

Spatial Statistics

PM599, Fall 2013

Lecture 3: September 16th, 2013

- ▶ More on anisotropy
- ▶ Review of stationarity and the variogram
- ▶ Fitting semivariogram model to data (theoretical semivariogram to empirical semivariogram)
- ▶ Covariance functions
- ▶ Kriging

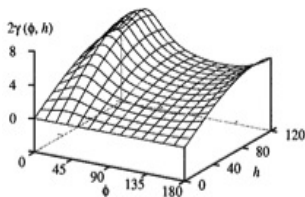
Anisotropy

- ▶ There are two types of anisotropy: Geometric and Zonal
- ▶ Geometric: directional semivariograms have the same shape and sill but different ranges. Sometimes called range anisotropy
- ▶ To make semivariograms isotropic, adapt our known isotropic semivariograms using elliptical geometry
- ▶ Rotate the coordinate axes so they are aligned with the major and minor axes of the ellipse
- ▶ Zonal isotropy: when sill changes with direction but the range remains constant. Sometimes called sill anisotropy
- ▶ See Eriksson and Siska (2000), posted on blackboard

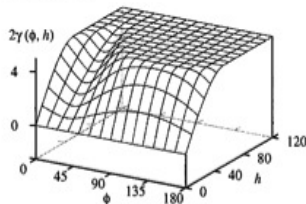
Geostatistical Data

Anisotropy

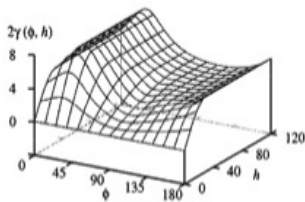
A: Sill Varies



B: Range Varies



C: Sill and Range Vary



D: Two Structures

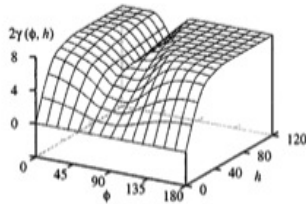


Figure: Eriksson, M and P.P Siska (2000)

We assume that our spatial process has a mean $E(Z(s)) = \mu(s)$ and the variance of $Z(s)$ exists.

- ▶ **Strong (also called strict):** the joint density is invariant under translation:
- ▶ this is often too restrictive for spatial applications

$$\begin{aligned} &P(Z(s_1) \leq z_1, Z(s_2) \leq z_2, \dots, Z(s_n) \leq z_n) \\ &= P(Z(s_1 + h) \leq z_1, Z(s_2 + h) \leq z_2, \dots, Z(s_n + h) \leq z_n) \end{aligned}$$

- ▶ **Weak (also called second order):** the moments (mean, variance) of the joint density are invariant. It has a constant mean $E(Z(s)) = \mu$ and covariance for all s is dependent only on distance $Cov(Z(s+h), Z(s)) = C(h)$ and we are interested in characterizing the covariance function. At $h=0$, we get $Cov(Z(s+0), Z(s)) = C(0) = Var(Z(s))$

- ▶ **Intrinsic:** when the difference (i.e. $Z(s+h) - Z(s)$) is second order stationary. $E(Z(s+h) - Z(s)) = 0$ Then we can write $Var[Z(s+h) - Z(s)] = 2\gamma((s+h) - s) = 2\gamma(h)$
- ▶ second order stationarity implies intrinsic stationarity but the reverse is not implied
- ▶ note the similarity between intrinsic stationarity and second order stationarity: intrinsic is defined in terms of the variogram and second order is defined in terms of the covariance function
- ▶ the variogram is a generalization of the covariance function and under second order stationarity the two functions are related

Geostatistical Data

Variogram and covariance

Relationship between semivariogram and covariance:

$$\begin{aligned}\gamma(h) &= \frac{1}{2}E[(Z(s+h) - Z(s))^2] \\&= \frac{1}{2}E[((Z(s+h) - \mu) - (Z(s) - \mu))^2] \\&= -E[(Z(s+h) - \mu)(Z(s) - \mu)] + \frac{1}{2}E[(Z(s+h) - \mu)^2] \\&\quad + \frac{1}{2}E[(Z(s) - \mu)^2] \\&= -C(h) + C(0) \\ \gamma(h) &= C(0) - C(h)\end{aligned}$$

- ▶ if $C(h)$ exists then we can get $\gamma(h)$, but can we get $C(h)$ from $\gamma(h)$?

Geostatistical Data

Variogram and covariance

- ▶ If $C(h) \rightarrow 0$ as $|h| \rightarrow \infty$
- ▶ Covariance goes to 0 as distance goes to infinity
- ▶ if we take the limit on both sides of $\gamma(h) = C(0) - C(h)$, get
$$\lim_{h \rightarrow \infty} \gamma(h) = C(0)$$
- ▶ but the limit may not exist, for example in the linear semivariogram case

Geostatistical Data

Variogram and covariance

- ▶ Example: Linear semivariogram
- ▶ $\gamma(h) = \tau^2 + \sigma^2 h$ if $h > 0$; 0 otherwise
- ▶ as $h \rightarrow \infty$ then $\gamma(h) \rightarrow \infty$
- ▶ thus is not a second-order stationary process and $C(h)$ does not exist

- ▶ We want to use the covariance function/semivariogram function to make spatial predictions
- ▶ Assumptions of prediction via kriging require that we have stationarity
- ▶ We may look at the semivariogram but then need a covariance function, thus we must have intrinsic stationarity which is also a second-order stationary process

Geostatistical Data

Fitting a semivariogram model

- ▶ Eyeballing the semivariogram is useful for exploratory purposes and to find the approximate shape of the spatial process, but we would rather find a valid theoretical semivariogram function that reflects the empirical semivariogram.
- ▶ We choose from our set of valid theoretical semivariograms and see how well the function fits to our data.
- ▶ We can't just pick any curve that looks to fit our data because the semivariogram model must be negative definite to ensure that results aren't off (i.e. the covariances of multiple points are inconsistent with each other, or could have negative variance for weighted averages)

Geostatistical Data

Fitting a semivariogram model

- ▶ Necessary condition of negative definiteness:
- ▶ $\sum_i \sum_j \lambda_i \lambda_j \gamma(s_i - s_j) \leq 0$
- ▶ where $\sum_i \lambda_i = 0$
- ▶ This is the analog of positive definiteness for covariance functions to ensure that all variances are non-negative
- ▶ $\lambda_1, \dots, \lambda_n$ are weights

Geostatistical Data

Fitting a semivariogram model

Non-linear Least Squares: find the parameters $\theta = (\tau^2, \sigma^2, \phi)$ that minimize the squared vertical distance between the empirical and theoretical semivariograms.

Use the binned semi-variogram

$$(\hat{\gamma}(h) - \gamma(h; \theta))^T (\hat{\gamma}(h) - \gamma(h; \theta))$$

where $\hat{\gamma}(h)$ is the empirical semivariogram and $\gamma(h; \theta)$ is the theoretical semivariogram with parameters θ

Geostatistical Data

Fitting a semivariogram model

- ▶ The relationship between $\hat{\gamma}(h)$ and h is usually nonlinear (semivariogram model is not a linear function)
- ▶ Use nonlinear least squares, solved numerically
- ▶ Minimize SSE $\sum_j^K [\hat{\gamma}(h_j) - \gamma(h_j)]^2$
- ▶ K bins from our empirical semivariogram

Geostatistical Data

Fitting a semivariogram model

Problem: OLS assumes:

- ▶ independence, but each observation enters into multiple bins (iid)
- ▶ equal variance of bins, but there are different numbers of pairs in each bin (homoskedasticity)

Solution: the usual adjustment to OLS when observations are correlated and heteroskedastic is generalized least squares (GLS)

GLS and Weighted Least Squares (WLS):

$$(\hat{\gamma}(h) - \gamma(h; \theta))^T V(h; \theta)^{-1} (\hat{\gamma}(h) - \gamma(h; \theta))$$

- ▶ Correlation among bins is accounted for with $V(h; \theta)^{-1}$
- ▶ Difficult to calculate this since θ are unknown, computationally intensive
- ▶ Use approximation and weighted least squares which accounts for unequal variance of bins (Cressie 1985)
- ▶ WLS still does not account for correlation, but is better than OLS as it gives more weight to bins having more data

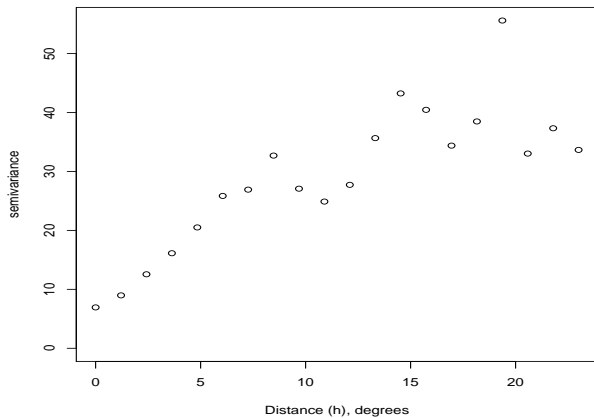
Geostatistical Data

Fitting a semivariogram model

- ▶ $V(h; \theta)^{-1} = I$ gives the OLS equation
- ▶ Taking $V(h; \theta)^{-1} = \text{diag Var}[\hat{\gamma}(h_1), \dots, \text{Var}[\hat{\gamma}(h_K)]]$ gives a weighted least squares estimator
- ▶ $\text{Var}[\hat{\gamma}(h_j)] \approx 2[\hat{\gamma}(h_j)]^2 / N(h_j)$
- ▶ now we minimize WSSE $\frac{1}{2} \sum_j^K \frac{N(h_j)}{\hat{\gamma}(h_j)} [\hat{\gamma}(h_j) - \gamma(h_j)]^2$

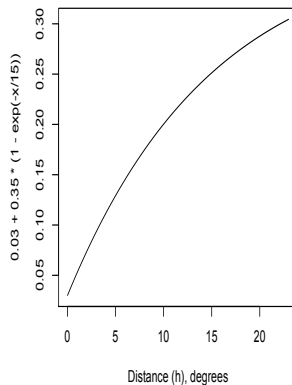
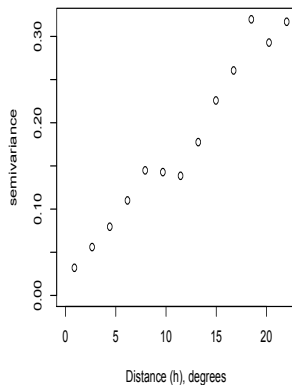
Geostatistical Data

Fitting a semivariogram model



Geostatistical Data

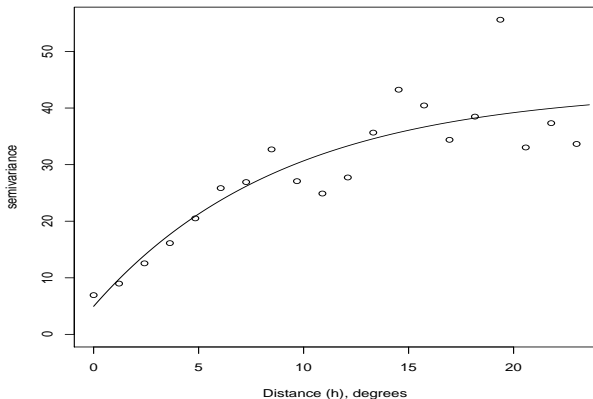
Fitting a semivariogram model



Geostatistical Data

Fitting a semivariogram model

Exponential OLS

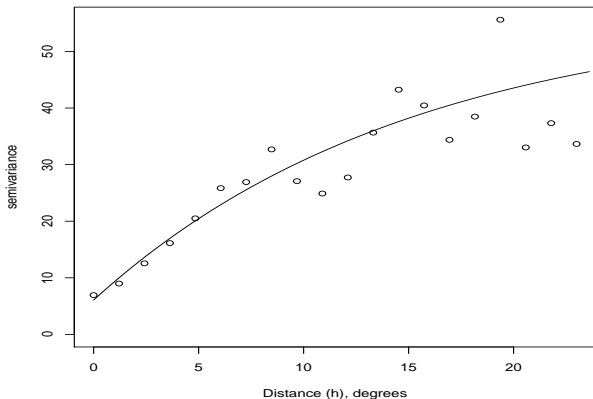


Estimated parameters: $\sigma^2 = 38.447$, $\rho = 9.049$, $\tau^2 = 4.954$; SSE: 578.356

Geostatistical Data

Fitting a semivariogram model

Exponential WLS

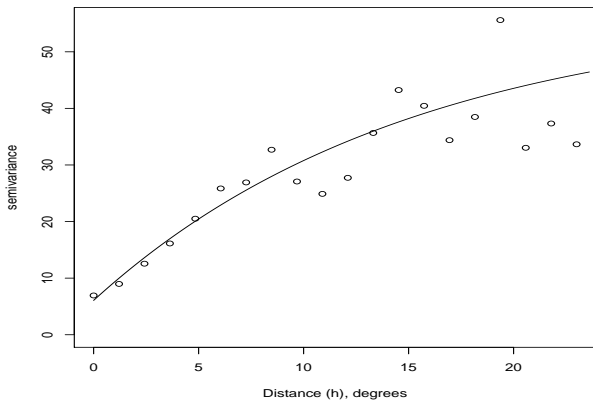


Estimated parameters: $\sigma^2 = 38.447$, $\rho = 9.049$, $\tau^2 = 4.954$; SSE: 948.569

Geostatistical Data

Fitting a semivariogram model

Gaussian WLS



Estimated parameters: $\sigma^2 = 26.733$, $\rho = 6.07$, $\tau^2 = 7.987$; SSE: 997.563

Geostatistical Data

Fitting a semivariogram model

- ▶ There is extensive literature on fitting semivariograms, but approach is somewhat arbitrary and unsatisfying statistically
- ▶ The objective that is minimized (deviation of empirical semivariogram values from semivariogram model) is based on pseudo-data
- ▶ Fitting is basically just curve fitting and is sensitive to the binning and maximum distance chosen
- ▶ However, calculations based on semivariograms are fast, even with many observations, which is not true for maximum likelihood
- ▶ Semivariogram modeling is not based on a particular probability model for the data, so it may be more resistant to violations of assumptions
- ▶ Compare SSE different models fit by one method (i.e. exponential vs spherical WLS). Don't compare SSE from WLS and OLS

Fitting covariance models by maximum likelihood

- ▶ The more standard statistical approach is to fit a covariance model by maximum likelihood (ML)
- ▶ ML is the most common approach in statistics to fit models by estimating parameter values
- ▶ Recall, in linear regression with normal (uncorrelated) errors, least squares is the same as maximum likelihood estimation.
- ▶ ML requires specification of a probability model (the likelihood) for the data
- ▶ Likelihoods involve unknown parameters that must be estimated from data
- ▶ Our spatial data must follow a multivariate Gaussian distribution and have second-order stationarity (Covariance exists)

Fitting covariance models by maximum likelihood

- ▶ Gaussian process: MVN, mean = μ , Cov = $\Sigma(\theta)$
- ▶ Can think of this as spatial regression:
$$E(Z(s)) = \mu + \epsilon(s) = X\beta + \epsilon(s)$$
- ▶ Where $\epsilon(s) \sim N(0, \Sigma(\theta))$
- ▶ The log-likelihood function has the form:
$$L(\beta, \theta; Z) = -\frac{1}{2} \log |\Sigma(\theta)| - \frac{1}{2} (Z - X\beta)^T \Sigma(\theta)^{-1} (Z - X\beta)$$
- ▶ Restricted maximum likelihood (REML) is an alternative and is based on maximizing the likelihood when the data are differences

Fitting covariance models by maximum likelihood

- ▶ The goal is to maximize the probability of the data relative to different parameter values
- ▶ The parameter values are treated as unknown and the data as fixed, and the parameter values that give the highest likelihood are chosen
- ▶ ML/REML is done in this case by numerical methods (there is no closed form solution) and can be intensive for large datasets (more than a few hundred observations)

Choosing among models fit by ML

- ▶ The traditional way is to use Akaike's Information Criterion, which in its general form minimizes:
- ▶ $AIC = 2 \log(\text{maximized likelihood}) + 2(\text{number of parameters})$
- ▶ AIC can be used for non-nested models. It compares the likelihoods of different models and penalizes models with more parameters:
- ▶ Models with smaller AIC are favoured

- ▶ Kriging is the spatial prediction of our process at unobserved locations
- ▶ Based on the fitted covariance function and the spatial regression model $E(Z(s)) = \mu + \epsilon(s) = X\beta + \epsilon(s)$
- ▶ Objective: To estimate the value of $Z(s)$ at one or more unsampled locations in our region D based on our observed samples $z(s_1), z(s_2), \dots, z(s_n)$

- ▶ The basic kriging recipe:
 1. Choose a parametric model for the semivariogram or covariance function
 2. Estimate the semivariogram/covariance parameters.
 3. Make predictions and uncertainty estimates given the parameter estimates.
- ▶ The kriging predictions are weighted averages of the observations. The covariance/semivariogram indicates the strength of spatial association and determines the weighting.
- ▶ The issue is how heavily to weight the observations based on distance from the location.

- ▶ basic kriging predictor at new location s_0 is
$$\hat{Z}(s_0) = \sum \lambda_i Z(s_i)$$
- ▶ Goal: to minimize squared error loss $E[(\hat{Z}(s_0) - Z(s_0))^2]$
- ▶ The best prediction of this is the conditional mean: $E(Z(s_0)|Z)$ which is the expected value of what you don't know given what you do know.
- ▶ This calculation assumes you know the covariance function (or have estimated it).

Geostatistical Data

Kriging

