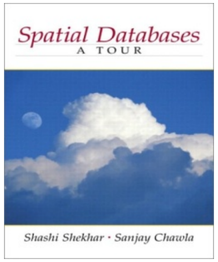
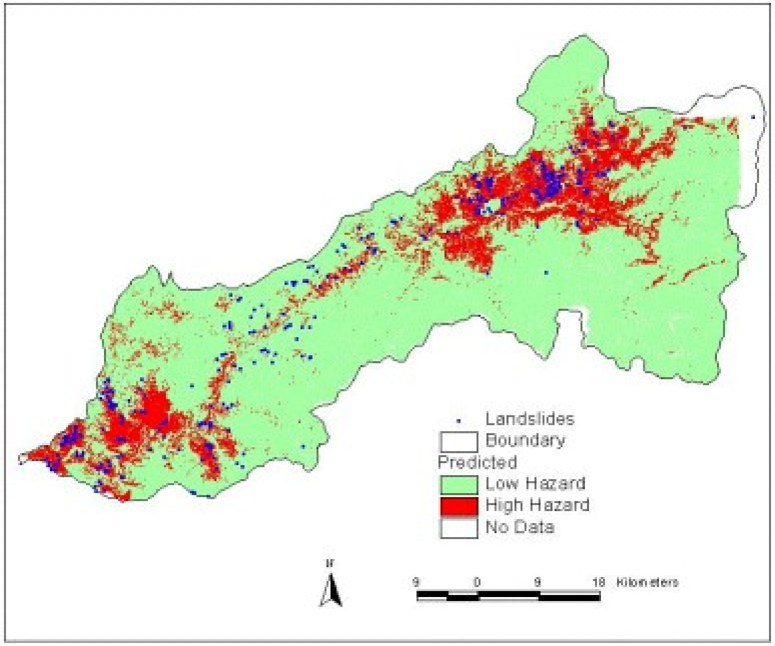
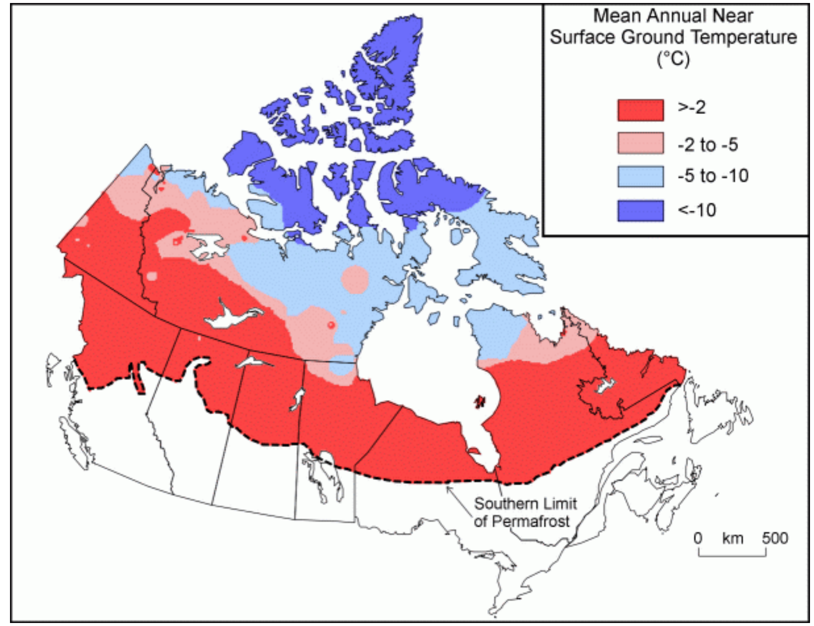


Spatial Data Mining

Regression and Classification Techniques

Spatial Regression and Classisfication

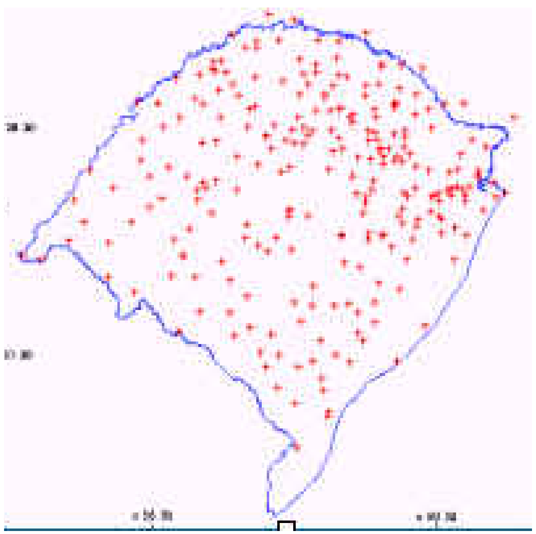




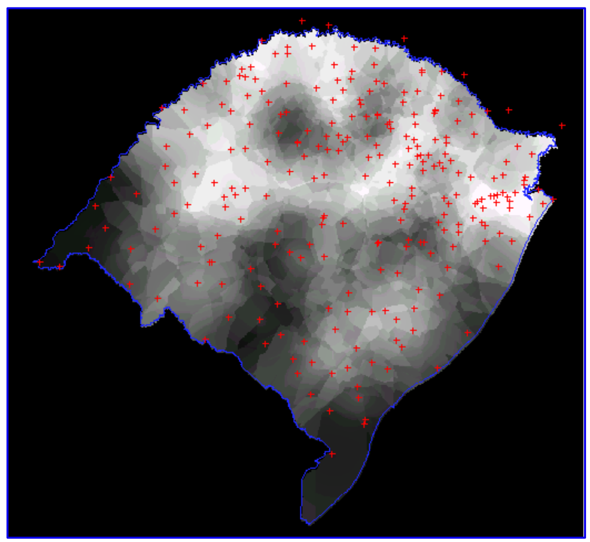
- Discrete class labels (left) vs. continues quantities (right) measured at locations (2D for geographic applications)
- Build a model for predicting the measured quantity at any location
- Additional (to spatial) attributes may exist

Geostatistics

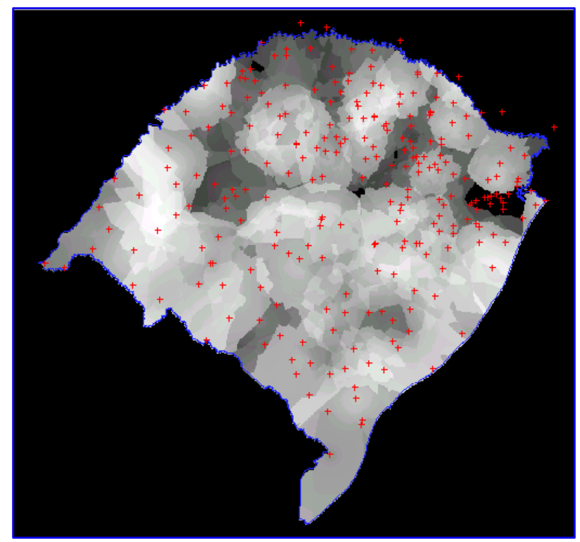
- Analysis and inference of continuously-distributed variables
  - Analysis: Describing the spatial variability of the phenomenon under study
    Inference: Estimating the unknown values
- Questions on measurements:
- How are they distributed? How are they related to each other? How can I infer a distribution from one sample?



Water Availabilty Index



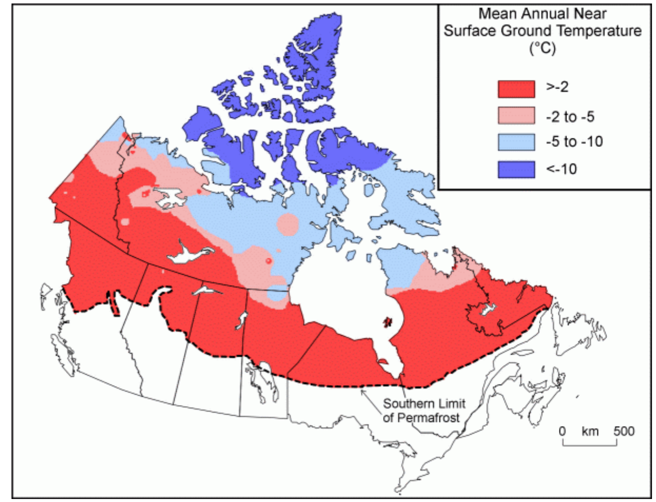
Estimated Surface

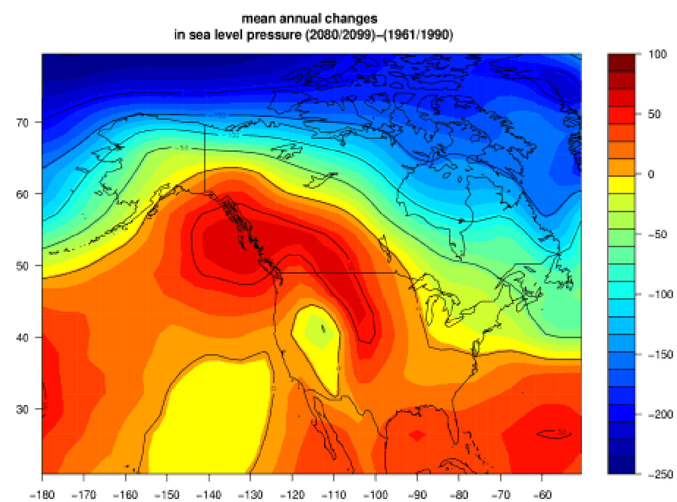


Estimated Uncertainty

Spatial continuity and stationarity

- Why prediction is possible?
  - **Continuity**: Spatial close measurements are more similar than distant ones
- What does it mean?
  - Model the underlying phenomenon with the model f(x,w), x the location vector and w the measurement
  - If not just noise, then continuity creates "smoothness" of w values that can be modeled by f(x,w)
- Can all locations be modeled by a single f(x,w)?
  - **Stationarity**: Measurements generated by a single distribution at all locations





Spatial Autocorrelation

- Continuity produces **autocorrelation**: correlation of a variable with itself through space
  - First law of geography: "everything is related to everything else, but near things are more related than distant things" – Waldo Tobler
- 3 possible cases:
  - If nearby or neighboring areas are more alike, this is *positive spatial autocorrelation*
  - *Negative autocorrelation* describes patterns in which neighboring areas are unlike
  - Random patterns exhibit *no spatial autocorrelation*

Why to bother about spatial autocorrelation?

- Most statistics/data mining methods are based on the assumption that the values of observations in each sample are independent of one another
- Positive spatial autocorrelation may violate this, if the samples were taken from nearby areas
  - Spatial Autocorrelation is a kind of redundancy: the measurement at a location constrains, or makes more probable, the measurement in a neighboring location
  - Models will be biased, since measurements tend to be concentrated and there are actually fewer number of independent observations than are being assumed

Measures of autocorrelation

- Objectives:
  - Measure the strength of spatial autocorrelation
  - Test the assumption of independence or randomness
- Measures
  - Moran's I
  - Variograms
  - other (Geary's C, Ripley's K)

Moran's I: A measure of spatial autocorrelation

- Compares the value of the variable at any one location with the value at all other locations

$$I = \frac{N \sum_i \sum_j W_{i,j}(X_i - \bar{X})(X_j - \bar{X})}{(\sum_i \sum_j W_{i,j}) \sum_i (X_i - \bar{X})^2}$$

- Similar to correlation coefficient, it varies between –1.0 and + 1.0
  - When autocorrelation is high, the coefficient is high
  - A high *I* value indicates positive autocorrelation

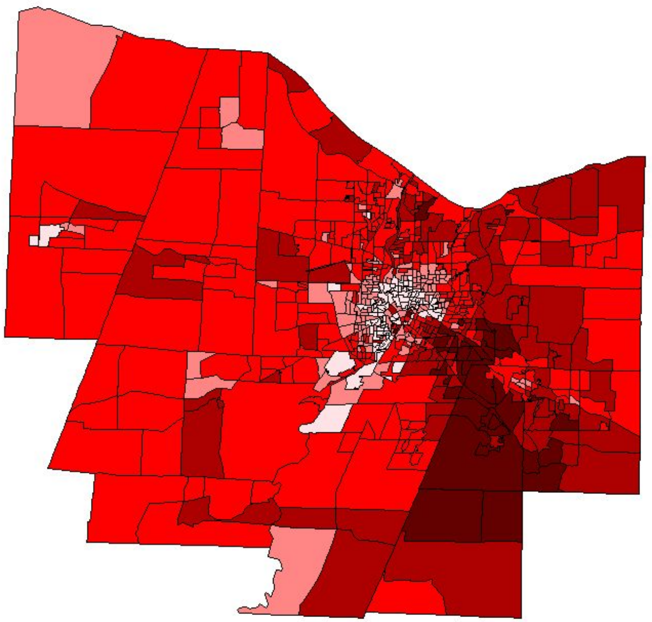
Symbols and Contiguity matrix

$$I = \frac{N\sum_i\sum_j W_{i,j}(X_i-\bar{X})(X_j-\bar{X})}{(\sum_i\sum_j W_{i,j})\sum_i (X_i-\bar{X})^2}$$

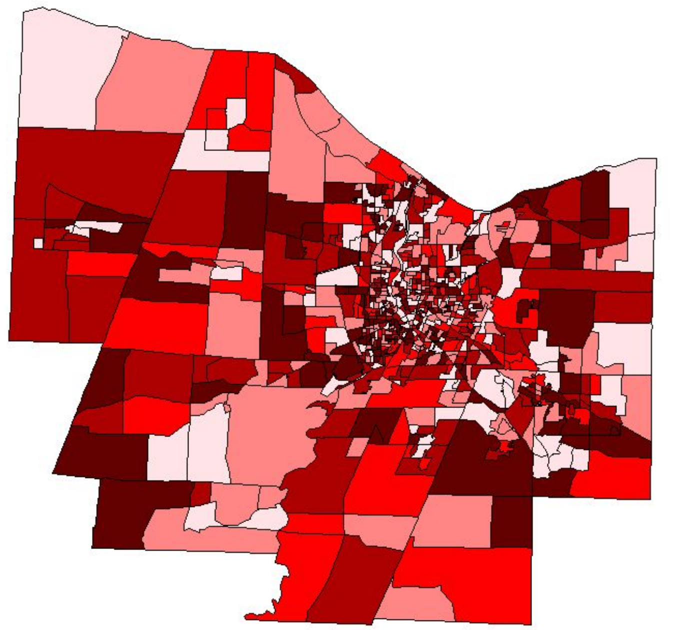
- *N* is the number of cases
  X_i is the variable value at location *i*
  X_j is the variable value at location *j*
  \bar{X} is the mean of the variable
  W_{ij} is a weight applied to the comparison between location *i* and location *j*

- W_{ij} is a contiguity matrix
  - If location *j* is adjacent to zone *i*, the interaction receives a weight of 1
  - Another option is to make W_{ij} a distance-based weight which is the inverse distance between locations I and j ($1/d_{ij}$)

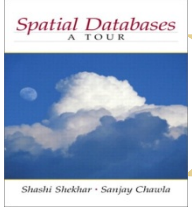
Example: Per Capita Income in Monroe County



Actual values: Moran's I: 0.66



Random values: Moran's I: 0.01



Local Moran's I

- Following Anselin's (1995) definition, a local Moran's I_i may be defined as:

$$I_i = \frac{z_i}{s^2} \sum_j w_{ij} z_j, \; i \neq j$$

- z_is are the deviations from the mean of y_is

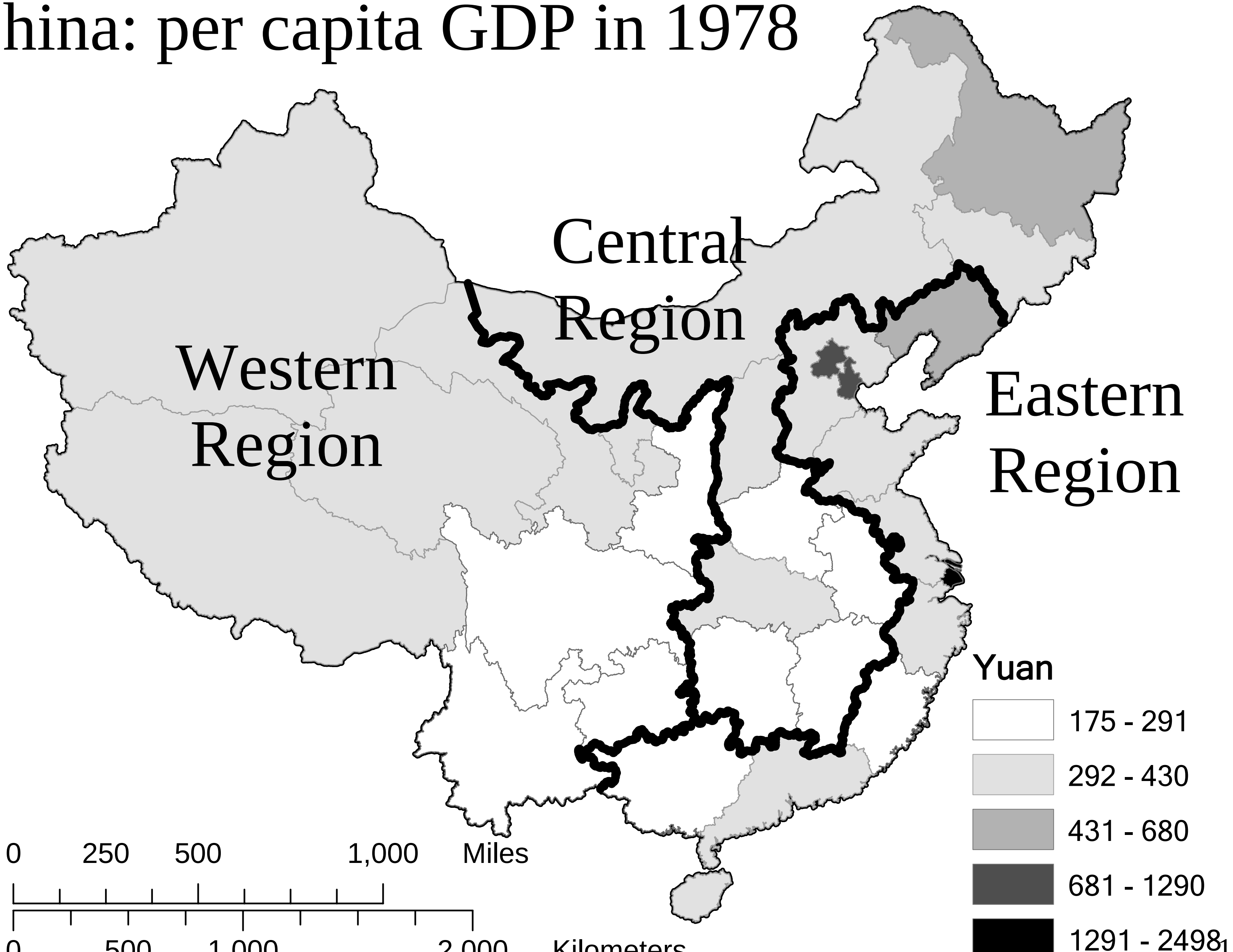
...	...	...	...	...
...	89	71	52	...
...	85	75	63	...
...	51	61	64	...
...	...	...	...	...

$$I_{75} = \frac{75 - 55.82}{675.32}\left[71 + 85 + 61 + 63 - 4 \times 55.82\right] = 1.61$$

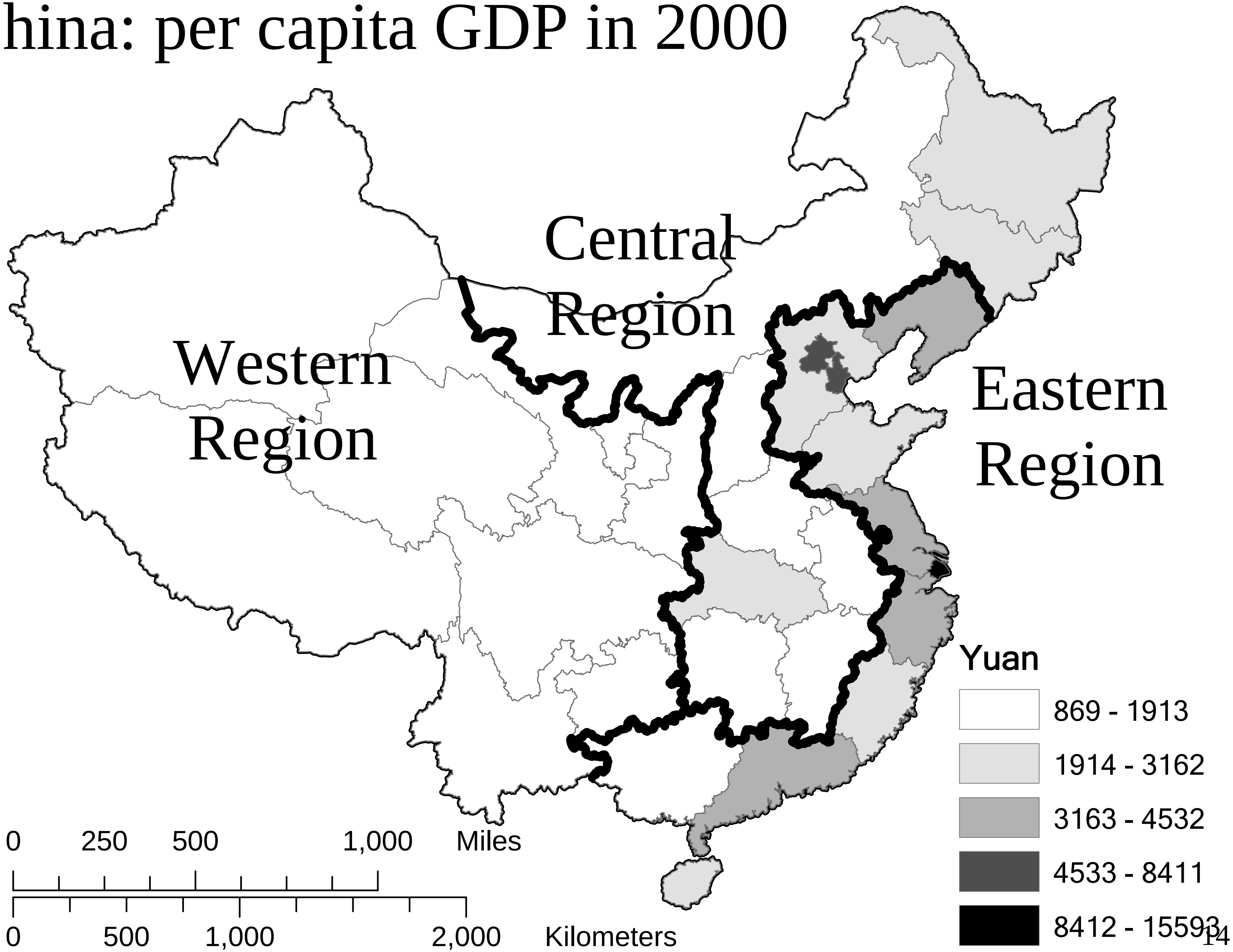
Global vs. Local Moran's I: example

- Spatial pattern detection in China's provincial development
- The variable used: per capita GDP
- Dynamic patterns – global Moran's *I*
- Specific local spatial process – local Moran's *I* and the Moran's scatterplot

China: per capita GDP in 1978

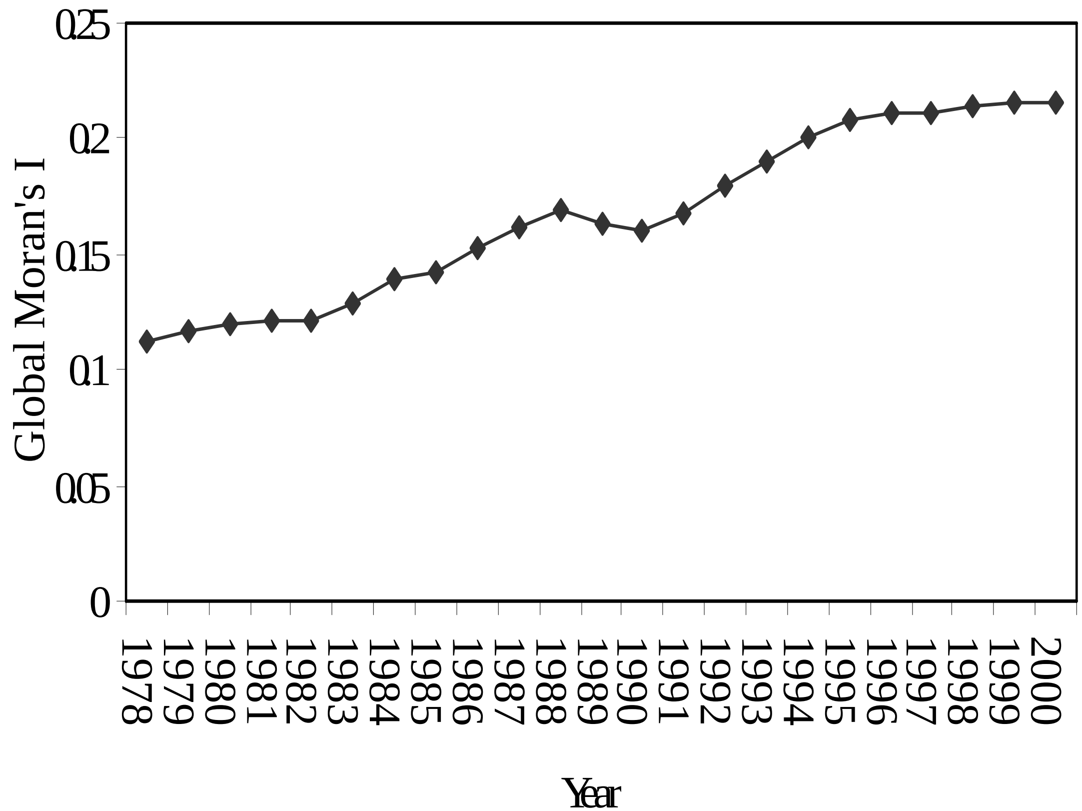


China: per capita GDP in 2000

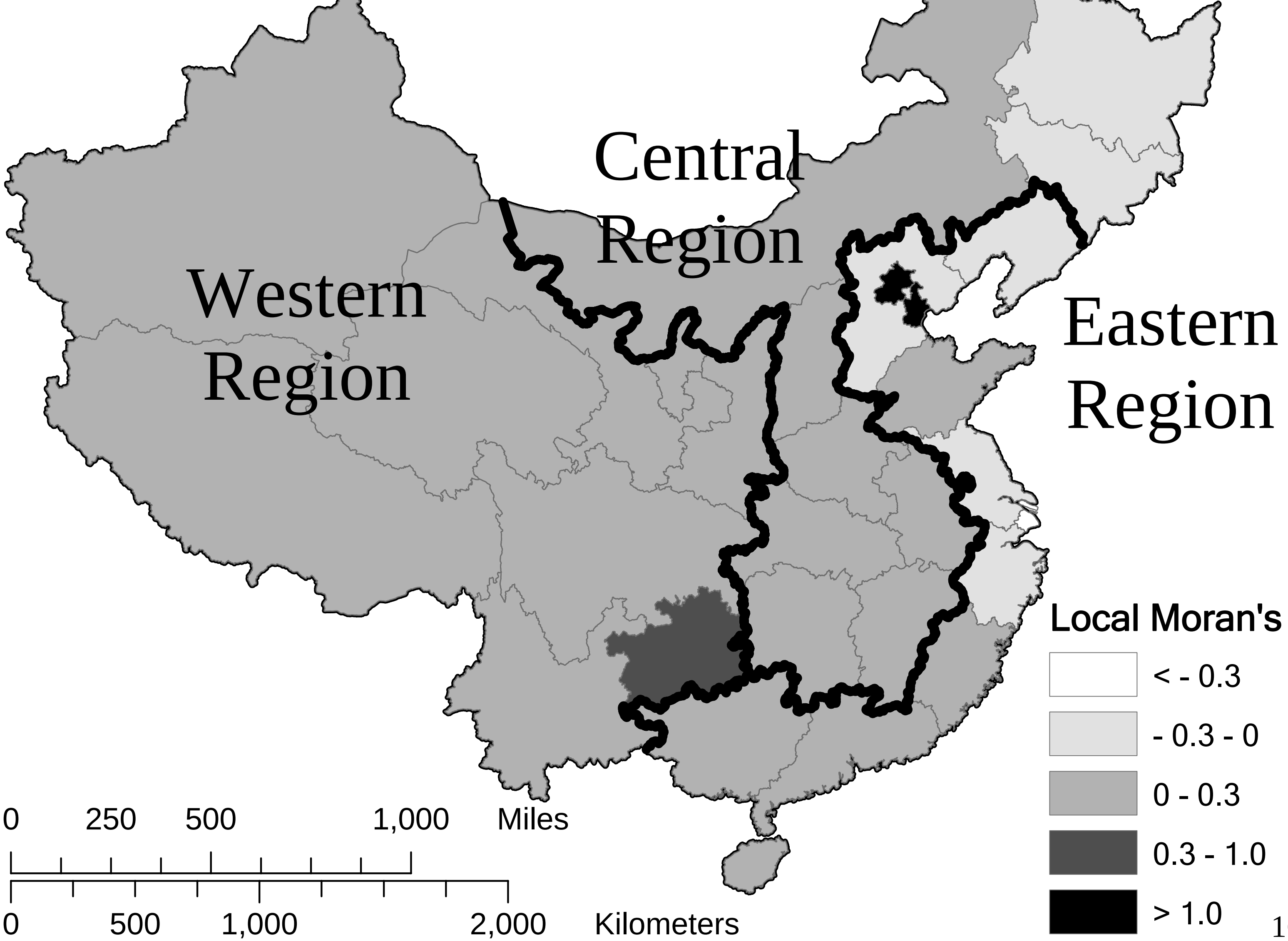


Global vs. Local Moran's I: example

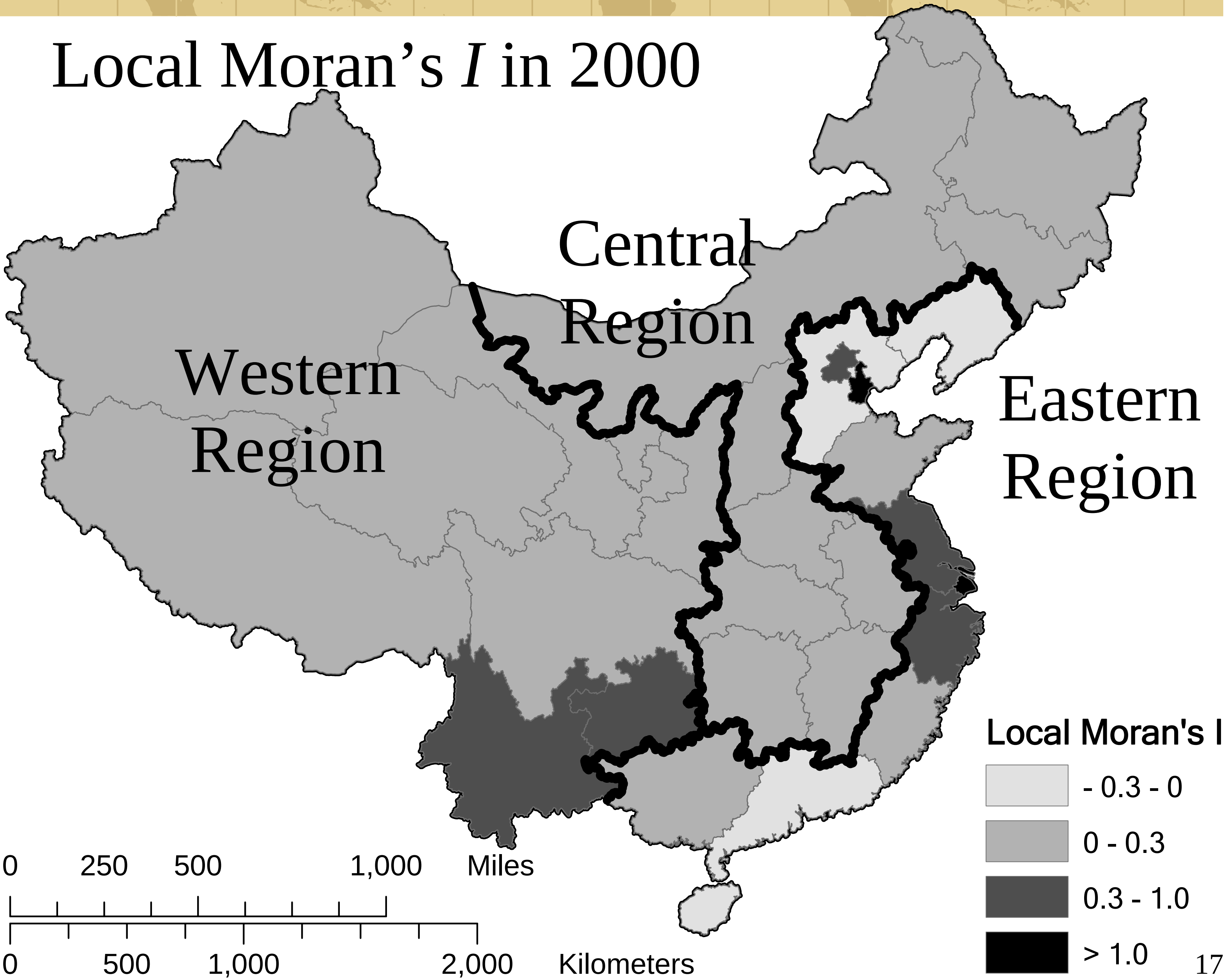
- There is a clustering trend in China's provincial level development (represented by per capita GDP
- But the global Moran's *I* can't tell on which side does the clustering trend take place



Local Moran’s *I* in 1978



Local Moran's *I* in 2000



More details to the Chine GDP example

- First, China's coast-interior divide persisted
  - Interior provinces exhibit great geographical similarity in economic development and spatial contributions to the global Moran's I
- Second, the municipalities (Beijing, Tianjin, Shanghai) always contribute the most
  - Shanghai's position is worth noting, it development changed the spatial pattern the most
- Third, Guangdong's contribution to the global index corresponds with its changing spatial behavior depicted in the Moran scatterplot
- Fourth, while most of the interior provinces have similar patterns, coastal provinces vary greatly
- Fifth, Shandong fell into the low-low quadrant, and contributed very little to the global index
- Sixth, Guizhou and Yunnan, two provinces in southwest China, contributed relatively highly to the global index in 2000
  - The poorest ones tend to form a poor cluster

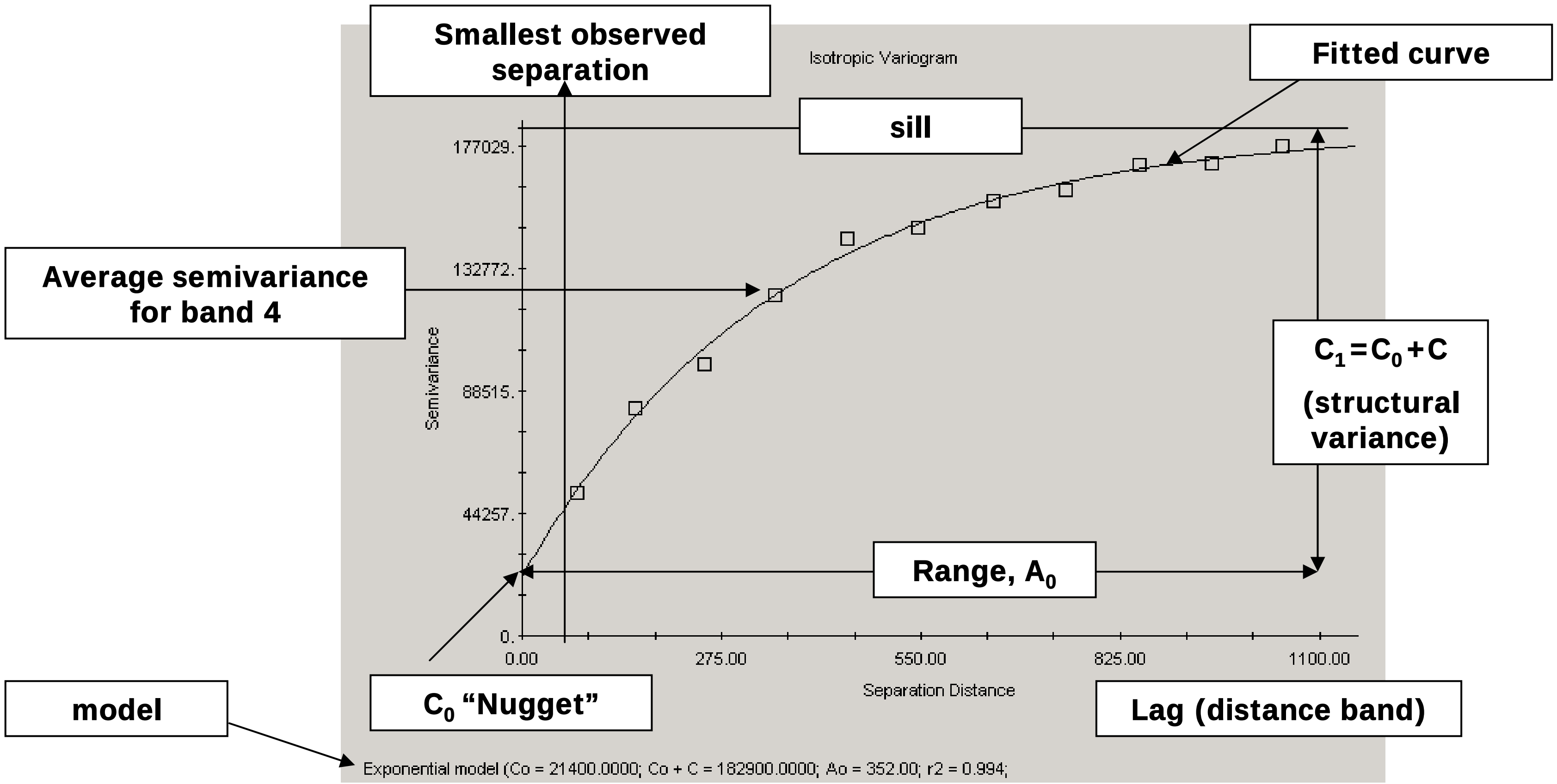
Variograms

- Analyse the observed variation in data values by distance bands using a spatial autocorrelation-like measure, γ:
  - Semivariance measure is most often used:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{d_{ij}=h-\Delta/2}^{d_{ij}=h+\Delta/2} (z_i - z_j)^2$$

  - Bands have width Δ. N(h) is the number of pairs in the band with mid-point distance h
- After building an experimental variogram, we need to fit a theoretical function in order to model the spatial variation

Variograms


Smallest observed separation
Isotropic Variogram
Fitted curve
sill
177029.
132772.
88515.
44257.
0.
Semivariance
Average semivariance for band 4
$C_1 = C_0 + C$
(structural variance)
Range, A_0
0.00
275.00
550.00
825.00
1100.00
Separation Distance
model
C_0 “Nugget”
Lag (distance band)
Exponential model (Co = 21400.0000; Co + C = 182900.0000; Ao = 352.00; r2 = 0.994;

Variograms

Model	Formula (Theoretical Fit)	Notes
Nugget effect	$\gamma(0) = C_0$	Simple constant. May be added to all models. Models with a nugget will not be exact
Linear	$\gamma(h) = C_1(h)$	No sill. Often used in combination with other functions. May be used as a ramp, with a constant sill value set at a range, a
Exponential Exp()	$\gamma(h) = C_1\left(1 - e^{-kh}\right)$	*k* is a constant, often *k*=1 or *k*=3. Useful when there is a larger nugget and slow rise to the sill
Spherical Sph()	$\gamma(h) = C_1\left(\frac{3h}{2} - \frac{1}{2}h^3\right), h < 1$ $\gamma(h) = C_1, h \geq 1$	Useful when the nugget effect is important but small. Given as the default model in some packages.

Approaches to spatial prediction

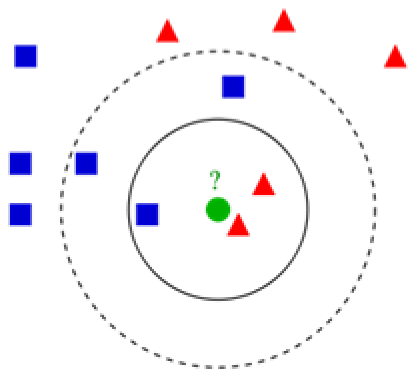
Value of the variable is predicted from **“nearby” samples**

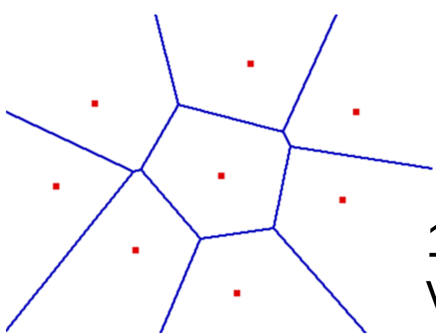
- Example: concentrations of soil constituents (e.g. salts, pollutants)
- Example: vegetation density

- Each interpolator has its own assumptions:
  - Nearest neighbor and variations:
    - Average within a radius
    - Average of *n* nearest neighbors
    - Distance-weighted average within a radius
    - Distance-weighted average of *n* nearest neighbours
  - “Optimal” weighting -> **Kriging**

Nearest Neighbor Methods

- k-NN Classification: assign the class label of the majority of the k-NN





1-NN:
Voronoi Diagram

- k-NN Regression: assign the mean value of the k-NN

$$\hat{Y}(x) = \frac{1}{k} \sum_{x_i \in N_k(x)} y_i$$

A common weighting scheme is to give each neighbor a weight of 1/*d*, where *d* is the distance to the neighbor

Nearest Neighbor Methods

Pros:

- Simple, no training (lazy)
- Benchmark:
  - $E_{1\text{-}NN} \le 2 E_B$
- Often as good as more sophisticated methods
- Per-se considerations of autocorrelation

Cons:

- Slow classification (lazy)
- Prone to noise
- High-variance
- Need to determine k
  - Cross validation

1	2	3	4	5
Train	Train	Validation	Train	Train

- Need to determine weights (for variations)

Nearest Neighbor Methods

- When no spatial autocorrelation (random data):

$$Z = f(X) + \varepsilon(X)$$

$$Z \approx \varepsilon(X)$$

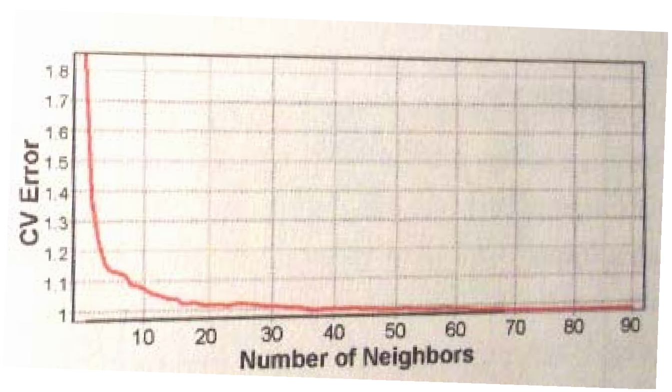
- CV (LOO) error is maximized for 1-NN:

$$E = \frac{1}{N}\sum_{j=1}^{N}\left(Z_j - Z_{j,1NN}\right)^2 \approx \frac{1}{N}\sum_{j=1}^{N}\left(\varepsilon_j - \varepsilon_{j,1NN}\right)^2 \propto 2Var\left(\varepsilon\right) \approx 2Var(Z)$$

Nearest Neighbor Methods

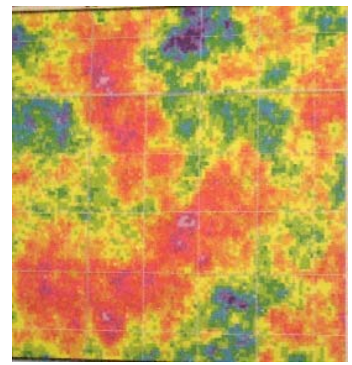
Random data

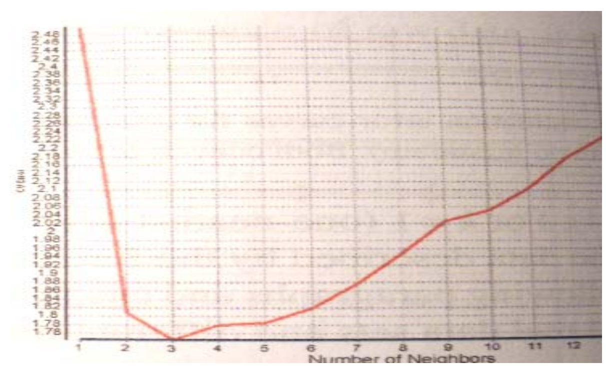
- No minimum occurs



Spatial autocorrelation

- Minimum occurs

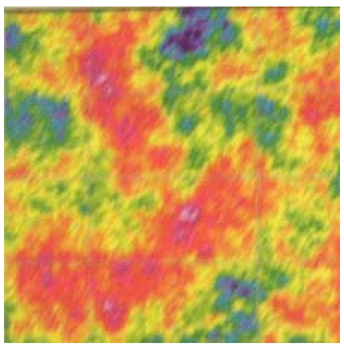




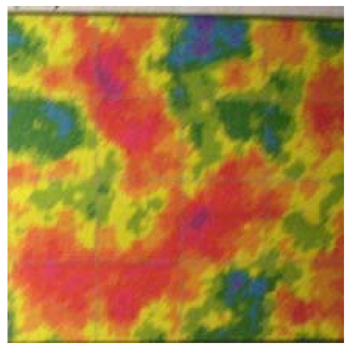
Nearest Neighbor Methods

- Bias-Variance decomposition: $$Err(x_0) = \sigma_\varepsilon^2 + \underbrace{\left(f(x_0) - \frac{1}{k}\sum_{n=1}^{k} f(x_n) \right)^2}_{\text{Bias}^2} + \underbrace{\frac{\sigma_\varepsilon^2}{k}}_{\text{Variance}}$$

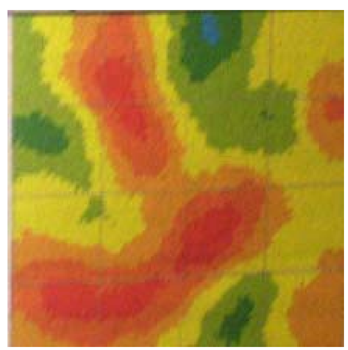
original



k=3



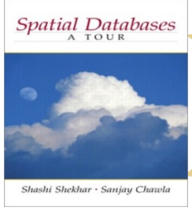
k=1
(hi var)



k=30
(hi bias)

“Optimal Weighting”: Kriging

- Characteristics of “optimality”:
  - Prediction is made as a **linear** combination of known data values (a **weighted average**)
    - Points closer to the point to be predicted have larger weight
  - Prediction is **unbiased** and **exact at known points**
  - Error estimate is based only on the sample configuration, not the data values
  - **Prediction error should be as small as possible**
- Why “optimal” and not optimal?
  - *“optimal” with respect to the chosen model!*



Overview of Kriging

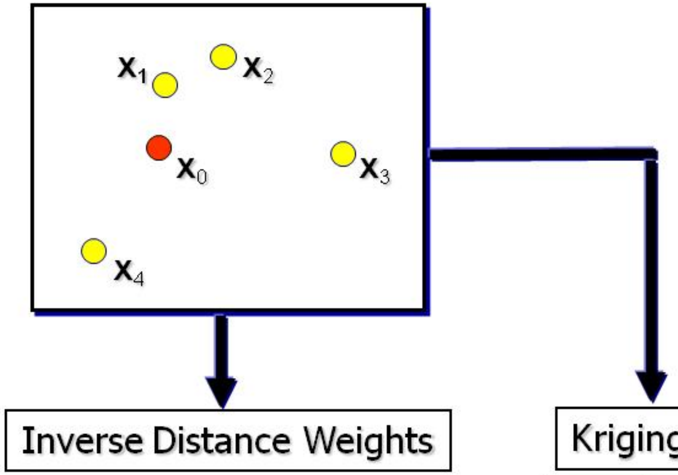
- 1. **Sample**, preferably at different resolutions
- 2. **Calculate** the **experimental variogram**
- 3. **Model** the variogram with one or more authorized functions
- 4. **Apply** the kriging system, with the variogram model of spatial dependence, at each point to be predicted
  - Predictions are often at each point on a **regular grid** (e.g. a raster map)
- 5. Calculate the **error** of each prediction; this is based only on the **sample point locations**, *not* their data values.

Ordinary Kriging (OK)

- In OK, we model the value of variable z at location s_i as the sum of a **regional mean** m and a **spatially-correlated random component** $e(s_i)$:

- $Z(s_i) = m + e(s_i)$

- The regional mean m is estimated from the sample, but not as the simple average, because there is spatial dependence
  - It is **implicit** in the OK system



$$Z^*_{\mathbf{x}_0} = \sum_{i=1}^{n} \lambda_i Z(\mathbf{x}_i)$$
$$\lambda_i = \frac{1}{d^2}$$

$$Z^*_{\mathbf{x}_0} = \sum_{i=1}^{n} \lambda_i Z(\mathbf{x}_i)$$
$$\lambda_i = ?$$

Ordinary Kriging: Solution

$$\begin{vmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \alpha \end{vmatrix} = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} & 1 \\ C_{21} & C_{22} & \cdots & C_{2n} & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix}^{-1} \begin{vmatrix} C_{10} \\ C_{20} \\ \vdots \\ C_{n0} \\ 1 \end{vmatrix}$$

- Covariance matrix elements

$$C_{ij} = C(\mathbf{0}) - \gamma(\mathbf{h}) = C_0 + C_1 - \gamma(\mathbf{h})$$

- Substituting the values we find the weights

- Kriging estimator: $Z^*_{\mathbf{x}_0} = \sum_{i=1}^{n} \lambda_i Z(\mathbf{x}_i)$

- Variance $\sigma^2_{ko} = (C_0 + C_1) - \lambda^T k$

Kriging usage

- Supported by many GIS
  - http://faculty.washington.edu/mlog/teaching/geostats/labs/ArcWizzard/wizzard_demo.shtml
- But be aware of polemics between classic statistics vs. geostatistics
  - spatial dependence may be assumed or be verified?
  - Kriging in scandal: Spatial dependence between borehole grades or blasthole grades was assumed at Bre-X's Busang property
  - More details:
    - http://en.wikipedia.org/wiki/Kriging#Controversy_in_climate_change.2C_mineral_exploration.2C_and_mining
    - http://www.geostatscam.com/


Bre-X fallout puts heat on regulators

Analysts warn Canadian junior miners will all pay penalty for Busang disaster. Far-reaching investigations likely.