

Geostatistics Basics

SERGIO REY

GEOG 384

Spatial Data Analysis

Center for Open Geographical Science

Department of Geography

San Diego State University

Fall 2023

Outline

- 1 Geostatistical Perspective
 - Spatial Random Field
 - Moment Conditions
- 2 Variogram and Correlogram
 - Semi-Variogram
 - Covariogram

Spatial Random Field

Continuous Spatial Process

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

- s is spatial index, continuous in R^2 (R^3)
- Sample of spatial locations
 - ▶ $\{s_1, s_2, \dots, s_n\}$: sample locations
 - ▶ $\{Z(s_1), Z(s_2), \dots, Z(s_n)\}$: random variable at sample locations

Geostatistical Data

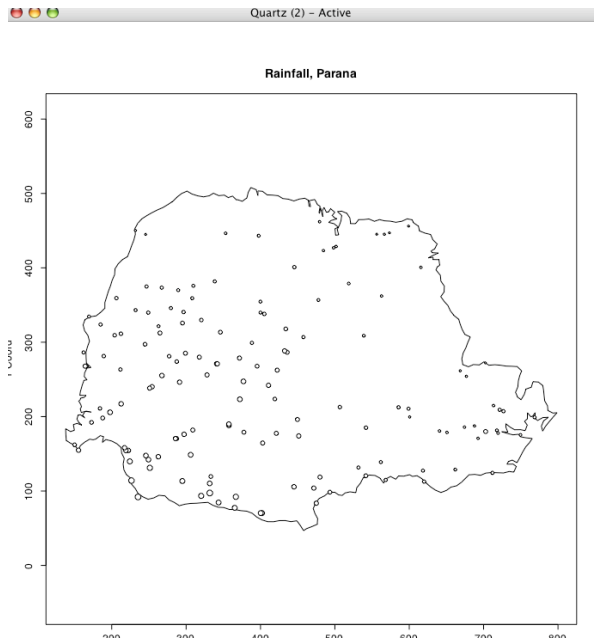
Spatial Domain: D

- A continuous and fixed set.
- Meaning $Z(s)$ can be observed everywhere within D .
- Between any two sample locations s_i and s_j you can theoretically place an infinite number of other samples.
- By fixed: the points in D are non-stochastic

Continuous Variation

- Because of the continuity of D
- Geostatistical data is referred to as “spatial data with continuous variation.”
- Continuity is associated with D .
- Attribute Z may, or may not, be continuous.

Geostatistical Data: Rainfall in Parana State Brazil



Geostatistical Data

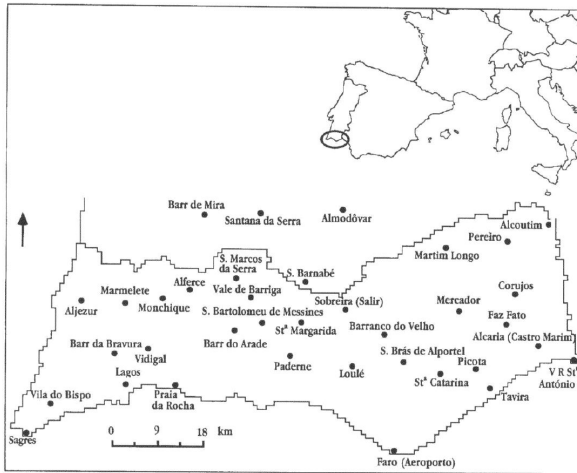
Continuous variation

- Potentially measurable anywhere in D
- Impossible to sample D exhaustively

Reconstruction of the surface from observed sites

- Tessellation based methods
- Interpolation
- Kriging

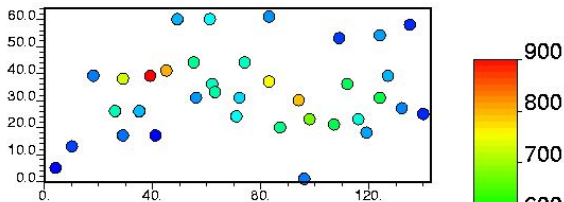
Surface Reconstruction: Example¹



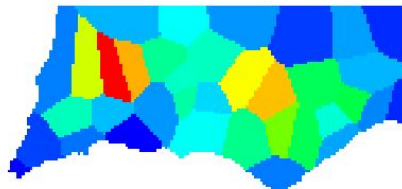
¹From Goovaerts, P. (1999) "Performance comparison of geostatistical algorithms for incorporating elevation into the mapping of precipitation". *Geocomputation '99*.

Surface Reconstruction: Tessellation Based Method

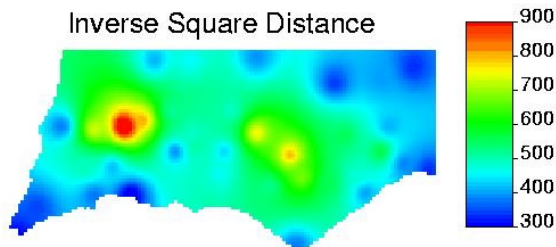
Annual rainfall data (mm)



Thiessen Polygons

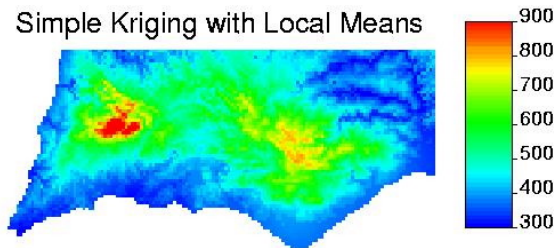


Surface Reconstruction: Spatial Interpolation

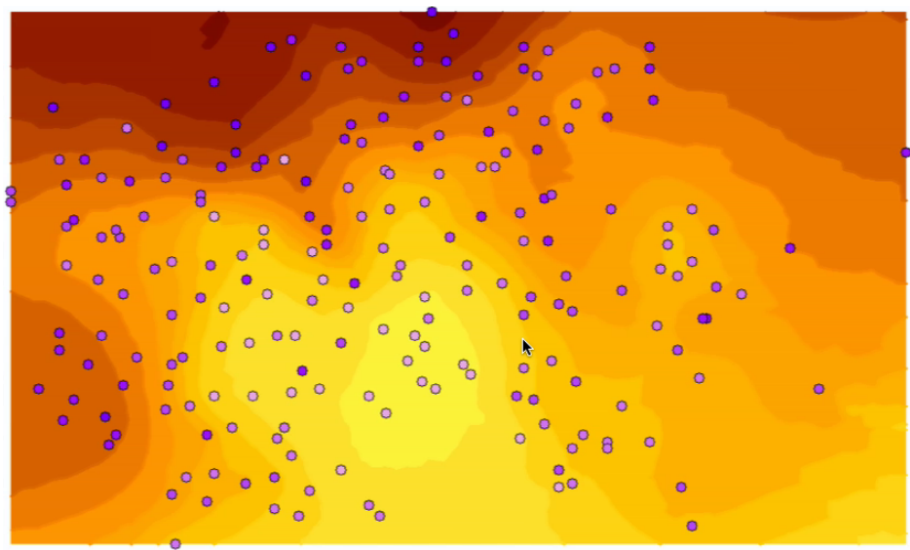


Surface Reconstruction: Kriging

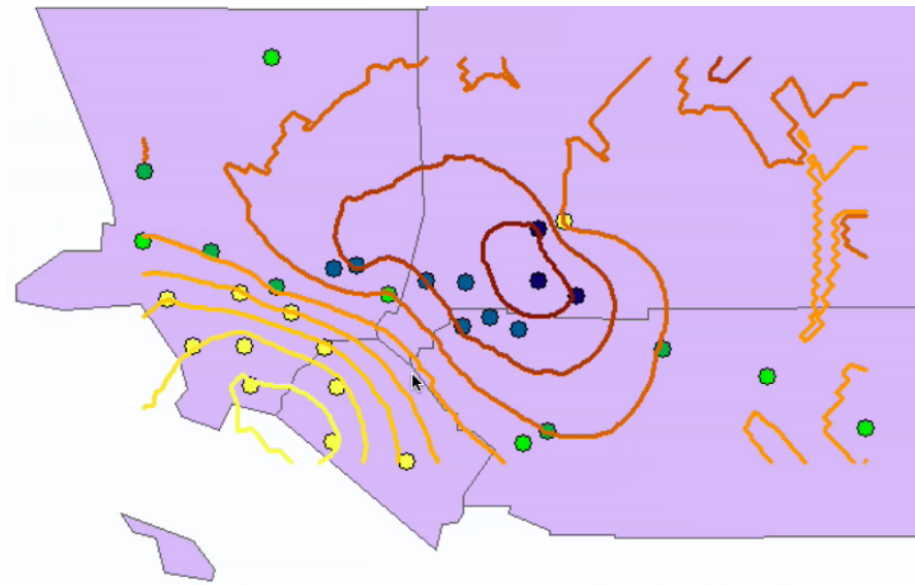
Simple Kriging with Local Means



Baltimore Sales Price, Baltimore MD



LA Basin Ozone Levels



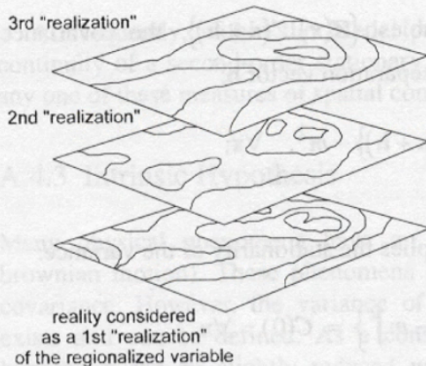
Conceptual Framework

Equilibrium

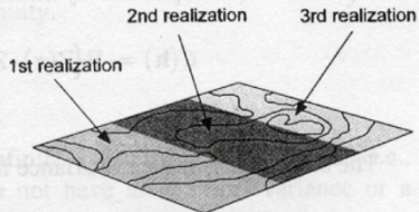
- Equilibrium = Stationarity
- Stochastic Process
 - ▶ not multiple realizations, but a single realization
 - ▶ the map is a single data point
- Notion of Stability
 - ▶ go from a single data point to acting as if there are multiple observations

Spatial Stationarity

Concept



Application (stationarity hypothesis)



Moment Conditions

- Constrain Variability
- Moments Must Exist
 - ▶ no infinite variance
- Moments Must Be Regular Over Space
 - ▶ restrictions on heterogeneity
 - ▶ restrictions on range of dependence

- Sample Realization is Representative
 - ▶ average obtained using a single realization is same as over all possible realizations
 - ▶ whether you observe one or many maps, the information should be the same

Strict Stationarity

- Pertains to the Complete Distribution
- Invariance Under Spatial Shift
 - ▶ joint density for two different spatial subsets is same
 - ▶ $\{z(s_l), \dots, z(s_k)\}$ and $\{z(s_{l+h}), \dots, z(s_{k+h})\}$
 - ▶ information about process is the same no matter where it is obtained
- Very Strict Requirement

Moment Stationarity

- Moments Invariant Under Spatial Shift
- Mean
 - ▶ no spatial trend
- Variance
 - ▶ no heteroskedasticity (no spatial regimes)
- Covariance
 - ▶ not a function of location
 - ▶ only spatial separation, angle

Intrinsic Hypothesis

No Spatial Trend

- if there is a trend, take it out
- residuals have no trend (mean = 0)

Constant Variance

- Variance of first difference only a function of displacement

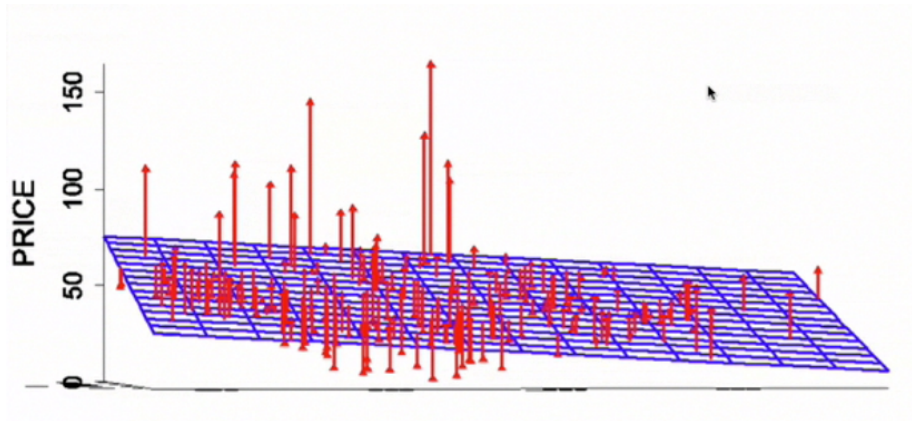
$$V\{Z_{s+h} - Z_s\} \quad (2)$$

a variability of the difference

Bubble Plot Baltimore House Sales Prices



Detrending Baltimore House Sales Prices



$$P = -166.02 - 0.148x + 0.634y$$

Semi-Variogram

Variogram Function

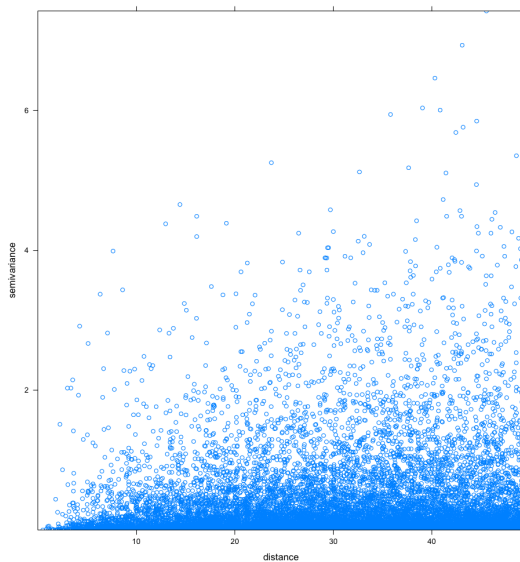
$$2\gamma(h) = V[Z_{s+h} - Z_s] \quad (3)$$

factor 2, so $\gamma(h)$ is half of the variogram, or semi-variogram

Constant Mean Assumption

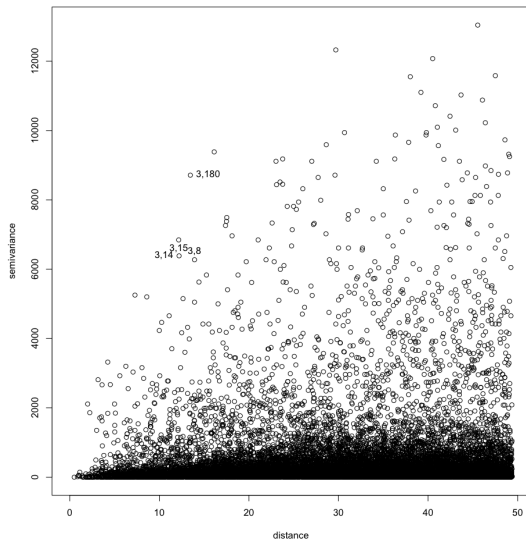
- $E[Z_{s+h} - Z_s] = E[Z_{s+h}] - E[Z_s] = 0$
- $V[Z_{s+h} - Z_s] = E[Z_{s+h} - E[Z_s]]^2 - 0$
- $\gamma(h) = (1/2)E[Z_{s+h} - E[Z_s]]^2$
- average of squared differences

Variogram Cloud

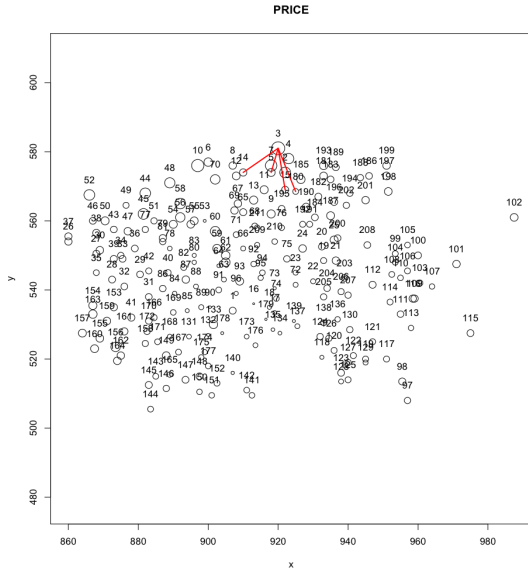


4

Outliers in Variogram Cloud



Outliers in Bubble Plot



Estimating a Variogram

Methods of Moments

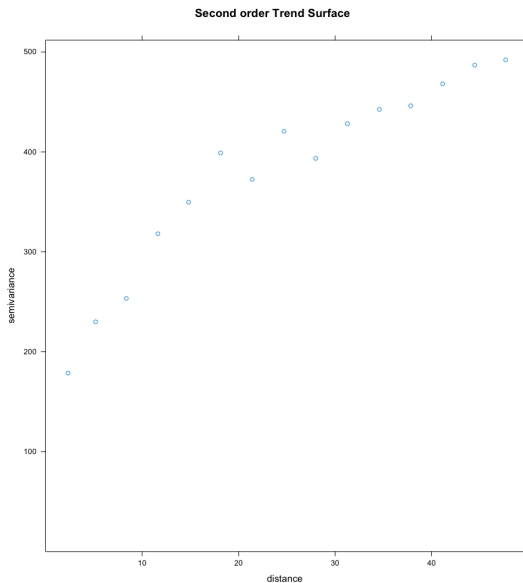
$$2\gamma(h) = (1/|N(h)|) \sum_h [Z_{s+h} - Z_s]^2 \quad (4)$$

- average of squared differences by distance bin h
- $N(h)$ number of pairs in distance bin h

Rules of Thumb

- at least 30 pairs in each bin
- $\max h < D/2$ (D is maximum distance)
- distance of reliability

Variogram on 2nd Order Trend Surface



Covariogram

Second Order Stationarity

- covariance regular over space
- stronger assumptions: regularity second order moments, cross-product

Covariance at h

$$C(h) = \text{Cov}[Z_{s+h}, Z_s] \quad (5)$$

Process Variance

- covariance at distance zero
- $C(0) = V[Z_s]$

Correlogram

Autocorrelation Function

$$\rho(h) = C(h)/C(0) \quad (6)$$

covariance standardized by process variance

Distance Decay

- Correlogram decreases with distance
- Tobler's law

Semi-Variogram and Covariogram

Two approaches to same concept

$$2\gamma(h) = E[Z_{s+h} - Z_s]^2 \quad (7)$$

$$= 2E[Z_s^2] - 2E[Z_{s+h}Z_s] \quad (8)$$

Semi-variance

- $\gamma(h) = C(0) - C(h)$
- semi-variance is variance of process less covariance at distance h