlation vs lag
  - Correlogram
- Semivariance vs lag
  - Semivariogram

Karakteristik Semi-Variogram

- Sill – nilai dimana grafik semivariogram turun (disebut ampliudo).
- Range – Nilai lag dimana grafik semivariogram mencapai sill.
- Nugget – Nilai semivariogram pada jarak yang mendekati 0.



- Jika empirical semivariogram bernilai naik dan stabil di luar dari nilai glpbal variance. Maka mengindikasikan ada suatu spatial trend yang significant

Pemodelan dari Semivariogram

- Perlu untuk kriging
  - Dapat mengakses nilai semivariogram sebagai fungsi bukan emperical.
- Pemodelannya yang umum (*licit semivariogram model*):

Nugget: $g(h) = \begin{cases} 0 & \text{if } h = 0 \\ c & \text{otherwise} \end{cases}$

Spherical: $g(h) = \begin{cases} c \cdot \left(1.5\left(\frac{h}{a}\right) - 0.5\left(\frac{h}{a}\right)^3\right) & \text{if } h \le a \\ c & \text{otherwise} \end{cases}$

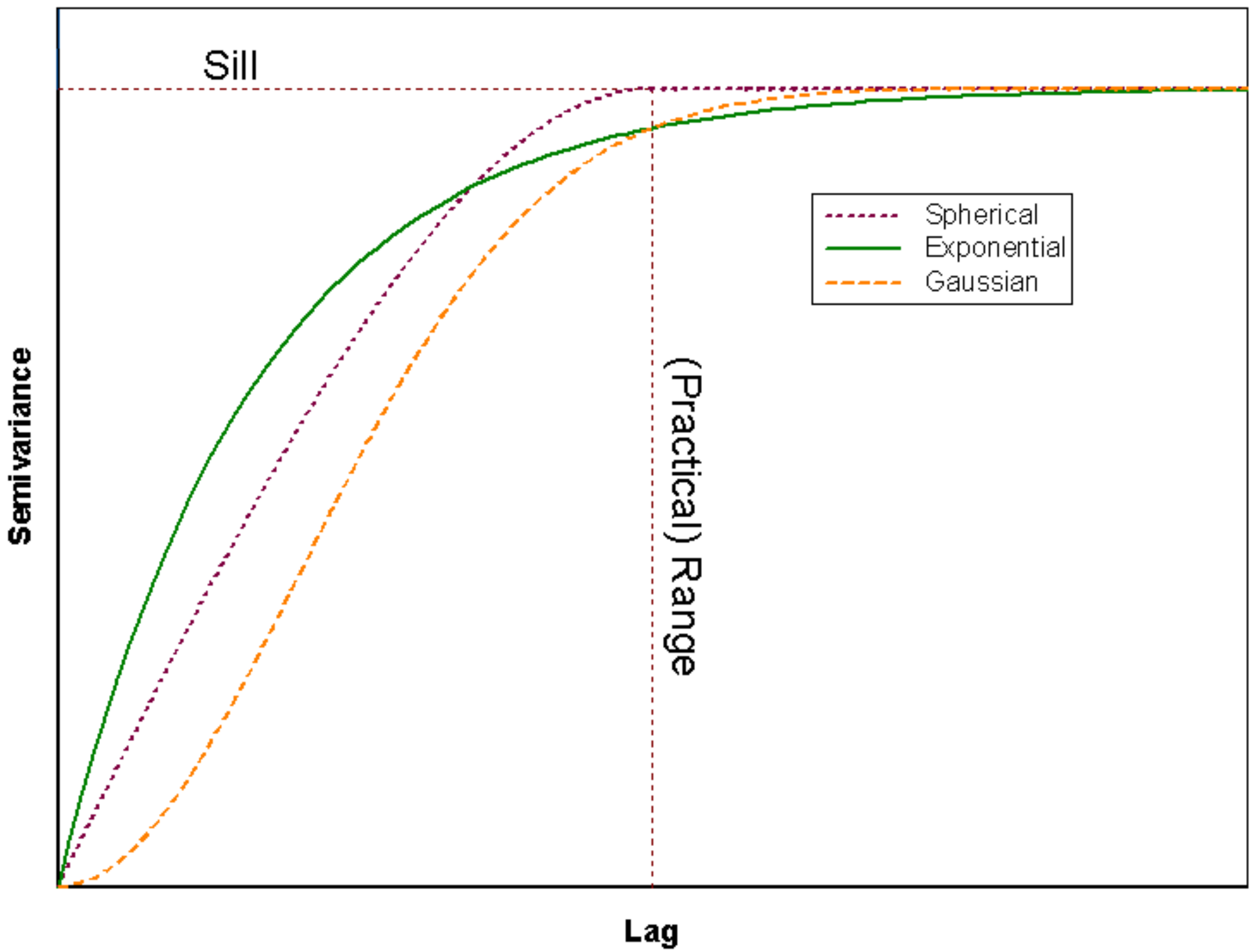
Exponential: $g(h) = c \cdot \left(1 - \exp\left(\frac{-3h}{a}\right)\right)$

Gaussian: $g(h) = c \cdot \left(1 - \exp\left(\frac{-3h^2}{a^2}\right)\right)$

Power: $g(h) = c \cdot h^{\omega} \quad with\ 0 < \omega < 2$

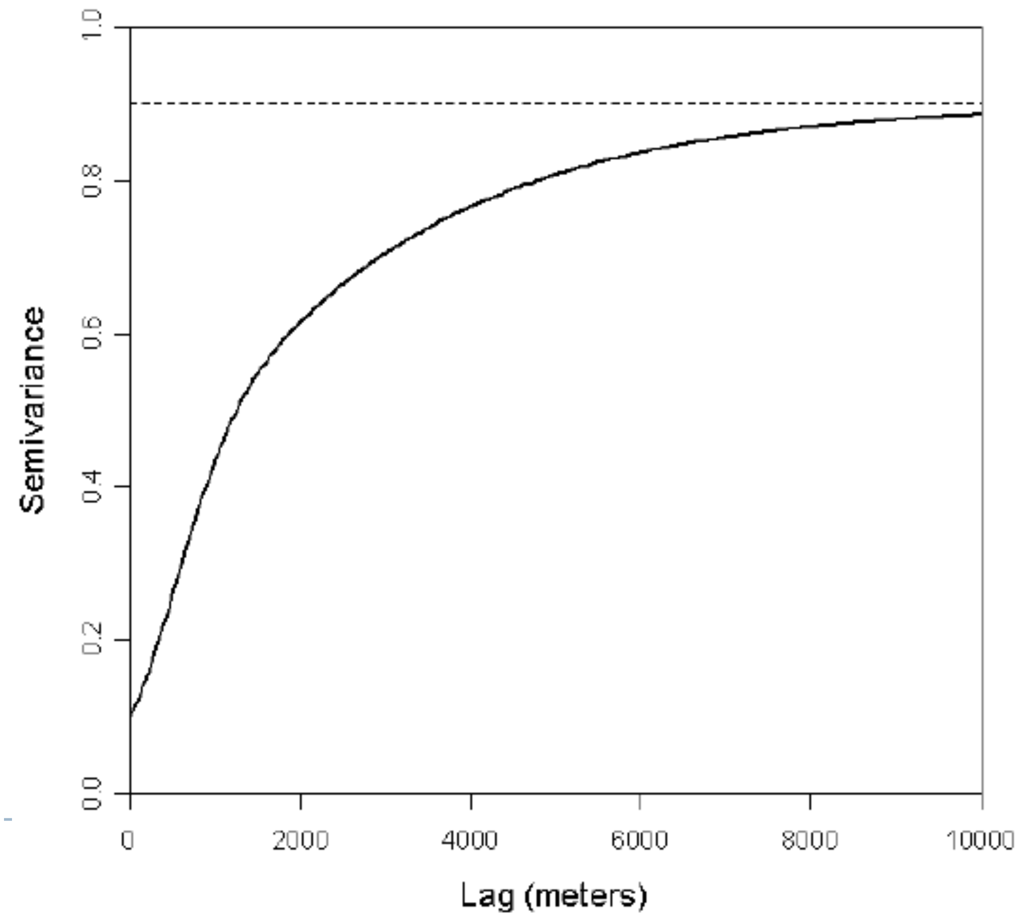
- h: lag
- a: practical range
- c: still

- Ilustrasi pemodelan semivariogram

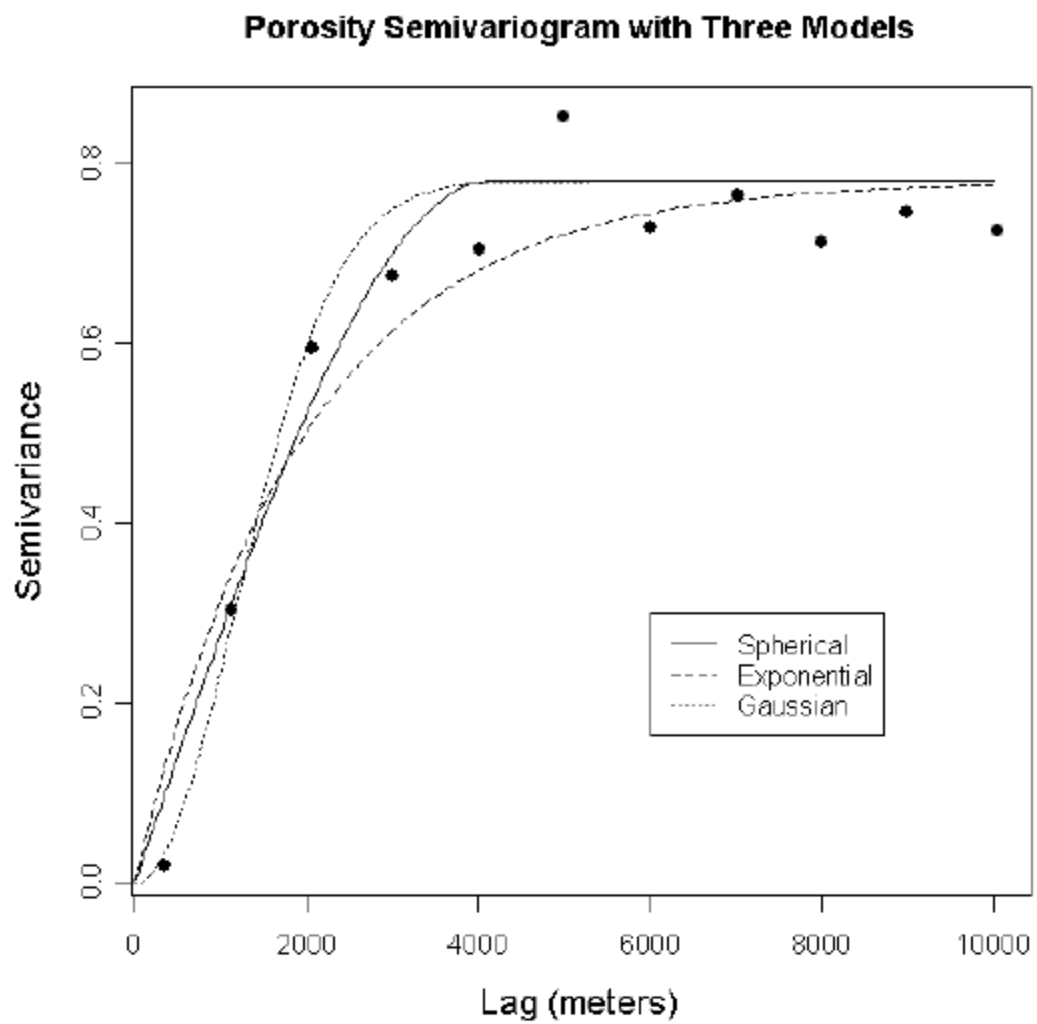


- Nugget mengindikasikan murni random variable dengan tidak ada spatial correlation
  - Discontinuity
- Gaussian → bervariasi dengan sangat halus.
  - Parabolic behaviour
- Spherical dan Exponential → bervariasi pada jarak yang pendek
  - Linear behaviour

- Gabungan linier dari licit semivariogram model adalah tetap sebahai licit model
- Nagget (c = 0.1) + Gaussian (a=1500,c=0.2) + Exponantial (a=8000, c=0.6)

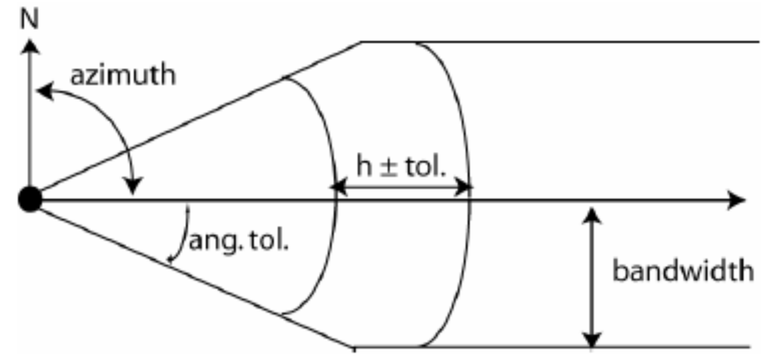


- Proses fitting model dari data observasi dengan regression

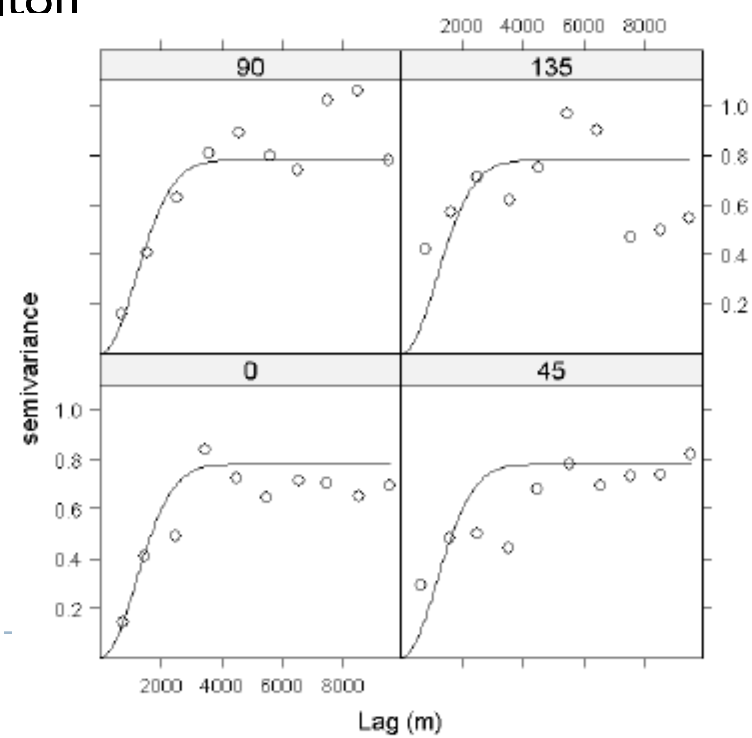


Omnidirectional

- Emperical semivariogram dihitung dengan memperhatikan arah.
- Anisotropic semivariogram
- Semivariance dihitung per directional band



- Contoh

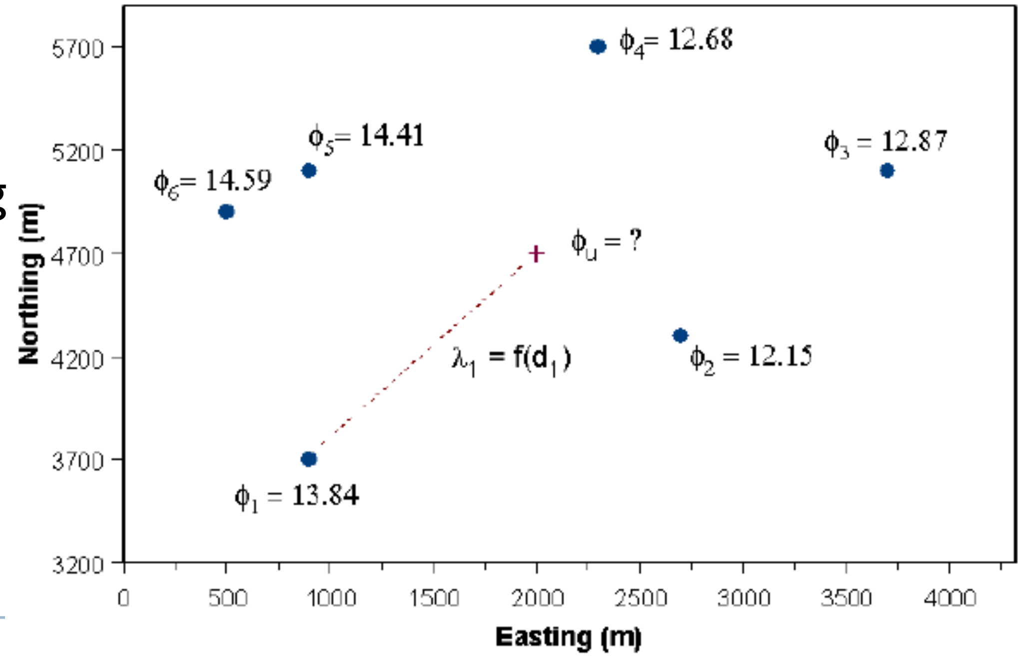


- Lebih bernoisi karena memotong jumlah pasangan data

Kriging

Interpolation

- Menghitung nilai suatu variable pada lokasi yang belum terhitung dengan menggunakan data observasi disekitarnya.
- Nilai dari suatu lokasi = jumlahan terbobot dari data disekitarnya.
- Jika data merata dan tredistibusi uniform maka hasil interpolasi akan baik tanpa mempedulikan metode yang dipakai
- Jika data tercluster dengan jarak antar cluster jauh maka akan mendapatkan hasil yang buruk tanpa mempedulikan metode yang dipakai



Definisi Kriging

- Optimal interpolation yang berdasarkan regresi pada data observasi disekitarnya dan diberi bobot sesuai dengan nilai spatial covariancenya.
- Kaunggulan
  - Membantu menanggulangi data tercluster dg memberikan nilai bobot lebih kecil dari pada isolated point.
  - Memberikan estimasi dari estimasi error (kriging variance)
  - Estimasi error ini menjadi basis untuk stochastic sumulation

Kriging estimator

- Basis linear estimator

$$\boxed{Z^*(u) - m(u) = \sum_{\alpha=1}^{n(u)} \lambda_\alpha [Z(u_\alpha) - m(u_\alpha)]}$$

- Dimana
- u : lokasi yang ingin diestimasi.
- u_α : data observasi disekirat u.
- $n(u)$: jumlah data observasi disekitar u
- $m(u)$: nilai ekspektasi dari $Z(u)$
- $m(u_\alpha)$: nilai ekspektasi dari $Z(u_\alpha)$
- $Z(u)$: random field dengan trend komponen $m(u)$ dan residual komponen $R(u)$.

$$\boxed{R(u) = Z(u) - m(u)}$$

- **Goal**
  - Menentukan nilai λ_α yang memiminalkan variance dari

$$\boxed{\sigma_E^2(u) = \sigma^2\{Z^*(u) - Z(u)\}}$$

dengan konstrain unbiasedness

$$\boxed{E\{Z^*(u) - Z(u)\} = 0}$$

- Residual Component
  - Stationary mean

$$\boxed{E\{R(u)\} = 0}$$

  - Stationary covariance

$$\boxed{Cov\{R(u), R(u+h)\} = E\{R(u) \cdot R(u+h)\} = C_R(h)}$$

Simple Kriging

- Trend nya konstan → *m(u)* = *m*
- Simple Kriging Estimator:

$$\boxed{Z^*_{SK}(u) = m + \sum_{\alpha=1}^{n(u)} \lambda_\alpha(u)[Z(u_\alpha) - m]}$$

- Error Estimation

$$\boxed{Z^*_{SK}(u) - Z(u) = [Z^*_{SK}(u) - m] - [Z(u) - m]}$$

$$\boxed{Z^*_{SK}(u) - Z(u) = R^*_{SK}(u) - R(u)}$$

$$\boxed{R^*_{SK}(u) = \sum_{\alpha=1}^{n(u)} \lambda_\alpha(u) R(u_\alpha)}$$

- Variance

$$\boxed{\sigma_E^2(u) = \sigma^2\{R_{SK}{}^*(u)\} + \sigma^2\{R_{SK}(u)\} - 2Cov\{R_{SK}{}^*(u), R_{SK}(u)\}}$$

$$\boxed{= \sum_{\alpha=1}^{n(u)}\sum_{\beta=1}^{n(u)} \lambda_\alpha^{SK}(u)\lambda_\beta^{SK}(u) C_R(u_\alpha - u_\beta) + C_R(0) - 2\sum_{\alpha=1}^{n(u)} \lambda_\alpha^{SK}(u) C_R(u_\alpha - u)}$$

- Untuk meminimalkan error maka $\sigma^2{}_E$ diturunkan terhadap setiap pembobotan kriging dan disamadengankan 0 sehingga

$$\boxed{\sum_{\beta=1}^{n(u)} \lambda_\beta^{SK}(u) C_R(u_\alpha - u_\beta) = C_R(u_\alpha - u\ )}$$

untuk

$$\boxed{\alpha = 1,\ldots,n(u)}$$

- Karena mean adalah konstant maka $\boxed{C(h) = C_R(h)}$ sehingga

$$\boxed{\sum_{\beta=1}^{n(u)} \lambda_{\beta}^{SK}(u) C(u_{\alpha} - u_{\beta}) = C(u_{\alpha} - u)} \qquad \boxed{\alpha = 1, \ldots, n(u)}$$

- Persamaan di atas dapat dibentuk dalam bentuk matrix;

$$\boxed{K\lambda_{SK} = k}$$

- Dimana *K* adalah covariance matrix antara data points dengan $\boxed{K(i,j) = C(u_i - u_j)}$, *k* adalah vector covariance antara data point dan estimation point di mana $\boxed{k(i) = C(u_i - u)}$ dan λ_{SK} adalah vector dari bobot kriging.

- Bobot kriging didapatkan dengan

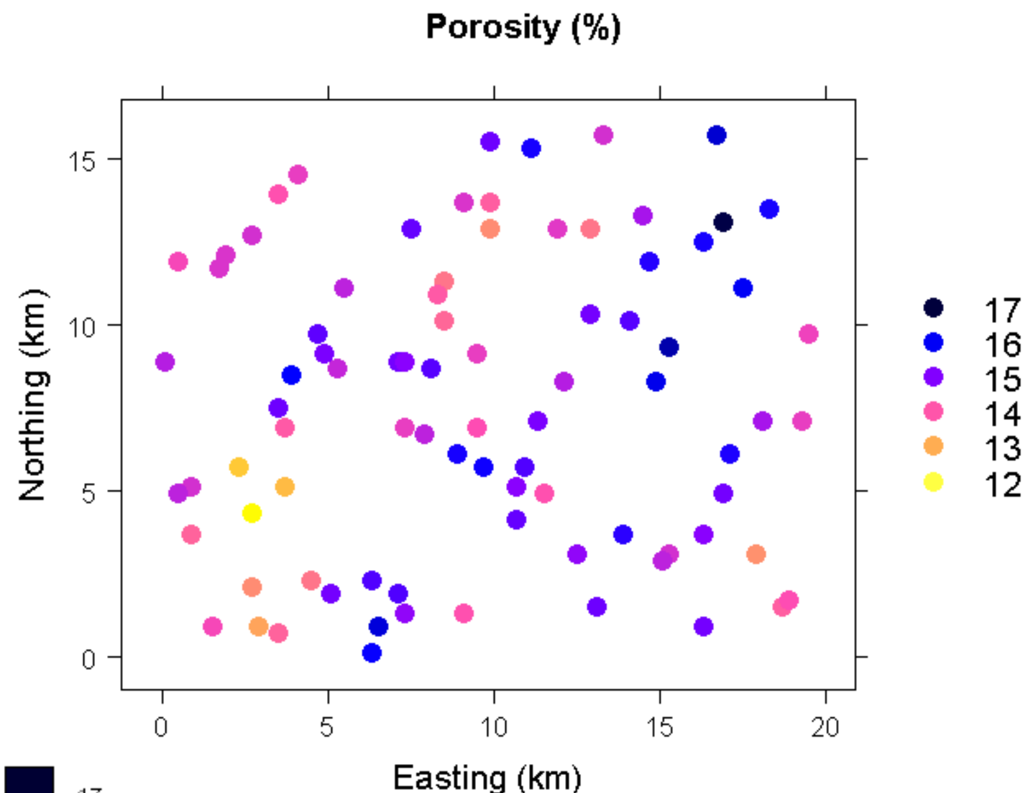
$$\boxed{\lambda_{SK} = K^{-1}k}$$

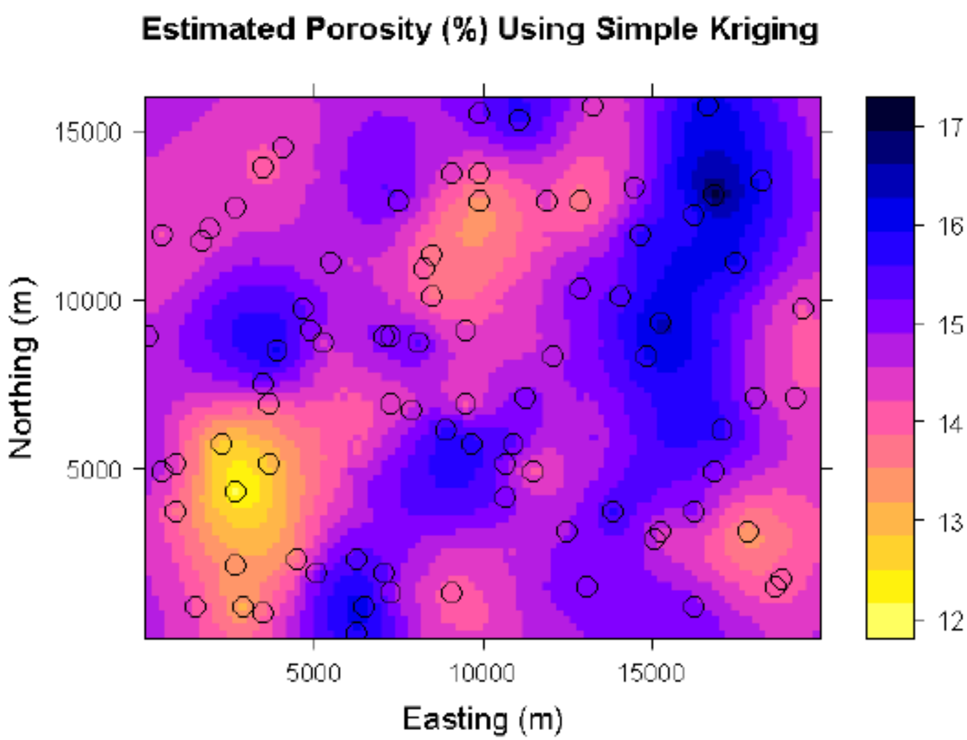
- Kringing variance dapat dihitung sesuai dengan persamaan

$$\boxed{\sigma_{SK}^{2}(u) = C(0) - \lambda_{SK}^{T}(u)k}$$

- Secara intiusi
  - Mencari set dari bobot untuk mengestimasi suatu nilai variable pada lokasi u dari nilai-nilai tetangganya.
  - Bobot akan bersesuaian dengan jarak dari u dan pengulangan data (data tercluster)

- Hasil simple kriging





FIN !