Spatial autocorrelation in ecological modelling

Workshop: Symposium for European Freshwater Sciences 2015

Avit Kumar Bhowmik

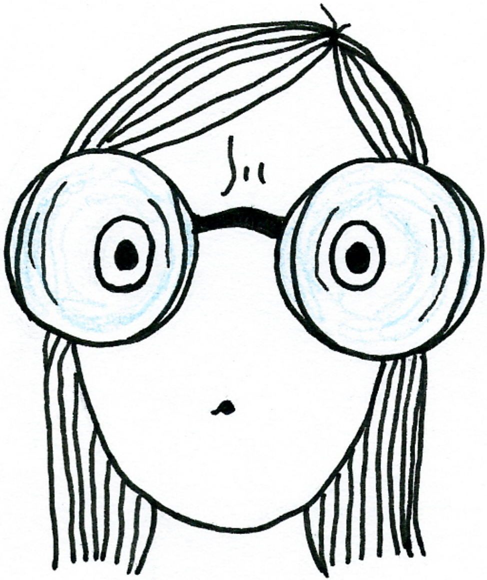
Institute for Environmental Sciences, University of Koblenz-Landau

bhowmik@uni-landau.de

July 4, 2015

We need spatially explicit models. Why?

Wear the GI glasses



World is spatially autocorrelated

World is spatially autocorrelated

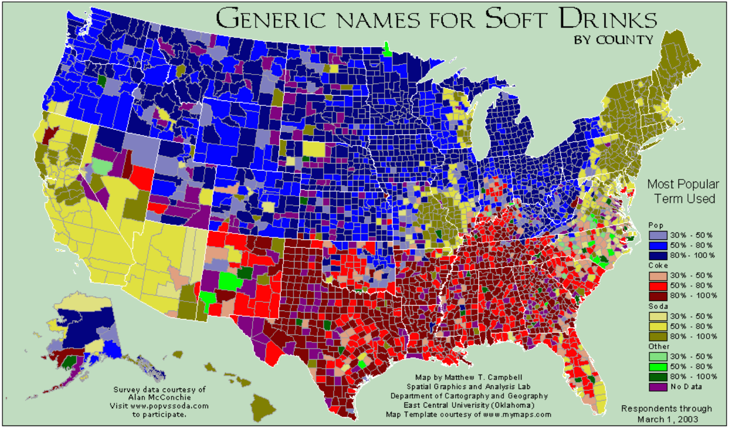
Tobler's first law of geography (Tobler, 1970. Eco.Geo)

"Everything is related to everything else, but near things are more related than distant things"

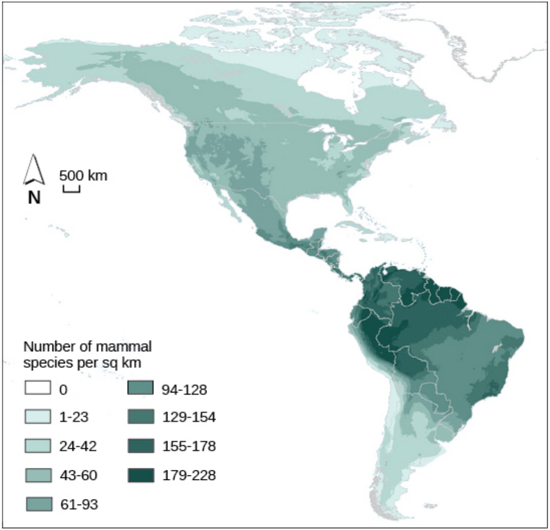
World is spatially autocorrelated

Tobler's first law of geography (Tobler, 1970. Eco.Geo)

"Everything is related to everything else, but near things are more related than distant things"



Autocorrelation is a “fact of life” for ecologists!



(NASA, CIESIN, Columbia University, 2010)

Autocorrelation is a "fact of life" for ecologists!

Definition (Legendre, 1993. Ecology)

"...the property of random variables taking values, at pairs of locations a certain distance apart, that are more similar (positive autocorrelation) or less similar (negative autocorrelation) than expected for randomly associated pairs of observations"

Autocorrelation is a "fact of life" for ecologists!

Definition (Legendre, 1993. Ecology)

"...the property of random variables taking values, at pairs of locations a certain distance apart, that are more similar (positive autocorrelation) or less similar (negative autocorrelation) than expected for randomly associated pairs of observations"

Fact (Fortin and Dale, 2005. Spatial Analysis)

"natural systems almost always have autocorrelation in the form of patchiness or gradients...over a wide range of spatial and temporal scales"

Two types of spatial autocorrelation

1. Endogenous
   - caused by biotic processes, e.g. dispersal
2. Exogenous
   - caused by functional dependence on spatially autocorrelated drivers, e.g. climate

Relevance for ecological models

- Spatial autocorrelation is a nuisance that complicates statistical hypothesis testing

Relevance for ecological models

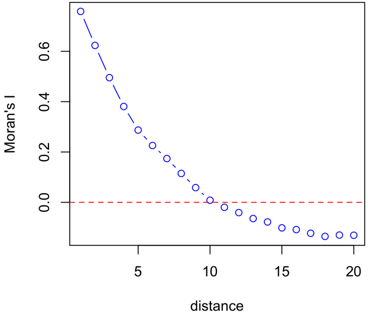
- Spatial autocorrelation is a nuisance that complicates statistical hypothesis testing
- Spatial autocorrelation is functionally important in many ecosystems, so we must revise our theories and models to incorporate spatial structure (Fortin and Dale, 2005. Spatial Analysis)

Quantification and visualization of spatial autocorrelation

- Morans I , Gearys c correlation coefficients over multiple distances

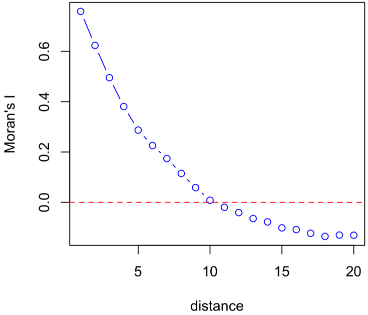
Quantification and visualization of spatial autocorrelation

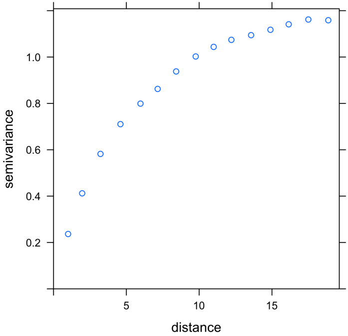
- Morans I , Gearys c correlation coefficients over multiple distances
- Correlogram plot distance on X-axis against correlation coefficient on Y-axis
    - Mantel correlogram for multivariate response
    - Observe distance decay!



Quantification and visualization of spatial autocorrelation

- Morans I , Gearys c correlation coefficients over multiple distances
- Correlogram plot distance on X-axis against correlation coefficient on Y-axis
    - Mantel correlogram for multivariate response
    - Observe distance decay!





- Semi-variogram or variogram
    - Inverse correlogram!

Introduce spatial autocorrelation structure to linear models

- Remember, linear models assume that data are one-dimensional, but spatial data are two-dimensional (Lat-long and XY-coordinates)

Introduce spatial autocorrelation structure to linear models

- Remember, linear models assume that data are one-dimensional, but spatial data are two-dimensional (Lat-long and XY-coordinates)
- Mind the distribution of your data

Introduce spatial autocorrelation structure to linear models

- Remember, linear models assume that data are one-dimensional, but spatial data are two-dimensional (Lat-long and XY-coordinates)
- Mind the distribution of your data
- Incorporate a variance-covariance matrix (C) that is affected by a proximity matrix (W) of neighbor weights

$$y = X\beta + \epsilon$$

$$y = X\beta + \rho W(y - X\beta) + \epsilon$$

$$C = \delta^2[(I - \rho W)^T (I - \rho W)]^{-1}$$

Introduce spatial autocorrelation structure to linear models

- Remember, linear models assume that data are one-dimensional, but spatial data are two-dimensional (Lat-long and XY-coordinates)
- Mind the distribution of your data
- Incorporate a variance-covariance matrix (C) that is affected by a proximity matrix (W) of neighbor weights

$$y = X\beta + \epsilon$$

$$y = X\beta + \rho W(y - X\beta) + \epsilon$$

$$C = \delta^2[(I - \rho W)^T (I - \rho W)]^{-1}$$

- I will show you two models:
  Generalized Least Sqaures (GLS) and Generalized Linear Mixed Models (GLMM)

Generalized Least Squares (GLS)

- Assumes normally distributed errors
  but errors are allowed to be correlated and/or have unequal variances

Generalized Least Squares (GLS)

- Assumes normally distributed errors
  but errors are allowed to be correlated and/or have unequal variances

- Incorporate a variance-covariance matrix (C) in a generalized linear model form

$$y = X\beta + \epsilon$$

$$\epsilon = N(0, C)$$

- Fits a parametric correlation function (e.g. exponential, Gaussian, spherical) directly to the variance-covariance matrix

Generalized Least Squares (GLS)

- Assumes normally distributed errors
  but errors are allowed to be correlated and/or have unequal variances
- Incorporate a variance-covariance matrix (C) in a generalized linear model form

$$y = X\beta + \epsilon$$

$$\epsilon = N(0, C)$$

- Fits a parametric correlation function (e.g. exponential, Gaussian, spherical) directly to the variance-covariance matrix

R package "nlme"

```
library(nlme)
?gls
?corClasses
```

Generalized Linear Mixed Models (GLMM)

- Assumes non-normally distributed errors
  and within-group errors may be spatially autocorrelated

Generalized Linear Mixed Models (GLMM)

- Assumes non-normally distributed errors
  and within-group errors may be spatially autocorrelated
- Linear predictor is transformed by a link function
- Linear predictor may contain random effects

$$E[Y_{i,j}|\zeta_i] = g_{-1}(\eta_{i,j})$$

$$\eta_{i,j} = x_{i,j}\beta + z_{i,j}\zeta_i$$

Generalized Linear Mixed Models (GLMM)

- Assumes non-normally distributed errors
  and within-group errors may be spatially autocorrelated
- Linear predictor is transformed by a link function
- Linear predictor may contain random effects

$$E[Y_{i,j}|\zeta_i] = g_{-1}(\eta_{i,j})$$

$$\eta_{i,j} = x_{i,j}\beta + z_{i,j}\zeta_i$$

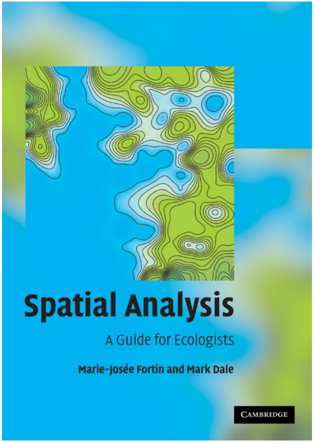
R package “MASS” and “lme4”

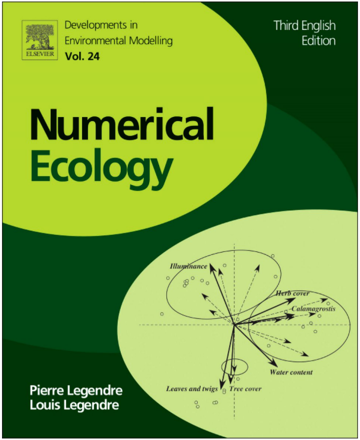
```
library(MASS)
?glmmPQL
library(lme4)
?glmer()
```

For more and on temporally autocorrelated data

Dormann et al. (2007)
Methods to account for spatial autocorrelation in the analysis of species distributional data: a review
Ecography 30, 609 – 628.

Dormann et al. (2007)
Methods to account for spatial autocorrelation in the analysis of species distributional data: a review
Ecography 30, 609 – 628.





Snouters!

http://encyclopedia.thefreedictionary.com/Snouter

Snouters!

- We will use Snouter data from Dormann et al., 2007. Ecogra provided in the GItHub repo

Snouters!

- We will use Snouter data from Dormann et al., 2007. Ecogra provided in the GItHub repo
- Questions
  - Do Snouter abundance and presence exhibit spatial autocorrelation?
  - What are the explained variances in Snouter abundance and presence by precipitation and distance to jungle?

Let's do it together!