

Spatial Statistics

PM569 Lecture 4: Geostatistical Data 3

Meredith Franklin

Division of Biostatistics, University of Southern California

September 20th, 2019

Geostatistical Data

- ▶ Kriging (Simple, Ordinary, Universal)
- ▶ Spatial regression
- ▶ Other interpolation and smoothing methods

Geostatistical Data: Kriging

Typically, the final stage of a classical geostatistical analysis is to predict values of $Z(s)$ at unobserved locations in $D \in \mathbb{R}^2$. Recall that unobserved locations could be specific x,y locations, or all locations (i.e. an infinitely small grid) in the domain. The method used for this purpose is called Kriging.

Kriging:

- ▶ is a technique for spatial prediction.
- ▶ uses estimated spatial process to predict values at unobserved locations.
- ▶ is based on the fitted covariance function and the spatial regression model
 $Z(s) = \mu + \epsilon(s)$ where μ can be $X\beta$.
- ▶ $\epsilon(s) \sim N(0, \Sigma(\theta))$

Objective of kriging: To estimate the value of $Z(s)$ at one or more unobserved locations in our region D based on our observed samples $z(s_1), z(s_2), \dots, z(s_n)$

Geostatistical Data: Kriging

- ▶ The basic kriging recipe:
 1. Choose a parametric model for the semivariance (covariance) function.
 2. Estimate the semivariogram (covariance) parameters (WLS, ML, or REML).
 3. Gather a set of unobserved locations (points or grid).
 4. 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.

Geostatistical Data: Kriging Theory

Best linear unbiased predictor (BLUP)

For an unobserved spatial location, s_0 we want to estimate $\hat{Z}(s_0)$. Kriging gives us the BLUP at any new location s_0 . It is BLUP because:

- ▶ the prediction variance $E[\hat{Z}(s_0) - Z(s_0)]^2$ is minimized ("best").
- ▶ it is a linear prediction based on a weighted average of the observations ("linear"). The weighted average is $\hat{Z}(s_0) = \sum_{i=1}^N \omega_i Z(s_i)$
- ▶ the expected value of the prediction $E[\hat{Z}(s_0)]$ is equal to the expected value at s_0 ("unbiased"). The unbiased condition $E[\hat{Z}(s_0)] = E[Z(s_0)]$ implies that $\sum_i \omega_i = 1$. This constraint also ensures non-negative and finite variance.

Geostatistical Data: Kriging Theory

- ▶ Goal is to minimize squared error $E[(\hat{Z}(s_0) - Z(s_0))^2]$ subject to the unbiasedness constraint $\sum_{i=1}^n \omega_i = 1$.
- ▶ The problem boils down to finding the set of coefficients $\omega_1, \dots, \omega_n$ that serve as the weights of our observations.
- ▶ The Lagrange multiplier method is used for finding the minimum subject to this constraint.
- ▶ This calculation assumes we know the moments, i.e. the mean $\mu(s)$ and the semivariance/covariance (or we have estimated it) $\gamma(h)$ or $C(h)$.

Geostatistical Data: Kriging Theory

Mathematically,

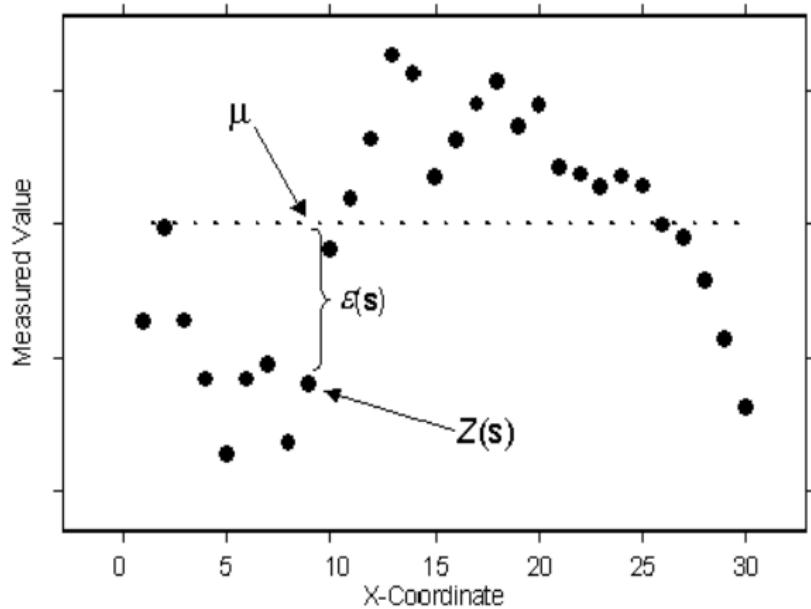
- ▶ $\hat{Z}(s_0) = \sum_{i=1}^N \omega_i Z(s_i)$
- ▶ $E[(\hat{Z}(s_0))] = \mu = E[Z(s_0)]$
- ▶ with constraint $\sum_{i=1}^N \omega_i = 1$
- ▶ Minimize $E[(\hat{Z}(s_0) - Z(s_0))^2]$
- ▶ See handout posted on blackboard.

Geostatistical Data: Kriging Types

There are different kriging types for different assumptions and analytical goals.
The types are:

- ▶ Simple Kriging: assumes a constant known mean $\mu = c$. This type is not often used because for unbiasedness constraint to be applicable in kriging equations, we must estimate the expected value.
- ▶ Ordinary Kriging: assumes a constant unknown mean (mean needs to be estimated) $Z(s) = \mu + \epsilon(s)$.
- ▶ Universal Kriging: assumes a trend in x and y , and may include other spatially varying covariates $Z(s) = \mu(s) + \epsilon(s)$ where $\mu(s) = \sum_{k=1}^p \beta_k x_k(s_i)$.

Geostatistical Data: Ordinary Kriging



Geostatistical Data: Ordinary Kriging

From the Lagrange multiplier approach, the kriging equations are $\sum_j \omega_j C_{ij} + \lambda = C_{i0}$ where C_{i0} and C_{ij} are evaluated based on our data and chosen covariance function.

$$\begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_N \\ \lambda \end{pmatrix} = \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1N} & 1 \\ C_{21} & C_{22} & \dots & C_{2N} & 1 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ C_{N1} & C_{N2} & \dots & C_{NN} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix}^{-1} \times \begin{pmatrix} C_{10} \\ C_{20} \\ \vdots \\ C_{N0} \\ 1 \end{pmatrix}$$

We have a linear kriging system of $N+1$ equations with $N+1$ unknowns. Solving for the weights, $\omega = \mathbf{C}^{-1}\mathbf{D}$.

Note that \mathbf{C} only needs to be estimated once, but \mathbf{D} is found for every prediction location s_0 .

Geostatistical Data: Ordinary Kriging

The variance of the prediction $\hat{Z}(s_0)$ is important as a measure of uncertainty in our predictions. It is obtained by the MSE:

$$\begin{aligned}MSE &= \sum_{i=1}^N \sum_{j=1}^N \omega_i \omega_j Cov[Z(s_i), Z(s_j)] + Var[Z(s_0)] \\&\quad - 2Cov\left[\sum_{i=1}^N \omega_i Z(s_i), Z(s_0)\right]\end{aligned}$$

which results in

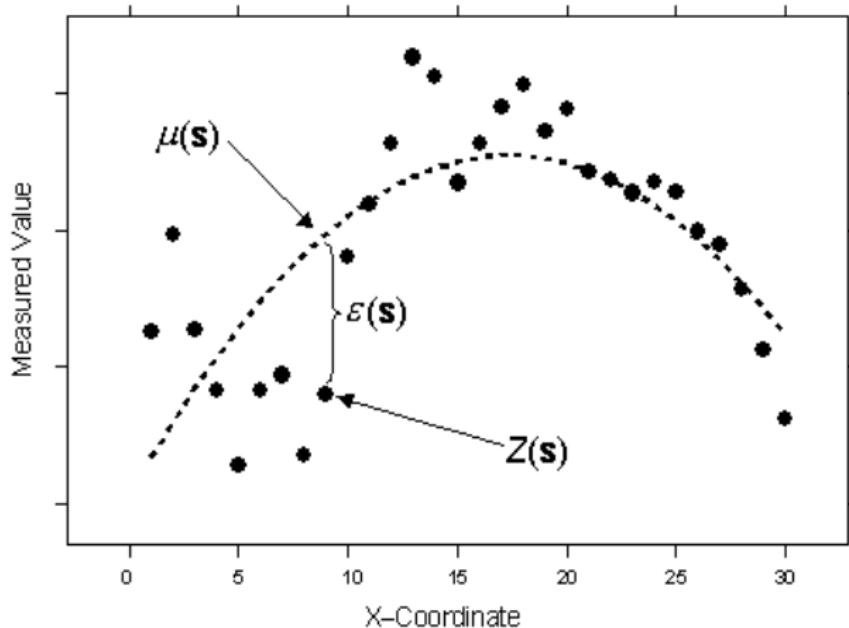
$$\sigma_{OK}^2 = \sigma^2 - \sum_{i=1}^N \omega_i (Cov[Z(s_i), Z(s_0)] - \lambda)$$

and in matrix form is

$$\sigma_{OK}^2 = \sigma^2 - \mathbf{w}' \mathbf{D} + \lambda$$

See handout for details.

Geostatistical Data: Universal Kriging



Geostatistical Data: Universal Kriging

In universal kriging we don't assume that μ is estimated as a constant, but rather we assume a trend in x and y , and we may possibly include other spatially varying covariates. The general equation is $Z(s) = \mu(s) + \epsilon(s)$ where

$$\mu(s) = \sum_{k=1}^p \beta_k x_k(s_i).$$

- ▶ In addition to estimating the parameters of the spatially varying covariates, we have spatial correlation in the data that is not explained by the covariates.
- ▶ $Z(s) \sim MVN(X(s)\beta, \Sigma)$
- ▶ The term $X(s)\beta$ is referred to as the mean structure (often describing large scale variation or trend) and is distinguished from Σ which is the residual small-scale variation.
- ▶ The residual process adjusts the model for any residual spatial variation after accounting for the covariates.
- ▶ Universal kriging is sometimes called spatial regression.

Geostatistical Data: Universal Kriging

With $Z(s) \sim MVN(X\beta, \Sigma)$, under intrinsic stationarity Σ is our spatial covariance function $C(h)$ for spatial lags $h = s_i - s_j$ and $X\beta$ represents the $k = 1, \dots, p$ covariates.

- ▶ It is possible to account for all spatial covariance in your response with spatially varying covariates. More often than not, there remains spatial covariance that cannot be explained, so we need to estimate Σ .
- ▶ There can be spatial confounding if your covariates vary spatially at the same scale. Be aware of this potential issue.

Geostatistical Data: Review of Regression

- ▶ Recall ordinary least squares, $\hat{\beta}_{OLS} = (X'X)^{-1}X'Y$
- ▶ Note: using $Y(s_i)$ here for regression versus $Z(s_i)$ in kriging equations, but they both represent the spatial process we wish to model.
- ▶ The variance $\hat{\sigma}_{OLS}^2 = \frac{(Y - X\hat{\beta}_{OLS})'(Y - X\hat{\beta}_{OLS})}{N-p}$
- ▶ The OLS estimators of our regression parameters are unbiased (and the confidence intervals on the estimates are correct) when the model is correctly specified. Our covariates correctly specify the model and the previous assumptions are met.

Geostatistical Data: Review of Regression

- ▶ The variance of our parameter estimates $Var(\hat{\beta}_{OLS}) = \hat{\sigma}_{OLS}^2(X^tX)^{-1}$
- ▶ Our variance-covariance matrix is $\Sigma = \sigma^2\mathbf{I}$
- ▶ There no covariance between errors
- ▶ These are all incorrect if there is residual spatial correlation unaccounted for in covariates. Our estimates of β_{OLS} will be biased

Geostatistical Data: Maximum Likelihood

- ▶ There is covariance between errors, i.e. $Cov(Y(s_i), Y(s_j))$
- ▶ Our variance-covariance matrix is $\Sigma = var(\epsilon) = \sigma^2 \mathbf{V}$
- ▶ We need to find the generalized least squares estimates $\hat{\beta}_{GLS}$
- ▶ $\hat{\beta}_{GLS} = (X^t \Sigma^{-1} X)^{-1} X^t \Sigma^{-1} Y$
- ▶ $\hat{\sigma}_{GLS}^2 = \frac{(Y - X \hat{\beta}_{GLS})^t \Sigma^{-1} (Y - X \hat{\beta}_{GLS})}{N-p}$

Geostatistical Data: Generalized Least Squares

- ▶ The variance of our parameter estimates $Var(\hat{\beta}_{GLS}) = (X^t \Sigma^{-1} X)^{-1}$
- ▶ Note, if we used OLS when there was spatial correlation, we would get the wrong estimates of uncertainty in our parameter estimates.
- ▶ The GLS estimator is the minimum variance linear unbiased estimator so it always has lower variance than the OLS estimator.

Geostatistical Data: Generalized Least Squares

- ▶ We want to use GLS but the problem is we don't know Σ
- ▶ As before, we need to parameterize our spatial covariance, but we don't know the parameters
- ▶ We need to estimate β s and Σ

Geostatistical Data: Generalized Least Squares

- ▶ There are two methods for approaching this problem:
 1. Iteratively Re-weighted Generalized least squares
 2. Maximum Likelihood (or Restricted Maximum Likelihood)

Geostatistical Data: Generalized Least Squares

Iteratively Re-weighted Generalized least squares

► STEPS:

1. Obtain a starting estimate of $\hat{\beta}$ from OLS
2. Compute the residuals $r = Y - X\hat{\beta}$
3. Estimate the semivariogram/covariance parameters $\theta = (\tau^2, \sigma^2, \phi)$ based on r .
This gives us $\Sigma(\hat{\theta})$
4. Obtain a new estimate of $\hat{\beta}$ using GLS with $\Sigma(\hat{\theta})$
5. Iterate steps 2-4 until the estimates of $\hat{\beta}$ are small (converge $< 10^{-5}$)

Geostatistical Data: Maximum Likelihood

The probability model for the data is $Y \sim MVN(X\beta, \Sigma)$. The likelihood

$$L(\beta, \theta; Y) = (2\pi)^{-n/2} |\Sigma(\theta)|^{-1/2} \exp\{-(1/2)(Y - X\beta)^t \Sigma(\theta)^{-1} (Y - X\beta)\}$$

We maximize the log likelihood

$$\log L(\beta, \theta; Y) = \frac{-n}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma(\theta)| - \frac{1}{2} (Y - X\beta)^t \Sigma(\theta)^{-1} (Y - X\beta)$$

Our covariance matrix $\Sigma(\theta) = Cov(s_i, s_j)$ is spatial as before, with parameters $\theta = (\tau^2, \sigma^2, \phi)$. This cannot be solved analytically so numerical methods are implemented (e.g. Newton Raphson).

For any fixed θ, θ_0 , the unique value of β that maximizes the log-likelihood is

$$\hat{\beta} = (X^t \Sigma^{-1}(\theta_0) X)^{-1} X^t \Sigma^{-1}(\theta_0) Y$$

For $\hat{\theta}$ we maximize the log-likelihood

$$\log L(\theta; Y) = \frac{1}{2} \log |\Sigma(\theta)| - \frac{1}{2} Y^t P(\theta) Y$$

And now

$$\hat{\beta} = (X^t \Sigma^{-1}(\hat{\theta}) X)^{-1} X^t \Sigma^{-1}(\hat{\theta}) Y$$

Geostatistical Data: Maximum Likelihood

From our log likelihood for finding $\hat{\theta}$

$$P(\theta) = \Sigma^{-1}(\theta) - \Sigma^{-1}(\theta)X(X^t\Sigma^{-1}(\theta)X)^{-1}X^t\Sigma^{-1}(\theta)$$

Note we have eliminated β from this part of the likelihood. $\log L(\theta; Y)$ is called the profile log-likelihood of θ . The solution is obtained numerically by Newton-Raphson or Fisher scoring (very similar).

$$\theta^{k+1} = \theta^k + (B^{(k)})^{-1}\nabla$$

The $k+1$ iterate is found by the previous iterate k . The matrix $(B^{(k)})^{-1}$ is the inverse of the information matrix (second partial derivatives of the log likelihood), and ∇ is the gradient (first partial derivatives of the log likelihood).

Geostatistical Data: Maximum Likelihood

In R we use the geoR package with the following:

- ▶ We can use the `proflik()` to get the profile likelihood estimates of $\theta = (\tau^2, \sigma^2, \rho)$.
- ▶ If we don't want to do the profile likelihood, we can fit the covariance function with `likfit()`.
- ▶ We set up the kriging with the `krige.control()` function. Here we specify trends and covariates and pass the profile likelihood fit for θ
- ▶ We execute the predictions to unobserved locations s_0 with `krige.conv()`, which stands for "conventional kriging".

Geostatistical Data: Hypothesis Testing and Model Comparison

Performing a hypothesis test of the parameters β and θ can be viewed as a comparison of two models: the "full" model and the "reduced" model. These correspond to the alternative and null hypotheses and can be carried out via the likelihood ratio test.

The assumptions of the likelihood ratio tests are that $-2\loglik$ is asymptotically distributed under the null hypothesis as a chi-squared random variable with q degrees of freedom (q being the difference in the degrees of freedom between the full and reduced models).

Non-nested models can be compared with the AIC (Akaike's Information Criterion) or the BIC (Schwarz's Bayesian Information Criterion).

$$\text{AIC} = -2 \log l(\hat{\beta}, \hat{\theta}) + 2(p + m)$$

$$\text{BIC} = -2 \log l(\hat{\beta}, \hat{\theta}) + 2(p + m) \log n$$

The smaller AIC or BIC indicates the better model.

Geostatistical Data: Alternate Universal Kriging

- ▶ We have the choice between detrending data and then using ordinary kriging versus implementing universal kriging
- ▶ Detrending means using GLS to estimate $\hat{\beta}_{GLS}$
- ▶ Then $\hat{Z}(s_0) = x(s_0)\hat{\beta}_{GLS} + \hat{\epsilon}(s_0)_{OK}$
- ▶ This should give the same predictions as universal kriging as long as the residuals are estimated using GLS and the covariance function for OK and UK is the same
- ▶ HOWEVER, the prediction errors are NOT the same.
- ▶ The kriging variance with detrended data does not reflect the uncertainty in estimating $\hat{\beta}_{GLS}$

Geostatistical Data: Other Types of Kriging

- ▶ We have gone through the primary types of spatial prediction via the covariance function
 - Simple kriging (known mean), ordinary kriging (unknown mean, estimated), universal kriging (unknown trend, covariates estimated)
- ▶ Spatial prediction using the above techniques can be seen as point interpolation, taking a set of points as inputs and enabling prediction on a grid (raster)
- ▶ Spatial prediction may also be done at a specific point or set of points if desired
- ▶ There are additional types of kriging:
 - Cokriging
 - Indicator kriging

Geostatistical Data: Cokriging

- ▶ Cokriging is a multivariate version of kriging
- ▶ Two interrelated variables measured at s , harness the correlation between them
- ▶ Often used when one variable of interest is measured more sparsely than another and it is desirable to predict the sparse variable
- ▶ Helps minimize the variance of the estimation error by exploiting the cross-correlation between variables because cross-correlated information contained in the secondary variable should help reduce the variance of the estimation errors

Geostatistical Data: Cokriging

- ▶ Cross-variograms

$$2\nu_{1,2}(h) = Cov(Z_1(s + h) - Z_1(s), Z_2(s + h) - Z_2(s))$$

$$2\gamma_{1,2} = Var(Z_1(s + h) - Z_2(s))$$

- ▶ Note that the first equation relies on Z_1 and Z_2 being measured at the same locations s and $s + h$
- ▶ The second equation does not require observations at the same locations, but only at distances h

Geostatistical Data: Cokriging

- ▶ Cokriging enables the use of observed values of two variables of interest to predict $Z_1(s_0)$ or $Z_2(s_0)$
- ▶ We must assume the following exist for Z_1 and Z_2

$$2\gamma_1(s_i - s_j)$$

$$2\gamma_2(s_g - s_h)$$

$$2\gamma_{1,2}(s_i - s_g)$$

for all s_i, s_j, s_g, s_h in the domain D

- ▶ Co-kriging can be done in R gstat (not geoR). See Bivand, Pebesma, Gomez-Rubio: Applied Spatial Data Analysis with R Chapter 8.

Geostatistical Data: Indicator Kriging

- ▶ When we have point locations but the outcome is binary, but based on an exceedance, for example $P(Z(s_0) > |Z_1, \dots, Z_N)$
- ▶ Categorization is based on a transformation of the original (continuous) data
- ▶ It is a way to make a probability map, whereby the probability is that of a certain value being exceeded

$$\begin{aligned} I(Z(s) > z) &= && 1 && \text{if } Z(s) > z \\ &= && 0 && \text{otherwise} \end{aligned}$$

Geostatistical Data: Indicator Kriging

- ▶ The probability can be estimated by kriging the indicator $I(Z(s_0) > z)$ from the indicator data $I(Z(s_1) > z), \dots, I(Z(s_N) > z)$
- ▶ Giving the estimate $E(I(Z_0) > z | I(Z(s_i) > z))$ which is an estimate of $P(Z(s_0) > | Z_1, \dots, Z_N)$

$$I_{OK}(s_0) = \sum_{i=1}^N \lambda_i I(Z(s_i) > z)$$

- ▶ Indicator kriging

Geostatistical Data: Interpolation & Smoothing

Other types of spatial prediction

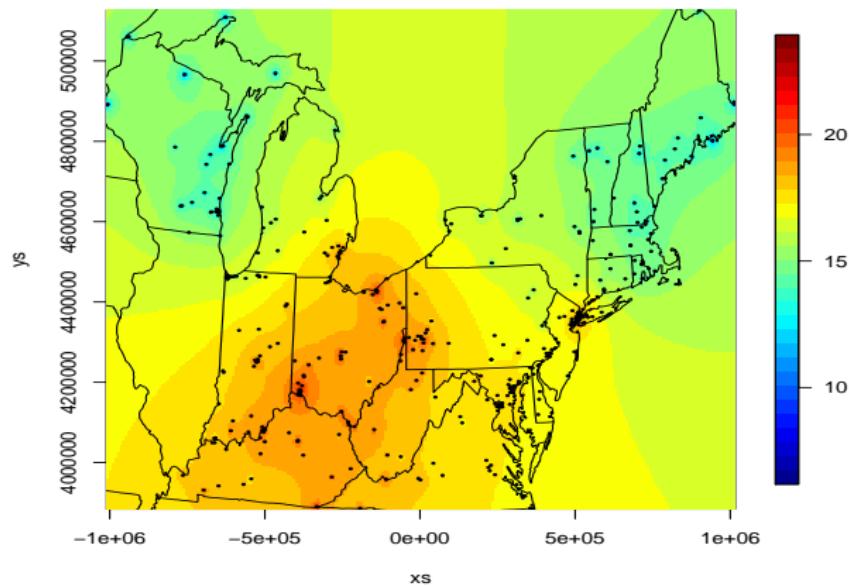
- ▶ **Inverse-distance weighted interpolation (IDW)**
- ▶ Kernel smoothing (local weighted averaging)
- ▶ Loess smoothing (local polynomial fitting)
- ▶ **Spline smoothing**

Geostatistical Data: IDW

- ▶ Inverse distance weighting
- ▶ Weight observations by the inverse of the distance between its location and the prediction location
- ▶ $\hat{z}_0 = \frac{\sum_i Z(s_i) d(s_i, s_0)^{-p}}{\sum_i d(s_i, s_0)^{-p}}$
- ▶ Larger p relates to a more localized weighting, and small p is smoother
- ▶ This is not considered statistical smoothing because there is no statistical estimation that gives standard errors for our predictions

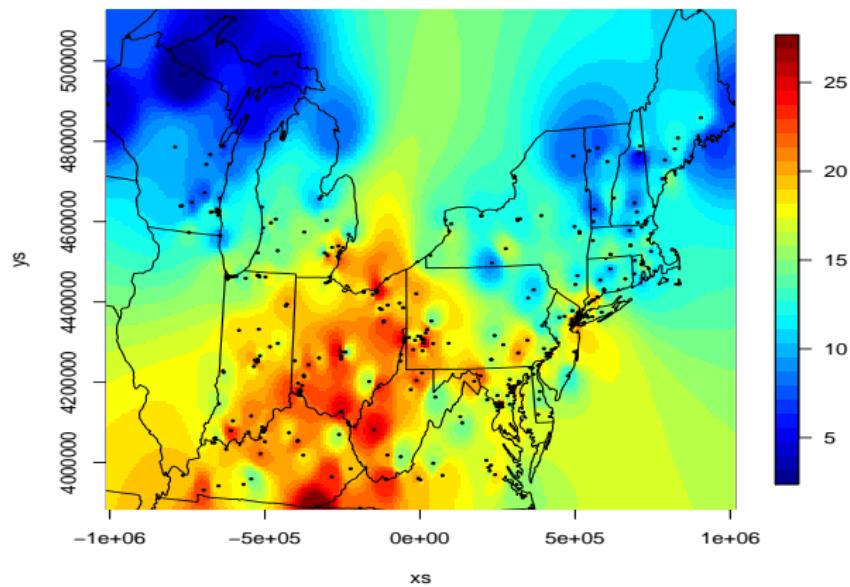
Geostatistical Data: IDW

p=1



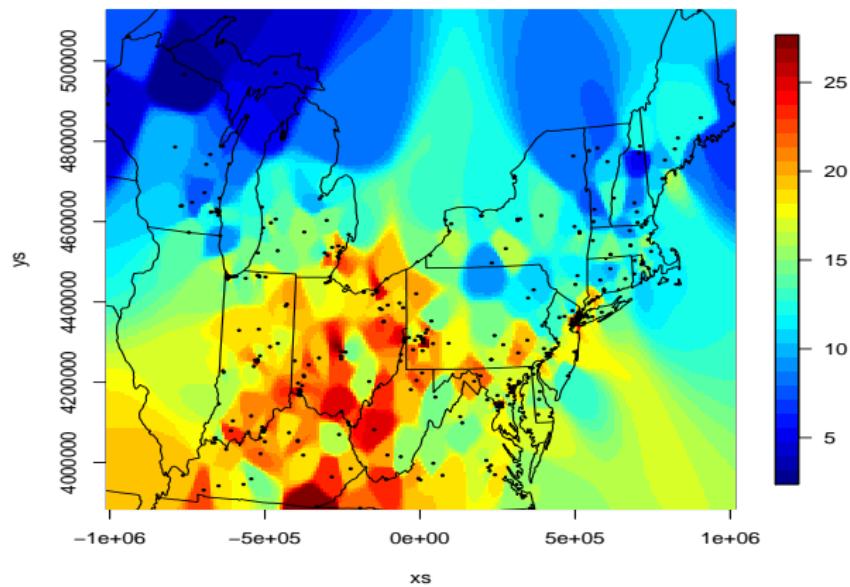
Geostatistical Data: IDW

p=3



Geostatistical Data: IDW

$p=12$



Geostatistical Data: Spline Smoothing

- ▶ Here we focus on modeling the spatial component in the mean part of a regression model rather than in the covariance
- ▶ In general, $Z(s) = f(s) + \epsilon$ where errors are not correlated
- ▶ $f(s)$ is a "smooth" function described by a basis function, and consists of 'non-linear' terms
- ▶ This is considered statistical smoothing because we can get an estimate of the prediction errors

Geostatistical Data: Basis Functions

Basics of Basis Functions

- ▶ We will start with a 1-dimensional, univariate case. For example this could be seen in time series, where we are modeling time (x) with basis functions.
- ▶ Polynomial bases are a good way to illustrate what is going on. Consider the regression model:

$$y_i = f(x_i) + \epsilon_i$$

and let's expand it out by a polynomial

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \epsilon_i$$

Here $f(x_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4$, a 4th order polynomial. So, $f(x)$ is a function represented by **five** basis functions

$f(x_i) = \sum_{j=1}^5 x^j \beta_j = \sum_{j=1}^5 b_j(x) \beta_j$ that are defined by:

$$b_1(x) = 1, b_2(x) = x, b_3(x) = x^2, b_4(x) = x^3, b_5(x) = x^4$$

Geostatistical Data: Basis Functions

- ▶ In general, a basis is a set of functions that can be added together in a weighted fashion to form a more complicated function
- ▶ Here our weights are the regression coefficients β_j
- ▶ In general, a basis function is represented by $f_i = \sum b_j(x_i)\beta_j$

$$\begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \end{pmatrix} = \begin{bmatrix} 1 & b_1(x_1) & b_2(x_1) & b_3(x_1) & b_4(x_1) & b_5(x_1) \\ 1 & b_1(x_2) & b_2(x_2) & b_3(x_2) & b_4(x_2) & b_5(x_2) \\ 1 & b_1(x_3) & b_2(x_3) & b_3(x_3) & b_4(x_3) & b_5(x_3) \\ 1 & b_1(x_4) & b_2(x_4) & b_3(x_4) & b_4(x_4) & b_5(x_4) \\ 1 & b_1(x_5) & b_2(x_5) & b_3(x_5) & b_4(x_5) & b_5(x_5) \end{bmatrix} \times \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix}$$

Geostatistical Data: Polynomial Basis

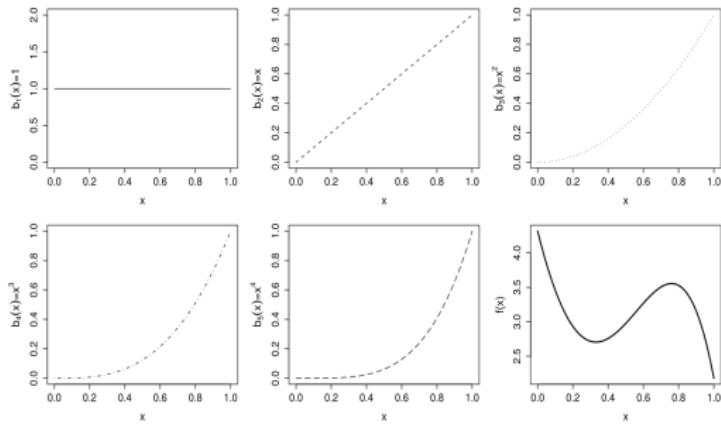
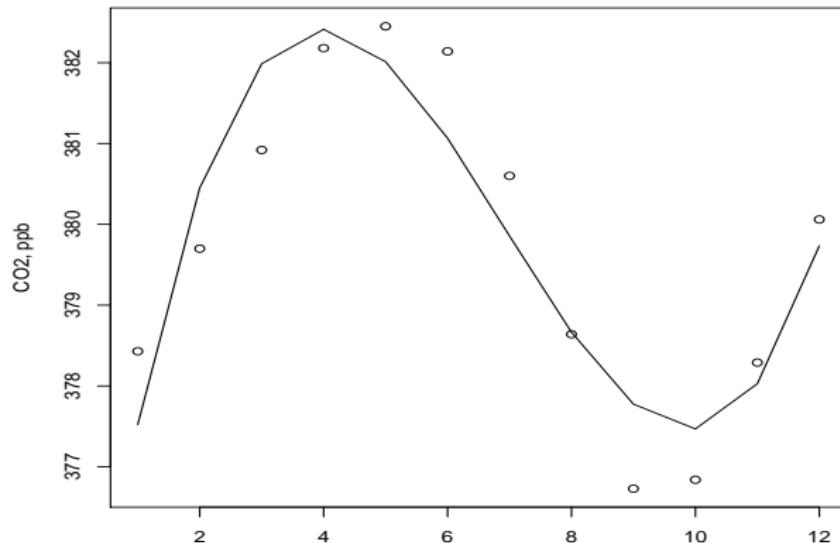


Figure 3.1 Illustration of the idea of representing a function in terms of basis functions, using a polynomial basis. The first 5 panels (starting from top left), illustrate the 5 basis functions, $b_j(x)$, for a 4th order polynomial basis. The basis functions are each multiplied by a real valued parameter, β_j , and are then summed to give the final curve $f(x)$, an example of which is shown in the bottom right panel. By varying the β_j , we can vary the form of $f(x)$, to produce any polynomial function of order 4 or lower. See also figure 3.2

From Wood, 2006 CRC

Geostatistical Data: Polynomial Basis

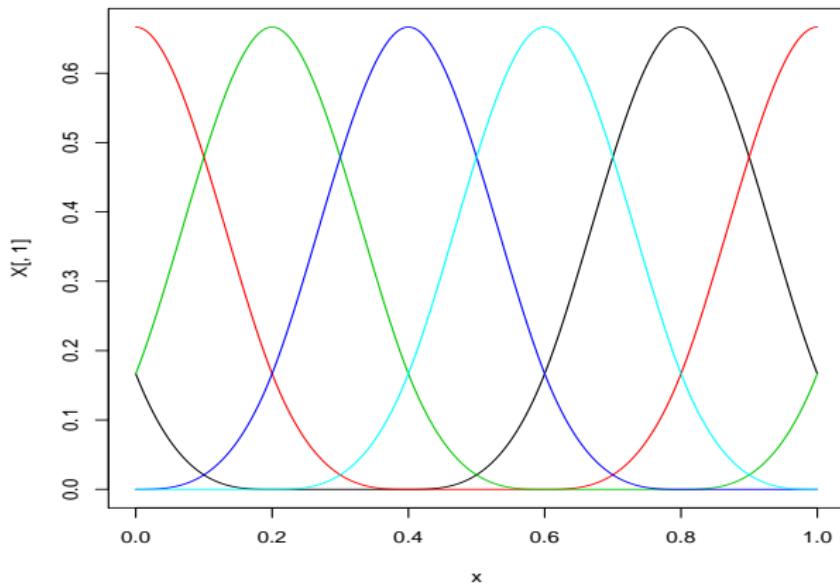
The basis functions are each multiplied by β_j and then summed to give the final curve $f(x)$. In the previous slide, this is shown in the bottom left figure. Below, we show this concept in terms of an example of CO₂ concentrations over a year (monthly data).



Geostatistical Data: Splines

- ▶ In general, splines are curves that are formed by combining pieces of a polynomial.
- ▶ There are several types of splines including natural, cubic, and b-splines (the b stands for basis).
- ▶ $f(t_i) = \sum_{j=1}^4 t^j \beta_j$
- ▶ B-spline curves are made up of polynomial pieces and are defined by a set of knots.
- ▶ Choosing the number of knots defines how smooth (few) or wiggly (many) your functions.

Geostatistical Data: Splines



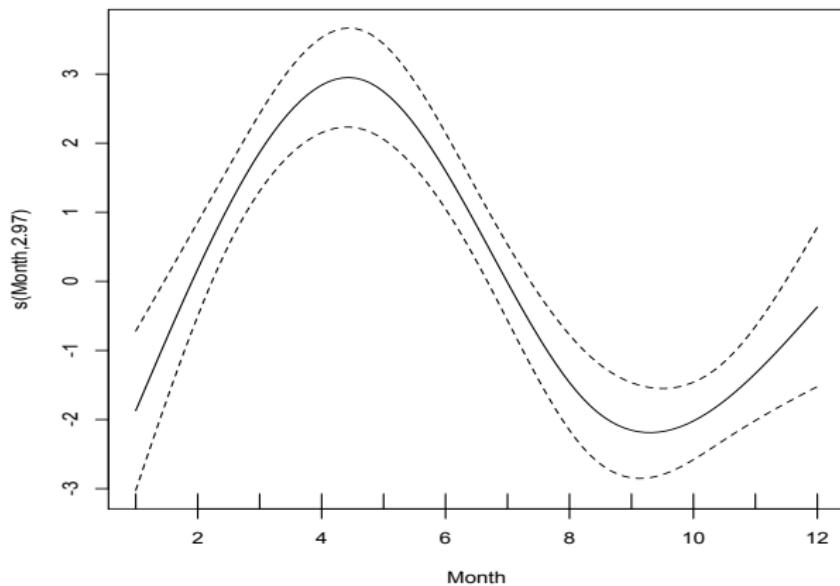
Geostatistical Data: Splines

- ▶ Smoothing splines with penalty allows us to estimate where to put the knots by penalizing the wigginess of the function
- ▶ Minimize the function $\sum_i (y_i - f(t_i))^2 + \lambda \int f''(t)^2 dt$
- ▶ Here, λ is a penalty parameter that controls how much to penalize wiggly functions (roughness penalty).
- ▶ Trade-off between the goodness of fit (the sum of squares) and the wigginess of the function (the integral).
- ▶ Start by putting a knot at every data point, then penalize
- ▶ It is an optimization problem where we minimize:

$$\sum_i (y_i - B_i^T \beta)^2 + \lambda \beta^T S \beta$$

- ▶ the matrix S is constructed by using the spline basis we chose, B is the basis matrix

Geostatistical Data: Splines



Geostatistical Data: Splines

Types of 1-D splines include:

- ▶ cubic splines (basically piecewise cubic polynomials)
- ▶ cyclic splines (cubic splines with connected ends)
- ▶ basis splines (B-spline) with other polynomial orders
- ▶ cardinal splines (where knot placement is always a certain distance)
- ▶ wavelets (often cardinal wavelet splines)

Geostatistical Data: 2-D Splines

- ▶ Thin plate splines are smoothing splines in 2-d
- ▶ Extend the 1-d case to:

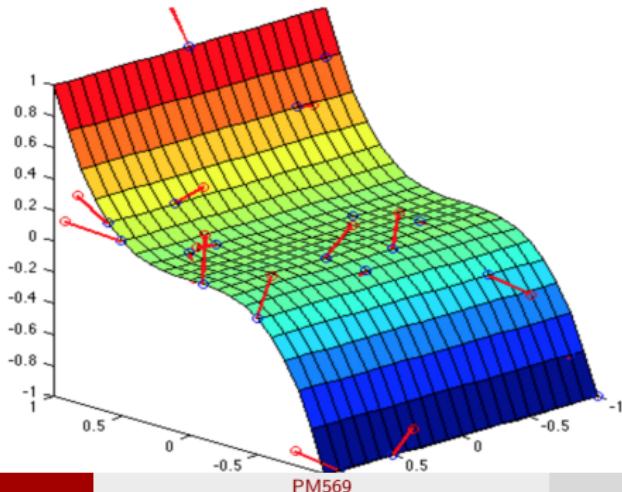
$$\sum_i (z_i - g(s_1, s_2))^2 + \lambda \iint g''(s_1, s_2)^2 ds_1 ds_2$$

- ▶ where the penalty breaks down to the sum of the partial second derivatives
- ▶ λ controls the "wiggliness" as in the 1-D spline (roughness penalty)

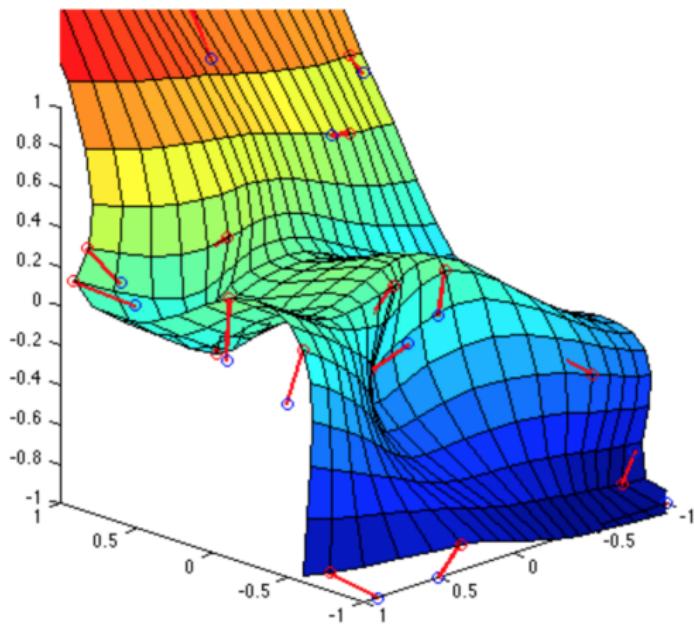
Geostatistical Data: Thin Plate Splines

The idea behind a thin plate spline is:

- ▶ Basically we put a bendable plane through over the space and the points in the space pull the plane (by way of knots)
- ▶ Where there are more points grouped, we expect the plane to be pulled more significantly
- ▶ If there is a very bumpy surface, there will be more knots used and a more wiggly surface

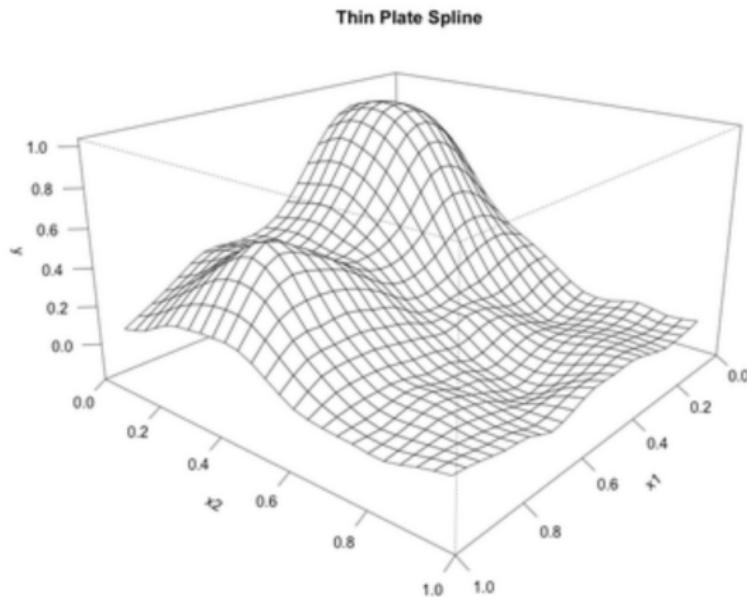


Geostatistical Data: Thin Plate Splines

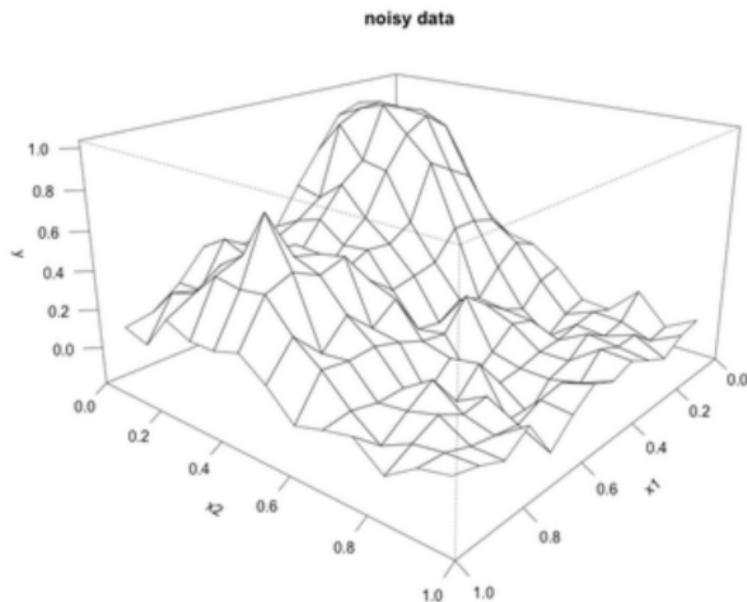


Geostatistical Data: Thin Plate Spline Regression

- The height of where the surface is pulled is going to depend on the magnitude of what we are modeling, $Z(s)$

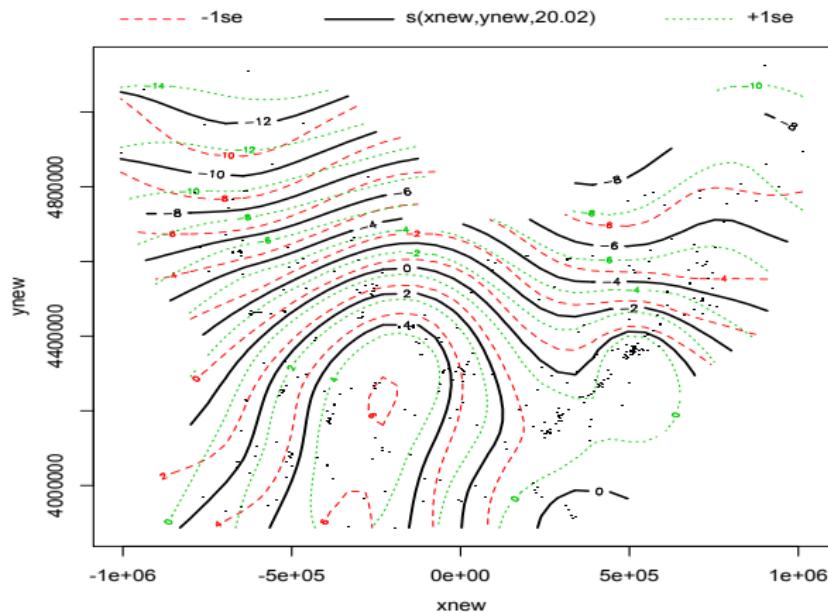


Geostatistical Data: Thin Plate Spline Regression



Geostatistical Data: Thin Plate Spline Regression

Applied to the PM data, we can see that a contour map shows where the points are and where the concentrations are higher

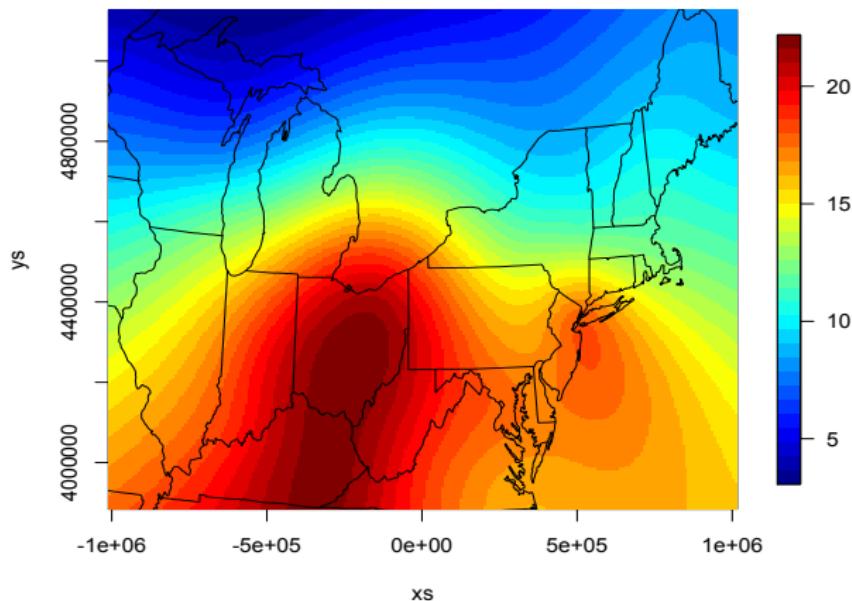


Geostatistical Data: Thin Plate Spline Regression

- ▶ Can use the fitted spline model to predict as we did with kriging
- ▶ Also useful as we can generate standard errors for our predictions
- ▶ Implemented in R using the `gam()` and `predict.gam()` functions in the `mgcv` library
- ▶ There are other 2-D splines, the most common is a tensor product

Geostatistical Data: Thin Plate Spline Regression

Predicted PM_{2.5} using a thin plate spline of the coordinates in a GAM model



Geostatistical Data: Thin Plate Spline Regression

Standard errors of prediction from the GAM model

