Overview of Statistical Analysis of Spatial Data
Geog 210C
Kriging and other Interpolation Approaches

Chris Funk
Outline

- Future labs and final project
- Interpolation Algorithms-General
- Review-Kriging systems
- Simple and Universal Kriging
- Recap
General Formulation for Interpolation

 propiedad lineal de interpolación: el valor previsto de la función objetivo \( y(t_m) \) en el soporte de la fuente \( t_m \) es expresado como una suma ponderada de los \( N \) datos de fuente \( y(s_n), n = 1, \ldots, N \):

\[
\hat{y}(t_m) = \sum_{n=1}^{N} w_{mn} y(s_n)
\]

\( w_{mn} \) = peso dado al \( n \)-ésimo dato de fuente \( y(s_n) \) para la predicción en el m-ésimo soporte de la fuente \( t_m \)

Cuando los soportes de la fuente son puntos y definidos en todo el espacio 2D, el correspondiente conjunto infinito de valores interpolados forma un atributo "superficie".
Weights: function of target-to-source distances, 1 for the closest neighbor to \( t_m \), 0 for all others:

\[
 w_{mn} = \begin{cases} 
 1 & \text{if } d_{mn} = \min\{d_{mn}, n = 1, \ldots, N\} \\
 0 & \text{if not}
\end{cases}
\]

Note that weights sum to 1, and lie in the [0, 1] interval.

Target prediction: interpolated value \( \hat{y}(t_m) \) is identified to the datum value \( y(s_n) \) of the source support \( s_n \) closest to \( t_m \):

\[
 \hat{y}(t_m) = \sum_{n=1}^{N} w_{mn} y(s_n) = 1 \cdot y(s_n) + \sum_{\substack{n' = 1 \\ n' \neq n}}^{N} 0 \cdot y(s_{n'})
\]
Nearest Neighbor Spatial Interpolation (II)

Polygon of influence of a point source support:
- region around source location $s_n$, in which any target location $t_m$ is closer to $s_n$ than to any other source location $s_n$; also known as Dirichlet tessellation, Voronoi or Thiessen polygon
- polygon boundaries formed by perpendicular bisectors between any source location $s_n$ and its $N - 1$ neighboring source locations
- polygons not uniquely defined for samples at edges of study region

Polygon of influence of a sample
Nearest Neighbor Spatial Interpolation: Example

Characteristics:
- Target predictions form discontinuous (patchy) surface,
- Convex interpolation: predicted values cannot be smaller (larger) than the smallest (largest) source datum,
- Exact interpolation: target prediction = source datum when a target and a source location coincide,
- Local & global versions are identical.
Spatial Interpolation via Triangulation (I)

- three source points form a Delaunay triangle if their respective (Voronoi) polygons of influence share a common vertex
- *Delaunay triangles are as close as possible to equilateral triangles*
- *linked to the Triangulated Irregular Network (TIN) surface representation*
Weights: are determined by “zooming” in the neighborhood of a target support \( t_m \), and considering only the 3 source points, say, \( s_1, s_2, \) and \( s_3 \), forming the Delaunay triangle in which \( t_m \) falls. The weight given to \( s_1 \) is proportional to the area of the opposite sub-triangle \( A_{m23} \):

\[
    w_{m1} = \frac{A_{m23}}{A_{123}}, \quad w_{m2} = \frac{A_{m13}}{A_{123}}, \quad w_{m3} = \frac{A_{m12}}{A_{123}}
\]

Note that the weights sum to 1 and lie in the [0, 1] interval. The smaller the distance between target location \( t_m \) and a source location \( s_{1} \), the larger the area \( A_{m23} \) of the opposite sub-triangle, the larger the weight \( w_{m1} \).
Target prediction: interpolated value $y(t_m)$ is a weighted sum of the source data, say, $y(s_1)$, $y(s_2)$ and $y(s_3)$, measured at the vertices of the Delaunay triangle in which $t_m$ falls:

$$\hat{y}(t_m) = \sum_{n=1}^{N} w_{mn} y(s_n) = \frac{A_{m23}}{A_{123}} \cdot y(s_1) + \frac{A_{m13}}{A_{123}} \cdot y(s_2) + \frac{A_{m12}}{A_{123}} \cdot y(s_3)$$

Characteristics: resulting interpolated “surface” has discontinuities at triangle edges, and consists of triangular “facets”, (ii) exact & convex interpolation, (iii) local interpolation
Weights: function of number $N(t_m)$ of source supports falling in a circle of radius $r$, or $r(t_m)$ for an adaptive version, centered at $m$-th target support $t_m$:

$$w_{mn} = \begin{cases} 
1/N(t_m) & \text{if } d_{mn} \leq r \\
0 & \text{if } \text{not}
\end{cases}$$

Note that weights sum to 1, and lie in the [0, 1] interval

Procedure: for a given fixed search radius $r$, or an adaptive radius $r(t_m)$:

1. consider a circle of radius $r$, or $r(t_m)$, centered at target support $t_m$
2. find the $N(t_m)$ source supports within that circle
3. each of these $N(t_m)$ source supports receives a weight of $1/N(t_m)$, and all other source supports a weight of 0
Spatial Interpolation via Local Averaging:

Example

- Target prediction: equal-weighted average of these $N(t_m)$ source data within circle of radius $r$ centered at target support $t_m$:

$$
\hat{y}(t_m) = \sum_{n=1}^{N} w_{mn} y(s_n) = \sum_{n=1}^{N(t_m)} \frac{1}{N} y(s_n) + \sum_{n'=1}^{N-N(t_m)} 0y(s_{n'})
$$

Characteristics: smoothness of resulting interpolated “surface” depends on radius $r$, (ii) non-exact & convex interpolation, (iii) only local version makes sense; adaptive (with spatially varying search radius) variants are available.
Weights: are made inversely proportional to (a power $p$) of target-to-source distances:

$$w_{mn} = \frac{d_{mn}^{-p}}{\sum_{n=1}^{N} d_{mn}^{-p}} = \frac{1/d_{mn}^p}{\sum_{n=1}^{N} 1/d_{mn}^p}$$

Note that the weights sum to 1 (due to the denominator), and lie in the $[0, 1]$ interval; the weight $w_{mn}$ is infinity when $t_m \equiv s_n$, i.e., when a target and a source location coincide.
Role of power parameter:

- $p \to 0$: weights become more similar: $w_{mn} \approx 1/N$, $\forall n$; target prediction approaches average of $N$ source data

- $p \to +\infty$: nearest source support, e.g., $s_n$, receives most weight: $w_{mn} \approx 1$, whereas $w_{mn} \approx 0$, $\forall n \neq n$; target prediction approaches value of nearest source datum $y(s_n)$ (nearest neighbor interpolation)
Characteristics:
- smoothness of resulting interpolated “surface” depends on power parameter $p$,
- exact & convex interpolation,
- both global & local variants, as well as adaptive (with spatially varying power $p$) variants are available
Target prediction: at \textit{m-th target support} $t_m$:

\[ \hat{y}(t_m) = \sum_{n=1}^{N} w_{mn} y(s_n) \]

- from left to right: the target prediction \textit{is comprised of weighted contributions} of \textit{N source data values}
- from right to left: the \textit{n-th source datum} $y(s_n)$ contributes amount $w_{mn} y(s_n)$ to the predicted value $y(t_m)$; in other words, any measurement at the \textit{n-th source support} $s_n$ is "dissipated" according to $w_{mn}$ to the particular target support $t_m$
Two closely related spatial interpolation methods exist, whereby a solution (interpolated) surface is sought under the following constraints:
- fidelity to observations at source supports,
- fidelity to smoothness (derivative) characteristics

Radial Basis Function (RBF) interpolation and Spline Interpolation both satisfy these constraints, and account for the interaction between source data.

The kernels selected in these two methods, however, are typically different: RBF often uses Gaussian kernels (but modifies them to account for spatial interactions between source data), whereas Spline Interpolation often uses pre-specified kernels encoded in the name of the spline function, e.g., cubic or thin-plate spline; such kernels are also modified as in the RBF case.

Both these methods solve exactly the same system of equations to get their weights for interpolation; they differ in what goes into that system of equations.
Spatial Interpolation by Splines: Example

Characteristics:
- Smoothness of resulting interpolated "surface" depends on name of spline function used (here bi-harmonic spline)
- Exact & non-convex interpolation
- Both global & local variants are available
- Smoothing (non-exact) variants are also popular, when measurement error exists in the source data
- Could yield unreasonable values at locations far away from the convex hull of the source supports
Spatial Prediction via Simple Kriging (SK)

- Known expected attribute values ("climatologies") at any (source & target) support. Often, in the absence of other information, it is assumed that the expected attribute value is the same everywhere (constant); this implies no first-order effects.

- Target prediction at target support = weighted sum of source data residuals ("anomalies") from known attribute expectations ("climatologies") + expectation at target support.

- Weights account for:
  - correlation between source and target supports
  - redundancy (inverse correlation = precision) between source supports
  - functional form, e.g., exponential, of semivariogram or correlogram model

- SK variance = reliability of target prediction = overall attribute variance reduced by
  - "weighted" influence of "nearby" source supports:
    - "nearby" = statistical proximity (correlation)
    - "weighted" = source redundancy (SK weights)
An intuitive analysis of OK

\[
\begin{align*}
C \cdot w &= D \\
C^{-1} \cdot C \cdot w &= C^{-1} \cdot D \\
I \cdot w &= C^{-1} \cdot D \\
w &= C^{-1} \cdot D
\end{align*}
\]

C terms provide information about the clustering of the data

Multiplying by \(C^{-1}\) ‘whitens’ the data, reducing the effects of clustering

D terms provide a weighting scheme similar to inverse distance weighting
Simple Kriging

SK prediction: \[ \hat{y}(t_m) - \mu(t_m) = \sum_{n=1}^{N} w_m(s_n)[y(s_n) - \mu(s_n)] = w_m^T r_s = w_m^T[y_s - \mu_s] \]

- \( w_m = [w_m(s_n), n = 1, \ldots, N]^T \): \((N \times 1)\) vector of SK-weights assigned to \(N\) source supports for prediction at target support \(t_m\).

- \( r_s = [y(s_n) - \mu(s_n), n = 1, \ldots, N]^T \): \((N \times 1)\) vector of residual data from known expectations \(\mu(s_n)\) at source supports.

\[ \begin{bmatrix} y(s_1) - \mu(s_1) \\ \vdots \\ y(s_n) - \mu(s_n) \\ \vdots \\ y(s_N) - \mu(s_N) \end{bmatrix} \]

Use semivariogram model to determine \(N\) weights at each target support \(t_m\); typically, we use the covariogram model (the kernel) due to computational reasons.
Input: Distance vector and matrix

Source-to-target & source-to-source distances:

\[ d_m = \begin{bmatrix} d_{1m} \\ \vdots \\ d_{nm} \\ \vdots \\ d_{Nm} \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 0 & \cdots & d_{1n'} & \cdots & d_{1N} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ d_{n1} & \cdots & 0 & \cdots & d_{nN} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{N1} & \cdots & d_{Nn'} & \cdots & 0 \end{bmatrix} \]

- as any other spatial interpolation method, one accounts for the proximity of the \( N \) source supports to the target support \( t_m \). \textbf{Note:} Vector \( d_m \) changes from one target support \( t_m \) to another, hence the subscript \( m \)

- unlike other interpolation methods, Kriging also accounts for the proximity between source supports themselves (sample configuration or data layout). \textbf{Note:} For global interpolation, matrix \( D \) of source-to-source distances is the same for all target supports
Covariance transform

From distance matrices to model covariance matrices: Take any distance value $d_{nm}$ and $d_{nn'}$, i.e., any entry in $d_m$ and $D$, and transform it, via the covariogram model or kernel, to a covariance value $\sigma(d_{nm})$ and $\sigma(d_{nn'})$

Source-to-target & source-to-source model covariances:

$$
\sigma_m = \begin{bmatrix}
\sigma(d_{1m}) \\
\vdots \\
\sigma(d_{nm}) \\
\sigma(d_{Nm})
\end{bmatrix}
$$

and

$$
\Sigma = \begin{bmatrix}
\sigma(0) & \cdots & \sigma(d_{1n'}) & \cdots & \sigma(d_{1N}) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
\sigma(d_{n1}) & \cdots & \sigma(0) & \cdots & \sigma(d_{nN}) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
\sigma(d_{N1}) & \cdots & \sigma(d_{Nn'}) & \cdots & \sigma(0)
\end{bmatrix}
$$

- source-to-target covariance vector $\sigma_m$: $(N \times 1)$ vector with model covariance values $\sigma(d_{nm})$ between $N$ source supports and target support $t_m$
- source-to-source covariance matrix $\Sigma$: $(N \times N)$ matrix with model covariance values $\sigma(d_{nn'})$ between any two source supports separated by distance $d_{nn'}$
the weights vector $w_m$ is obtained by solving the SK system anew for each target support $t_m$ since the entries of $\sigma_m$ change from one target support to another.

the SK system has a unique solution (there is one and only one weights vector $w_m$) if and only if the source-to-source covariance matrix $\Sigma$ is positive definite; for 2nd-order stationarity, this implies that a valid covariogram model $\sigma(d; \theta)$, e.g., exponential distance decay, with $\theta$ containing the sill and range, is used to populate $\Sigma$; in this case, $\sigma(d; \theta) = \sigma(0)\rho(d; \theta)$ and the weights do not depend on the sill $\sigma(0)$:
The Simple Kriging System of Equations

\[
\begin{bmatrix}
\sigma(s_1) & \ldots & \sigma(s_1, s_N) \\
\vdots & \ddots & \vdots \\
\sigma(s_N, s_1) & \ldots & \sigma(s_N)
\end{bmatrix}
\begin{bmatrix}
w_m(s_1) \\
\vdots \\
w_m(s_N)
\end{bmatrix}
= 
\begin{bmatrix}
\sigma(s_1, t_m) \\
\vdots \\
\sigma(s_N, t_m)
\end{bmatrix}
\]

\[\sum w_m = \sigma_m\]

→ a system of \(N\) equations in \(N\) unknowns (the weights in \(w_m\)) for prediction at support \(t_m\); there are \(M\) such systems for \(M\) target supports, since \(\sigma_m\) changes from one target support to another

→ a version of the system of normal equations used in multiple linear regression. For Kriging, the dependent variable pertains to the target support \(t_m\), and there are \(N\) predictor (lagged) variables pertaining to the \(N\) source supports
Kriging as regression

- System of normal equations: Per classical (ordinary) least squares regression theory, the vector $\beta$ of regression coefficients can be estimated as:

$$X^TX\beta = X^Ty$$

- Covariance based version: Since all variables are assumed to have 0-mean, this system can be simplified to

$$\sum \beta = \sigma$$

Covariances between predictors

Covariances between predictors and the predicted values
Solving the Simple Kriging (SK) System of Equations

\[
\begin{bmatrix}
w_m(s_1) \\
\vdots \\
w_m(s_N)
\end{bmatrix}
= \begin{bmatrix}
\sigma(s_1) & \cdots & \sigma(s_1, s_N) \\
\vdots & \ddots & \vdots \\
\sigma(s_N, s_1) & \cdots & \sigma(s_N)
\end{bmatrix}^{-1}
\begin{bmatrix}
\sigma(s_1, t_m) \\
\vdots \\
\sigma(s_N, t_m)
\end{bmatrix}
\]

\[w_m = \Sigma^{-1}\sigma_m\]

- The weights vector $w_m$ is obtained by solving the SK system anew for each target support $t_m$ since the entries of $\sigma_m$ change from one target support to another.

- The SK system has a unique solution (there is one and only one weights vector $w_m$) if and only if the source-to-source covariance matrix $\Sigma$ is positive definite.
Once the SK weights are computed as $w_m = \Sigma^{-1}\sigma_m$, they are substituted in the equations below to compute the SK prediction $y(t_m)$ and associated error variance $\sigma(t_m)$:

**SK prediction:**

$$\hat{y}(t_m) = \mu(t_m) + w_m^T r_s = \mu(t_m) + [w_m(s_1) \cdots w_m(s_N)]$$

**SK prediction error variance:**

$$\hat{\sigma}(t_m) = \sigma(t_m) - w_m^T \sigma_m = \sigma(t_m) - [w_m(s_1) \cdots w_m(s_N)]$$

$$\hat{y}(t_m) = \mu(t_m) + \sum_{n=1}^N w_m(s_n)[y(s_n) - \mu(s_n)] \text{ and } \hat{\sigma}(t_m) = \sigma(t_m) - \sum_{n=1}^N w_m(s_n)\sigma(s_n, t_m)$$
Simple Kriging with Local Search Neighborhoods

- Same procedure as before but: at each target support $t_m$, use only closest $N_m << N$ source data $\{y(s_n), n = 1, \ldots, N_m\}$ within a neighborhood around $t_m$ to compute the $(N_m \times N_m)$ source-to-source covariance matrix $\Sigma$ and the $(N_m \times 1)$ source-to-target covariance vector $\sigma_m$

- **Pros:**
  - no need to store and invert a large $(N \times N)$ source-to-source covariance matrix $\Sigma$ in the case of large $N$ (many source data), only $M$ much smaller ($N_m \times N_m$) sub-matrices of $\Sigma$
  - considering a local neighborhood can lead to the inclusion of more relevant source data into the prediction exercise

- **Cons:**
  - need to define rules for specifying the search neighborhood;
  - not a big issue: use circle with radius $\sim$ range of covariogram model
  - too small a search neighborhood reduces the number $N_m$ of source data considered for prediction, and might lead to more uncertain predictions

- Kriging with local neighborhoods is widely used in applications of geostatistics
Spatial Prediction with Kriging Example

- Precipitation data and semivariogram:

- Kriging predictions and error variances:
Statistics of predicted precipitation values:

- Proportions of high or low values not the same as those of source data

Smoothing effect of interpolation:

- Proportions of high or low values not the same as those of source data
- Map of target predictions exhibits much smoother spatial variability than that quantified by variogram model used for interpolation (Kriging);
  - The amount of excess smoothness depends on the source support configuration and density
Kriging with Covariates - I

- **Ordinary Kriging**
  - Add a vector of 1s to your covariance matrix
  - Automatically solves for a constant but unknown mean
  - All weights sum to 1

- **Universal kriging**
  - Add columns to your covariance matrix with spatial locations, regression coefficients for these are estimated during the solution process

- **Kriging with external drift**
  - Add columns corresponding to exogenous predictor variables to your covariance matrix, regression coefficients for these are estimated during the solution process
Kriging with Covariates - II

- Ordinary Kriging
  \[ C \ C \ C \ C \ 1 \]
  \[ C \ C \ C \ C \ 1 \]
  \[ C \ C \ C \ C \ 1 \]
  \[ 1 \ 1 \ 1 \ 0 \]

- Universal Kriging
  \[ C \ C \ x \ y \]
  \[ C \ C \ x \ y \]
  \[ x \ x \ 0 \ 0 \]
  \[ y \ x \ 0 \ 0 \]

- Kriging with External Drift
  \[ C \ C \ r \ l \]
  \[ C \ C \ r \ l \]
  \[ r \ r \ 0 \ 0 \]
  \[ l \ l \ 0 \ 0 \]
Recap

Geostatistical spatial prediction via Simple Kriging accounts for:

- known expected values ("climatologies") at any source or target support
- statistical proximity (correlation) between source and target supports (for the residual component)
- redundancy (inverse correlation) between source supports, distributing weights among nearby supports (again for the residual component)

Simple Kriging is a (no-intercept) linear regression model with (0-mean) lagged variables

Fundamental differences from other methods:

- Everything depends on the model covariogram (which is typically built from a sample semivariogram), thus reflecting the notion that weights should account for the nature of residual spatial variability (smooth versus rough) of a particular attribute
- An output of Kriging is a set of reliability measures for the predicted values. Such measures are encoded in the prediction error variances, which are independent of the actual data values and only depend on the sample configuration and covariogram model; Kriging error variances are often used in sampling design applications