Quantifying Form: Real-life Examples

- *Objectives:*

Discuss generic autocorrelation – 2-D analysis of form

Showcase some examples of Moran's I applications

- *Learning Outcomes:*

Review and criticize available autocorrelation metrics

Discuss how to display / interpret autocorrelation results

Criticize published papers (e.g., identify common mistakes)

Autocorrelation Coefficient

$$\frac{\sum w_{ij}(x_i - \bar{x})(x_j - \bar{x}) \;/\; \sum w_{ij}}{\sqrt{\sum (x_i - \bar{x})^2}\sqrt{\sum (x_i - \bar{x})^2}}$$

What are these weights? W

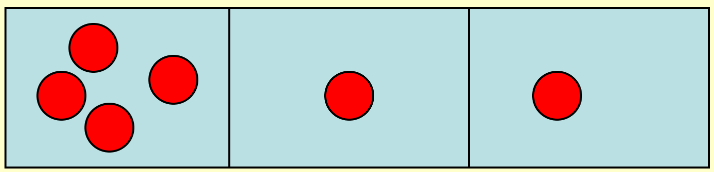
Weighted Average

Weighted Average = Σ (n * w) / Σ w

w = weight ***n = value***

What is the average fish abundance ? (fish / hr)

2 hr 1 hr 1 hr



Weighted mean: 1.5

sample	fish	effort (hr)	abund. (fish / hr)	(w * n) (effort * abund.)
1	4	2	2	4
2	1	1	1	1
3	1	1	1	1

Σ (w) = 4

Σ (n * w) = 6

Autocorrelation Coefficient

$$\frac{\sum \underbrace{w_{ij}}_{} \overbrace{(x_i - \bar{x})}^{Zi}\overbrace{(x_j - \bar{x})}^{Zj} \;/\; \sum w_{ij}}{\sqrt{\sum (x_i - \bar{x})^2}\sqrt{\sum (x_i - \bar{x})^2}}$$

Let's try a change in notation – to simplify the equation:

Zi is the deviation from the mean for value at location i
(i.e., Zi = xi – x for variable x)

Zj is the deviation from the mean for value at location j
(i.e., Zj = xj – x for variable x)

Autocorrelation Metrics – Moran's I

- Cross-product statistic that used to describe autocorrelation
- Compares deviations at pairs of locations

$$I_{(d)} = \frac{n \sum_i \sum_j w_{ij} Z_i Z_j}{W_{ij} \sum_i Z_i^{\,2}}$$

Numerator is the covariance (cross-product)

Denominator is the variance

Where:

n is the number of pairs

$\mathbf{Z_i}$ is the deviation from the mean for value at location i (i.e., $Z_i = x_i - \bar{x}$ for variable x)

$\mathbf{Z_j}$ is the deviation from the mean for value at location j (i.e., $Z_j = x_j - \bar{x}$ for variable x)

$\mathbf{w_{ij}}$ is the weight at given distance **d**

(e.g. wij = 1, if *j* is in distance class d from point *i*, otherwise = 0)

$\mathbf{W_{ij}}$ is the sum of all weights (for the number of pairs in distance class)

Autocorrelation Metrics – Moran's I

Moran's I Values range from: [-1, +1]

$$\boxed{(y_i - \overline{y})\ (y_j - \overline{y})}$$

Value = +1 **Positive correlation**

Value = -1 **Negative correlation**

Value = -1 / (n-1) **No correlation**

Autocorrelation Metrics – Geary's C

- Squared differences used to assess autocorrelation
- Considers differences between pairs of observations

$$C_{(d)} = \frac{[(n-1)[\sum_i \sum_j w_{ij}(y_i - y_j)^2]}{2W_{ij} \sum_i Z_i^{\,2}}$$

Numerator is the squared difference for each pair

Denominator is the variance

Where:

n is the number of pairs

Z_i is the deviation from the mean for value at location i (i.e., $Z_i = x_i - \bar{x}$ for variable x)

w_{ij} is the weight at given distance **d**

(e.g. wij = 1, if *j* is in distance class d from point *i*, otherwise = 0)

W_{ij} is the sum of all weights (for the number of pairs in distance class)

Autocorrelation Metrics – Geary's C

Geary's I Values range from: [0, 3]

$$\boxed{(y_i - y_j)^2}$$

Value = 0 **Positive correlation**

Value = 1 **No autocorrelation**

Value > 1 **Negative correlation**

So… Do We Use Moran's I or Geary's C ?

Ecological Pattern	Moran's I	Geary's C
Positive correlation	**+1**	**0**
Negative correlation	**-1**	**+3**
No autocorrelation	**-1 / (n-1)**	**+1**

- **Good aspects of Geary's C:**
  More sensitive to extreme values & clustering than Moran's I

- **Bad aspects of Geary's C:** **Confusing !!!**

  "no correlation" > "positive correlation"
  "negative correlation" > "positive correlation"

Moran's I Example – Reid et al. 2004

The relationship between autocorrelation and sampling distance (lags) used to identify the characteristic scales of distribution of krill predators and fisheries

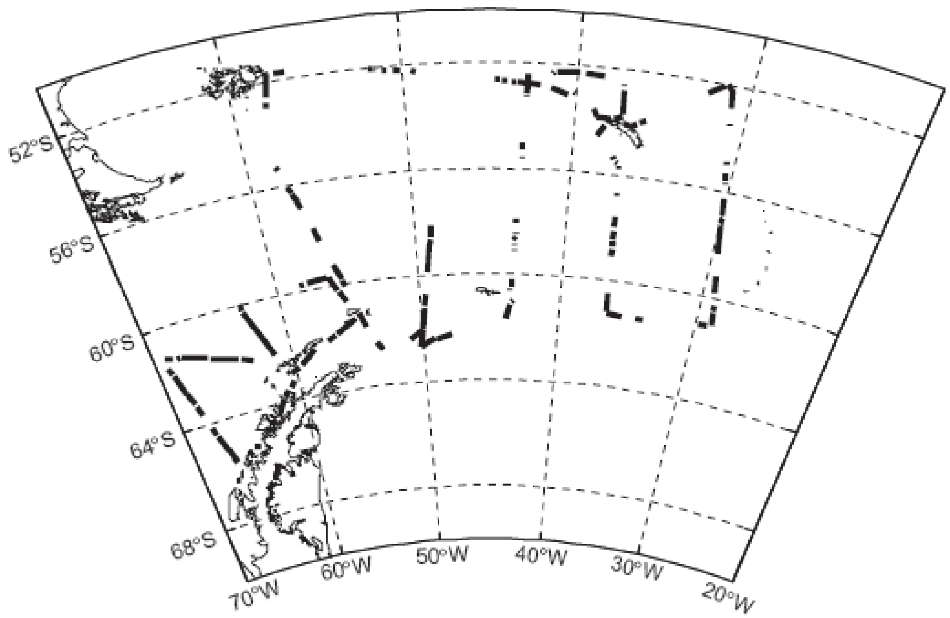


Fig. 1. Location of predator observations from the R.R.S. *James Clark Ross*.

How it Works:

- Pick Spatial Resolution

- Select Range of Lags

Species	n	Width (km)	Lo (lo) (km)
Krill	898	24	213 (213)
Fulmar	275	14	37 (35)
Fur seal	279	13	131 (72)

Moran's I Example – Reid et al. 2004

Two Different Auto-Correlation Definitions:

Species	Width (km)	Lo (km)	lo (km)
Krill	24	213	213

L0 (L-zero):
Lag where correlogram first crosses the x-axis

I0 (I-zero):
First lag where correlogram not significantly different from 0 *(open / filled circles)*

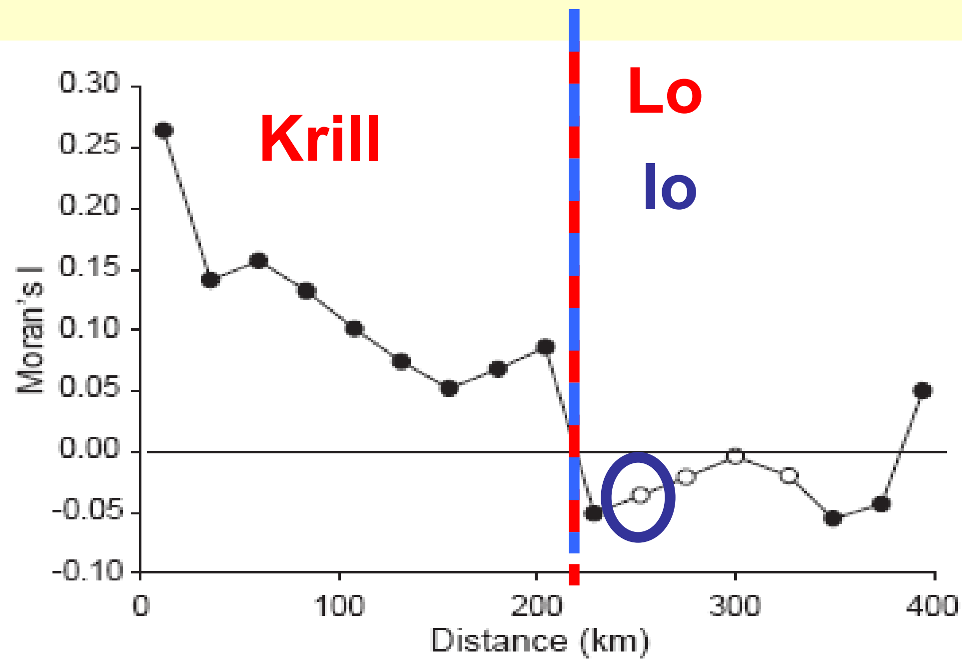


Fig. 5. Spatial autocorrelation function for krill biomass. Open circles indicate Moran's *I* not significantly different to zero.

Moran’s I Example – Reid et al. 2004

Species	Width (km)	Lo (km)	lo (km)
Fulmar	14	37	35
Fur seal	13	131	72

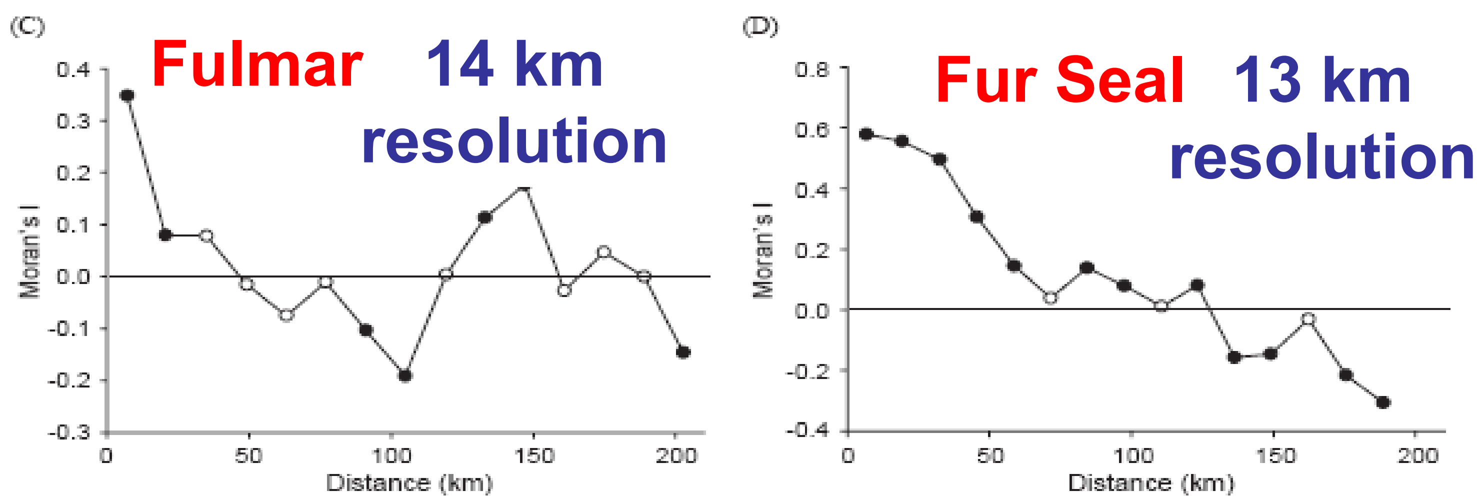


Fig. 3. Spatial autocorrelation function for (A) (Antarctic) prion; (B) chinstrap penguin; (C); Antarctic fulmar and (D) (Antarctic) fur seal. Open circles indicate Moran's *I* not significantly different to zero.

Sturges Rule

Sturges H (1926). The choice of a class-interval. *J. Amer. Statist. Assoc*., **21**, 65-66.

A histogram is constructed by dividing up the range of possible values in a data set into nonoverlapping intervals or classes called *bins* and then counting the number of observations that fall into each bin.

A good rule of thumb for choosing the number of bins (and thereby determining the bin width) is *Sturges' rule*.

This rule gives K number of bins, where:

K = 1 + (3.322*log10(n)) ***(where n is the number of samples)***

Sturges Rule

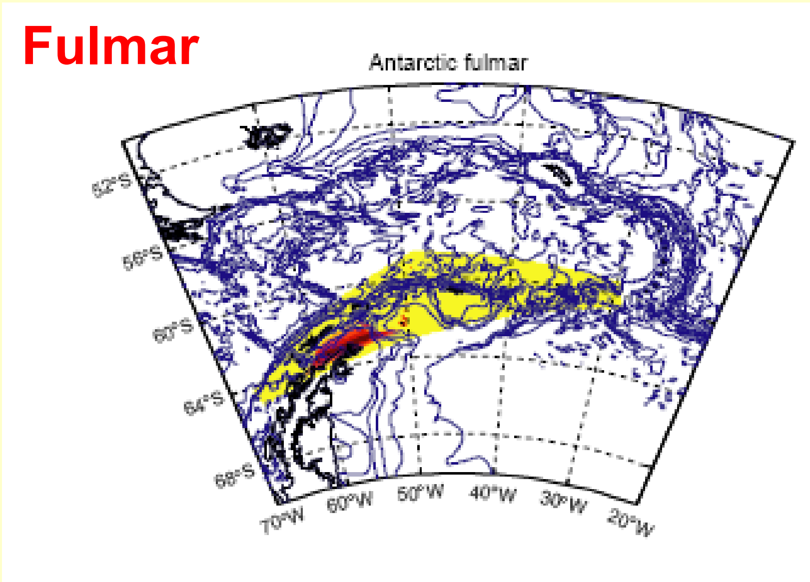
This rule gives K number of bins, where:

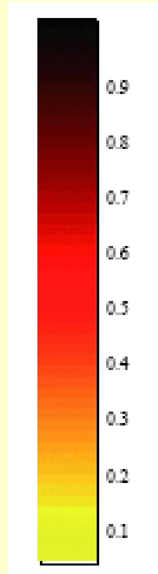
K = 1 + (3.322*log10(n)) ***(where n is the number of samples)***

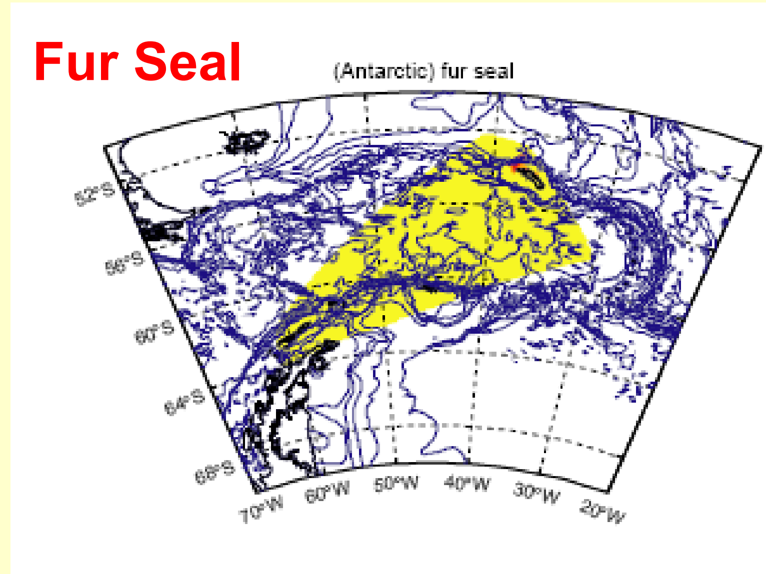
(n * (n-1)) / 2

species	n	K	spatial range	spatial scale
krill	367653	19.5	400	21
fulmar	37675	16.2	200	12.3
fur seal	38781	16.2	200	12.3

Moran's I Example – Reid et al. 2004







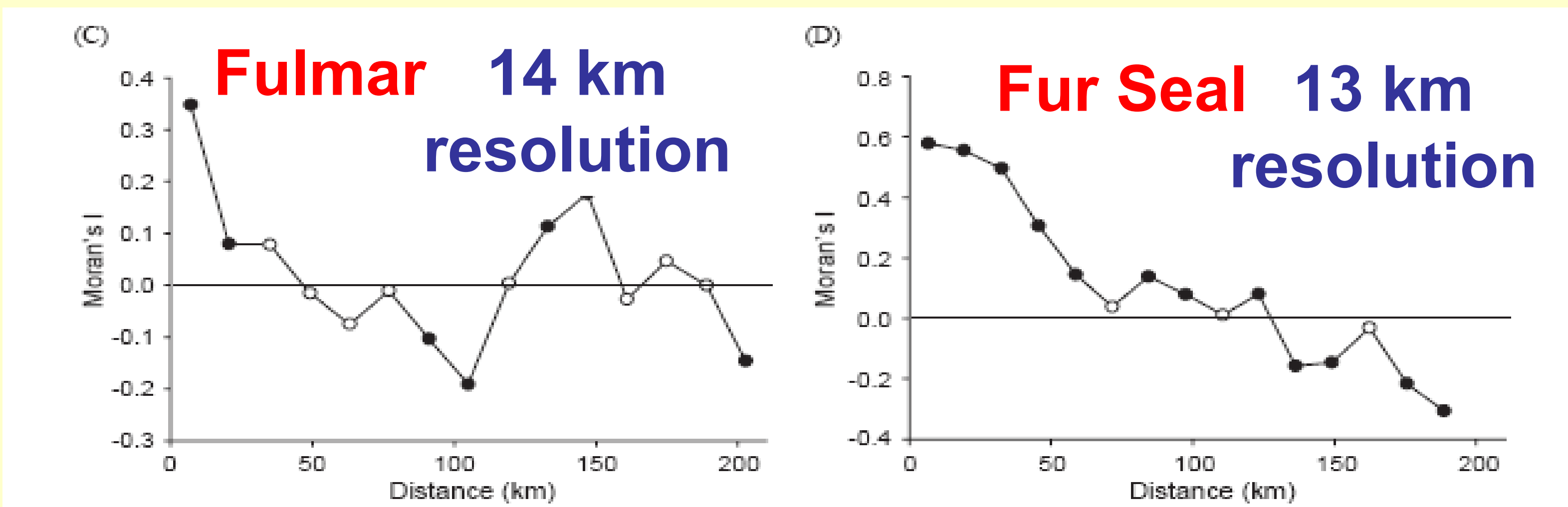


Fig. 3. Spatial autocorrelation function for (A) (Antarctic) prion; (B) chinstrap penguin; (C); Antarctic fulmar and (D) (Antarctic) fur seal. Open circles indicate Moran's I not significantly different to zero.

Moran's I Example – Rosenberg et al. 1999

- Atlas of 40 cancer mortalities in 355 areas in Western Europe

(A) Global Autocorrelation

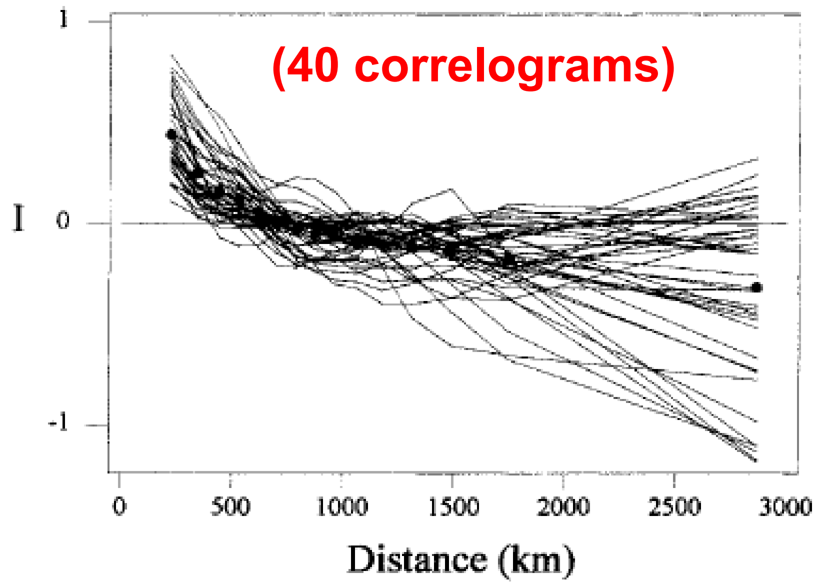


Figure 1. One-dimensional correlograms of Moran's I for mortalities of 40 cancers. Filled circles represent the average correlogram.

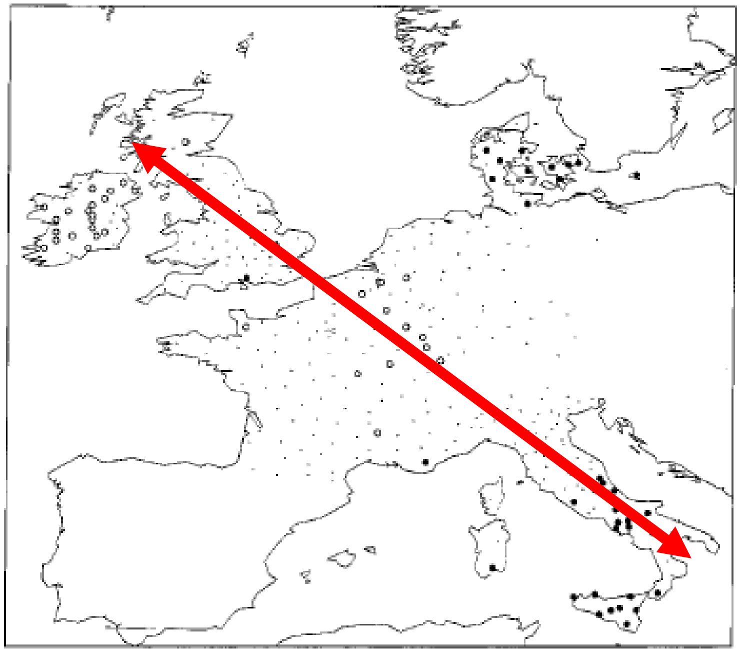


Figure 4. Map of the top (positive local SA; filled circles) and bottom (negative local SA; hollow circles) of the ranked conditional permutational probabilities of I_i for ovarian cancer mortality. We aimed at 10% for both ends of the distribution, but were unable to achieve that exactly. There are more positive areas, because of tied values of I_i and fewer negative areas, because the number of local coefficients below their expected value was less than 10%. The remaining localities are shown as dots.

Moran's I Example – Rosenberg et al. 1999

(B) Synthesize Results

Are there similar patterns ?

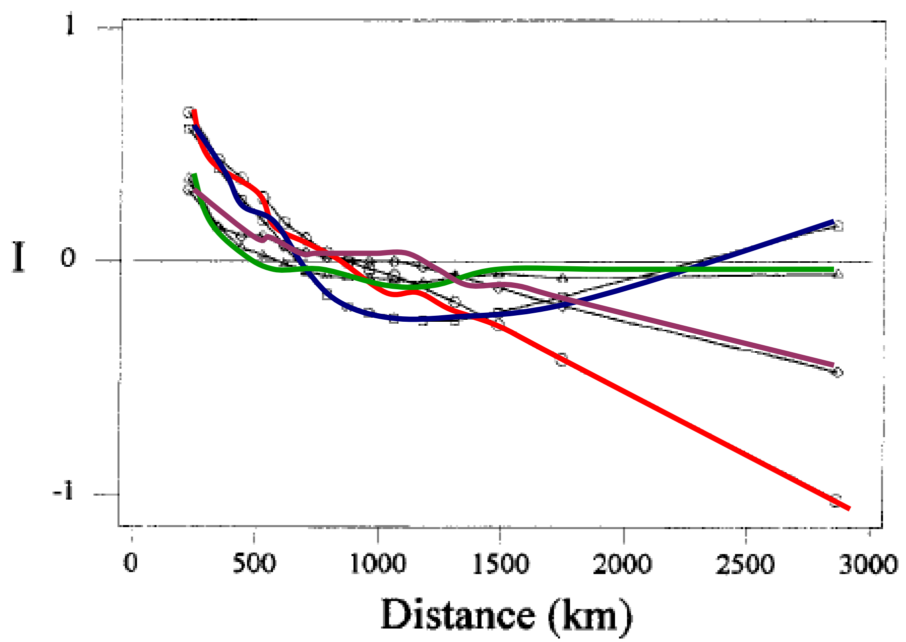


Figure 2. Average one-dimensional correlograms of the 4 *k*-means clusters of the Moran's *I* correlograms. Circles: Group 1, strong cline; diamonds: Group 2, moderate cline; triangles: Group 3, patchy; squares: Group 4, imperfect bowl.

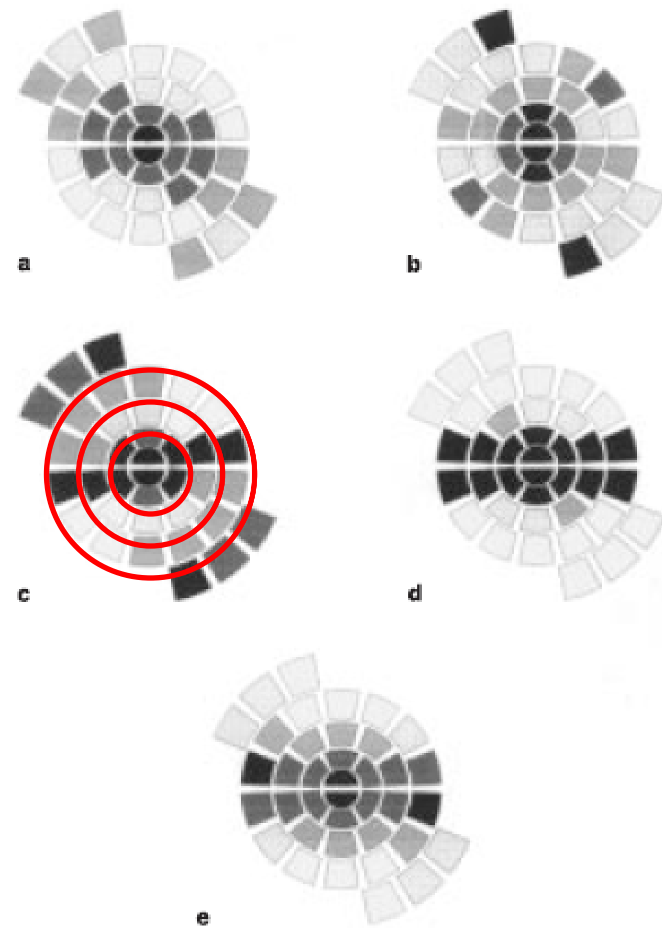


Figure 3. Average two-dimensional correlograms based on 5 clusters. Values of *I*: white: −2.44 to −0.32; pale gray: −0.29 to −0.13; gray: −0.12 to 0.02; dark gray: 0.03 to 0.18; black: 0.19 to 1.21. Upper limits of distance class annuli: 150, 600, 1350, 2400, 3750 km. a, Group 1; b, Group 2; c, Group 3; d, Group 4; e, Group 5.

Moran's I Example – Rosenberg et al. 1999

(C) Local Autocorrelation

Distinguishes "hot" / "cold" spots

$$G_{(d)} = \frac{\sum w_{ij}(d) x_i x_j}{\sum x_i x_j}$$

G ranges from 0 to infinite:

large if high values cluster
low if low values cluster

Numerator calculated "within" a distance bound (d)

Denominator, expressed relative to the entire study area

Where:
d = distance class
W_{ij} = weight matrix **(1 when |i-j| $\leq$ d, 0 when |i-j| $\geq$ d)**

Summary

Several metrics are available to calculate autocorrelation

They all have good / bad features. **Moran's I is Safe Pick**

Carefully consider resolution / extent: **use same scales**

Carefully consider the lags: **too coarse for interpretation**

Be aware of complementary perspectives: **global vs local**

Depending on your hypothesis, you can develop very complicated "connectedness" rules:

- 1-D and 2-D
- circular plots (heading)
- direction (upstream / downstream)