Spatial Modeling in Ecology

Marie-Josée Fortin

Ecology and Evolutionary Biology
University of Toronto

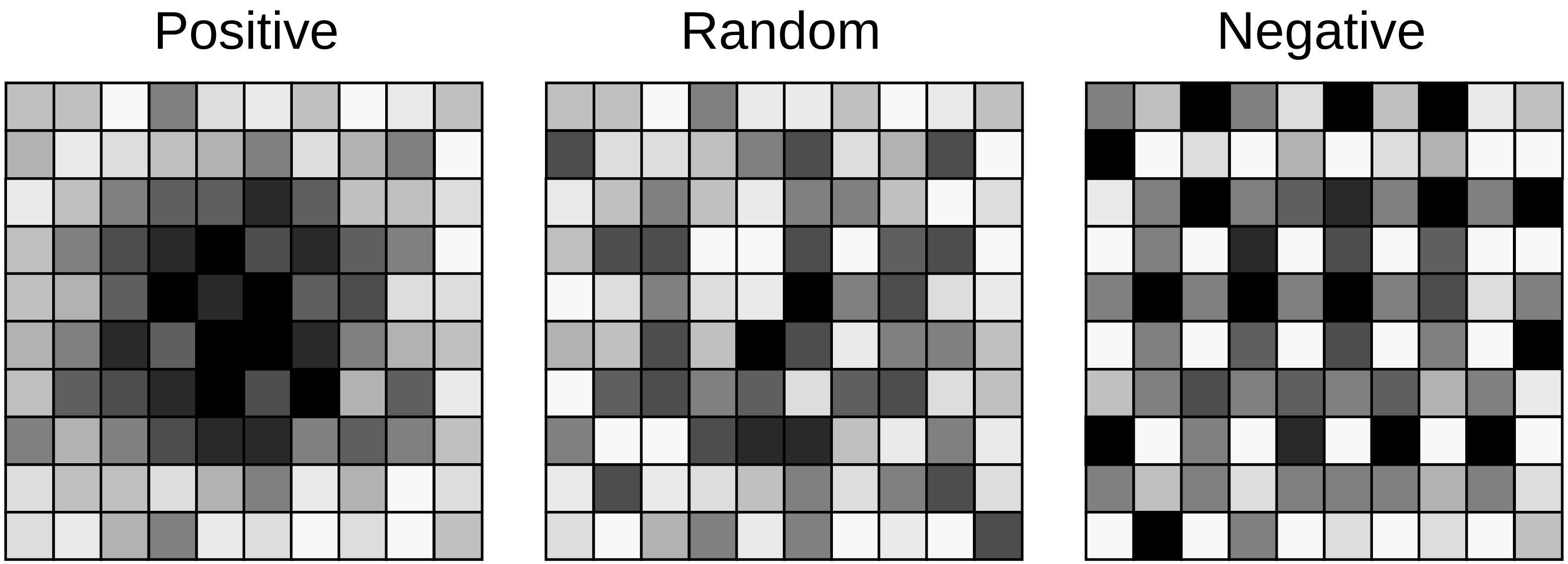
Dependencies in Data due to Space

Tobler's First Law of Geography:
"Everything is related to everything else, but near things are more related than distant things."

Dependencies in Data due to Space

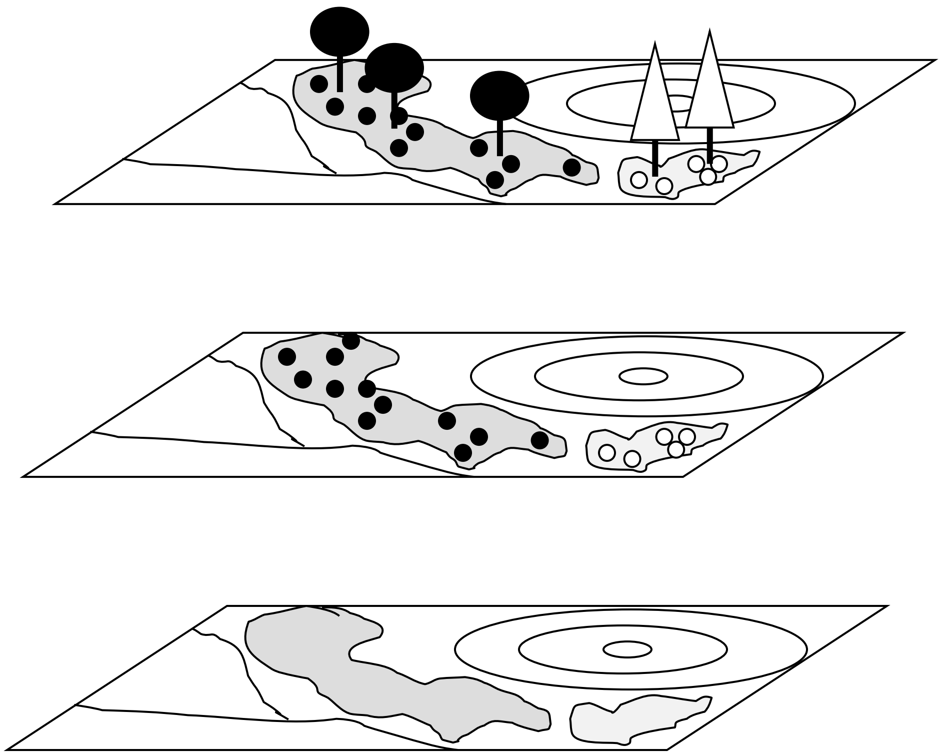
Spatial Autocorrelation

Correlation of the values of a variable according to distance



Fortin & Dale 2009

Spatial Structure = Autocorrelation + Dependence



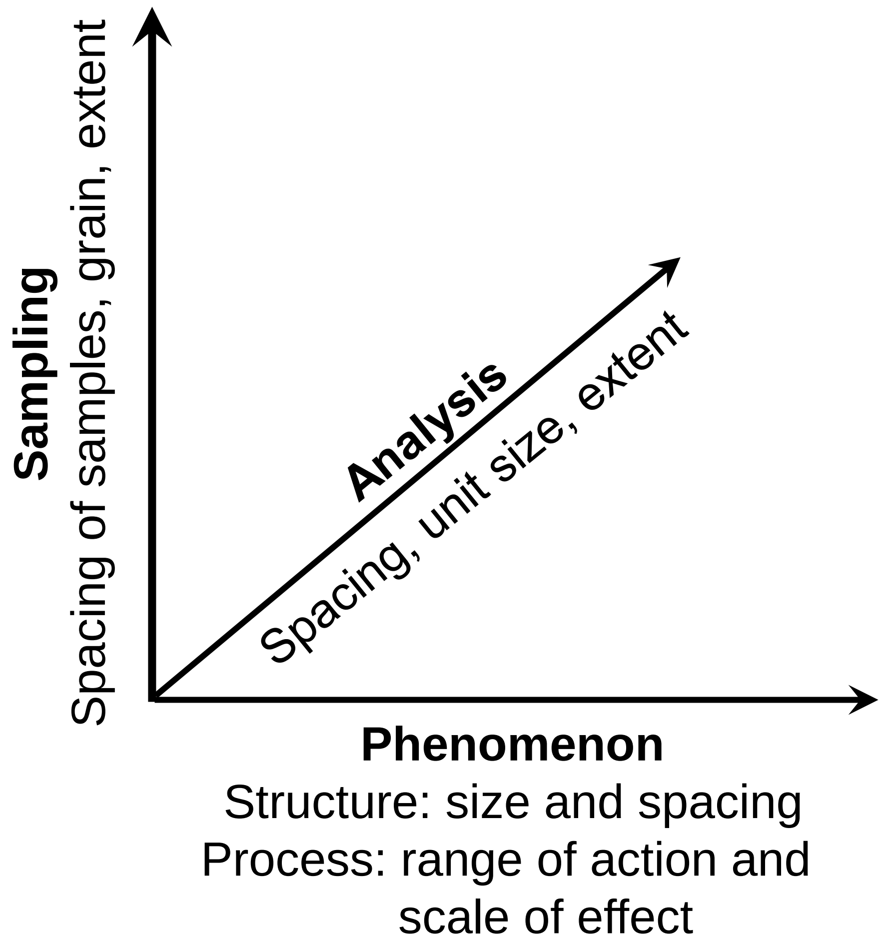
Spatially Distributed Seeds Due to Dispersal Processes from the Trees and the Spatially Distributed Environmental Factors

Spatially Distributed Seeds Due to the Spatially Distributed Environmental Factors

Spatially Distributed Environmental Factors

Fortin & Dale 2009

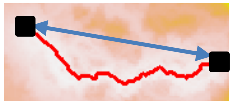
Other Dependencies: Processes and Analysis



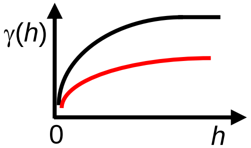
Fletcher & Fortin 2019

Spatial Aspects

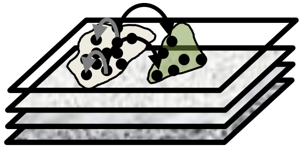
x-y Coordinates
Euclidean Distances
Least-cost Distances



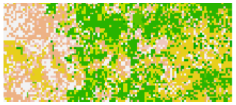
Spatial Autocorrelation
Spatial Dependence



Spatial Relationship

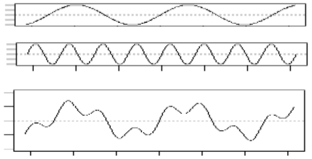


Spatial Legacy
Spatial Contingency

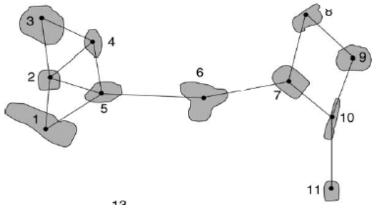


Spatial Perception

Multiscale Analysis



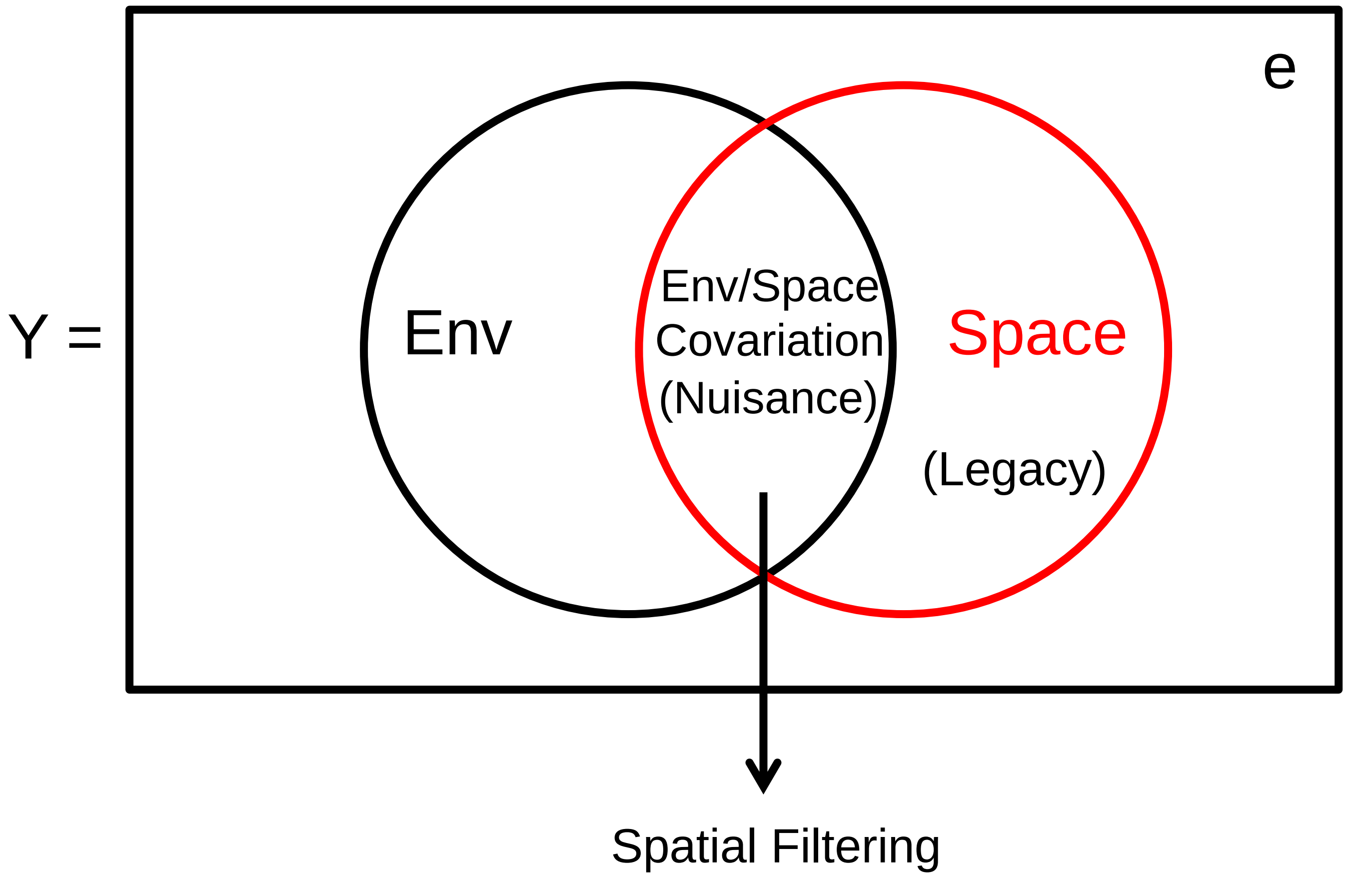
Metapopulation
Metacommunity
Metaecosystem
Metanetwork



Fortin et al 2012

Spatial Dependence + Spatial Autocorrelation

Environmental and Spatial Contributions



Dale & Fortin 2014

Spatial Structure

SPATIAL NUISANCE (Space is your Enemy)
Space as a weight: Spatial autocorrelation as part of residual variation

G	D	C/W	E
x-y	Euclidean Distance	Connectivity	Environment

SPATIAL LEGACY (Space is your Friend)

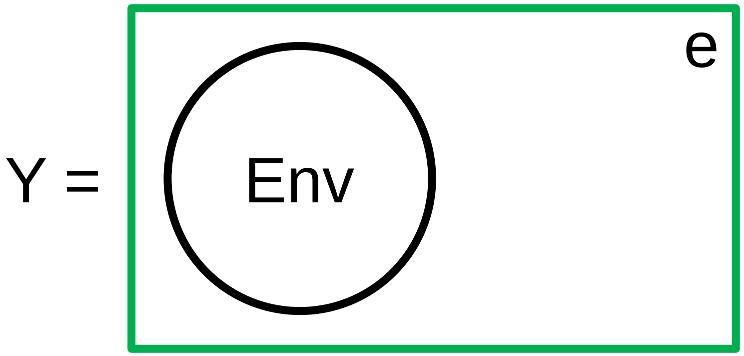
Space as a term: Spatial autocorrelation as a function

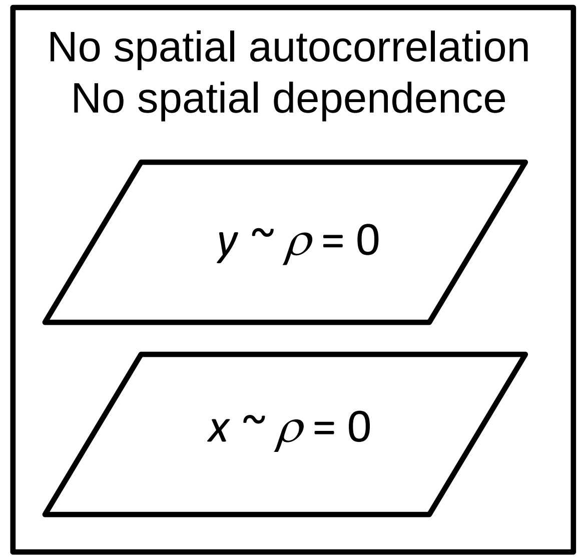
C/W = Spatial Matrix

Peres-Neto & Legendre 2010

Regression

Name	Model
Independent Errors (no spatial dependence)	$\mathbf{Y} = \mathbf{X}\beta + \varepsilon$

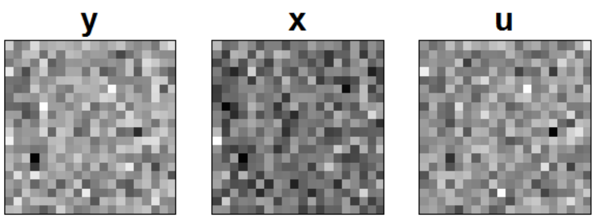




Dale & Fortin 2014

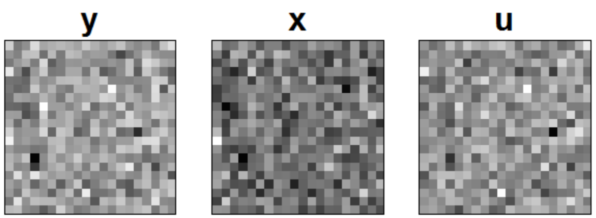
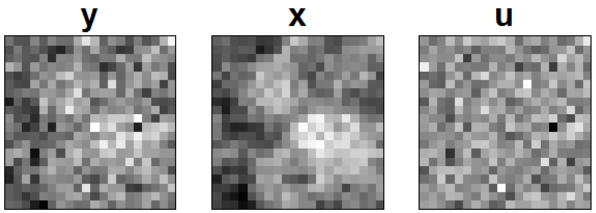
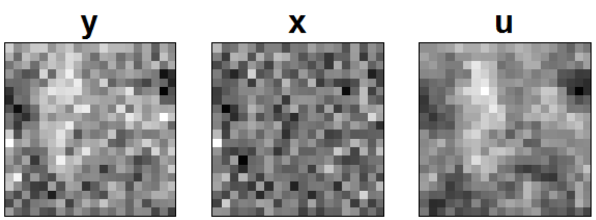
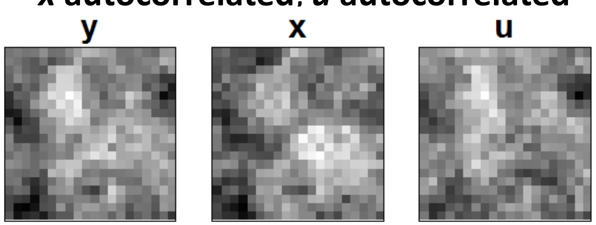
Linear Regression

$$y = 0.5x + 0.5u$$

x random, *u* random  	Predictor & Response: spatially independent Regression is ***correct***: • Moran's *I* of residuals = -0.002 • Slope estimate (*b* ± SE) = 0.500 ± 0.025 • True SE of slope: 0.025 • Type I error rate: 0.050

Wagner & Fortin 2016

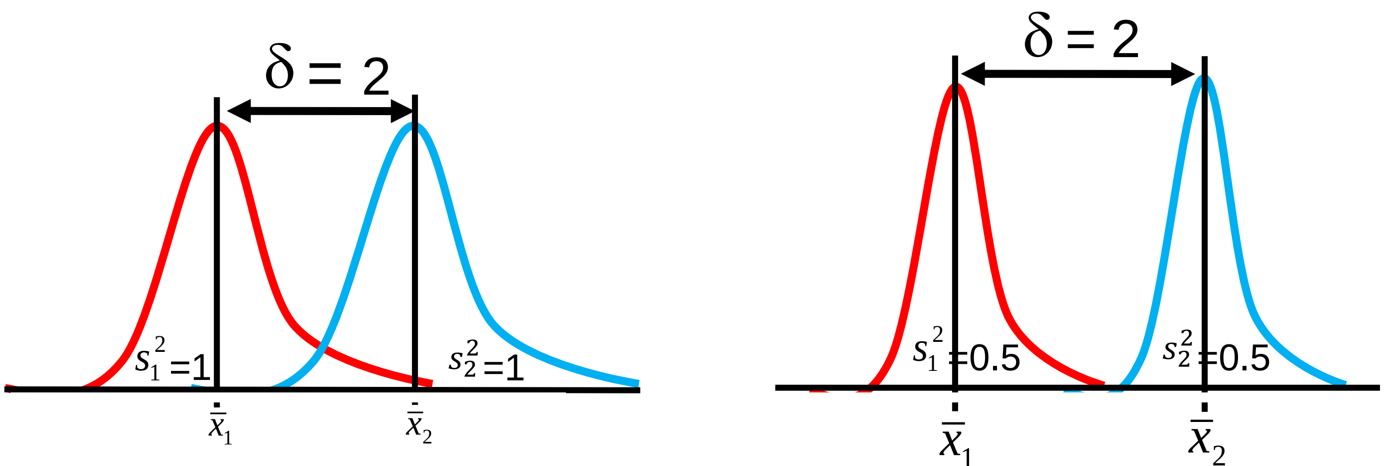
Linear Regression

x random, u̲ random  	Predictor random, Response random Regression is ***correct***: • Moran's *I* of residuals = -0.002 • Slope estimate (b ± SE) = 0.500 ± 0.025 • True SE of slope: 0.025 • Type I error rate: 0.050
***x* autocorrelated**, *u* random  	**Predictor autocorrelated**, Response random Regression is ***correct***: • Moran's *I* of residuals = -0.003 • Slope estimate (b ± SE) = 0.500 ± 0.025 • True SE of slope: 0.025 • Type I error rate: 0.048
x random, ***u* autocorrelated**  	Predictor random, **Response autocorrelated** Regression is ***correct***, but residuals spatially autocorrelated: • Moran's *I* of residuals = 0.640 • Slope estimate (b ± SE) = 0.501 ± 0.025 • True SE of slope: 0.025 • Type I error rate: 0.054
***x* autocorrelated**, ***u* autocorrelated**  	**Predictor autocorrelated**, **Response autocorrelated** Regression is ***incorrect***: • Moran's *I* of residuals = 0.629 • Slope estimate (b ± SE) = 0.501 ± 0.025 • True SE of slope: 0.071 • Type I error rate: 0.504

Wagner & Fortin 2016

Spatial Autocorrelation

Spatial Autocorrelation may:	*What does this mean?*
Inflate type I error rates	*Risk of false positives*
Bias parameter estimates	*Effect size may be over-estimated*
Inflate variability of parameter estimates	*Confidence intervals are calculated too small*



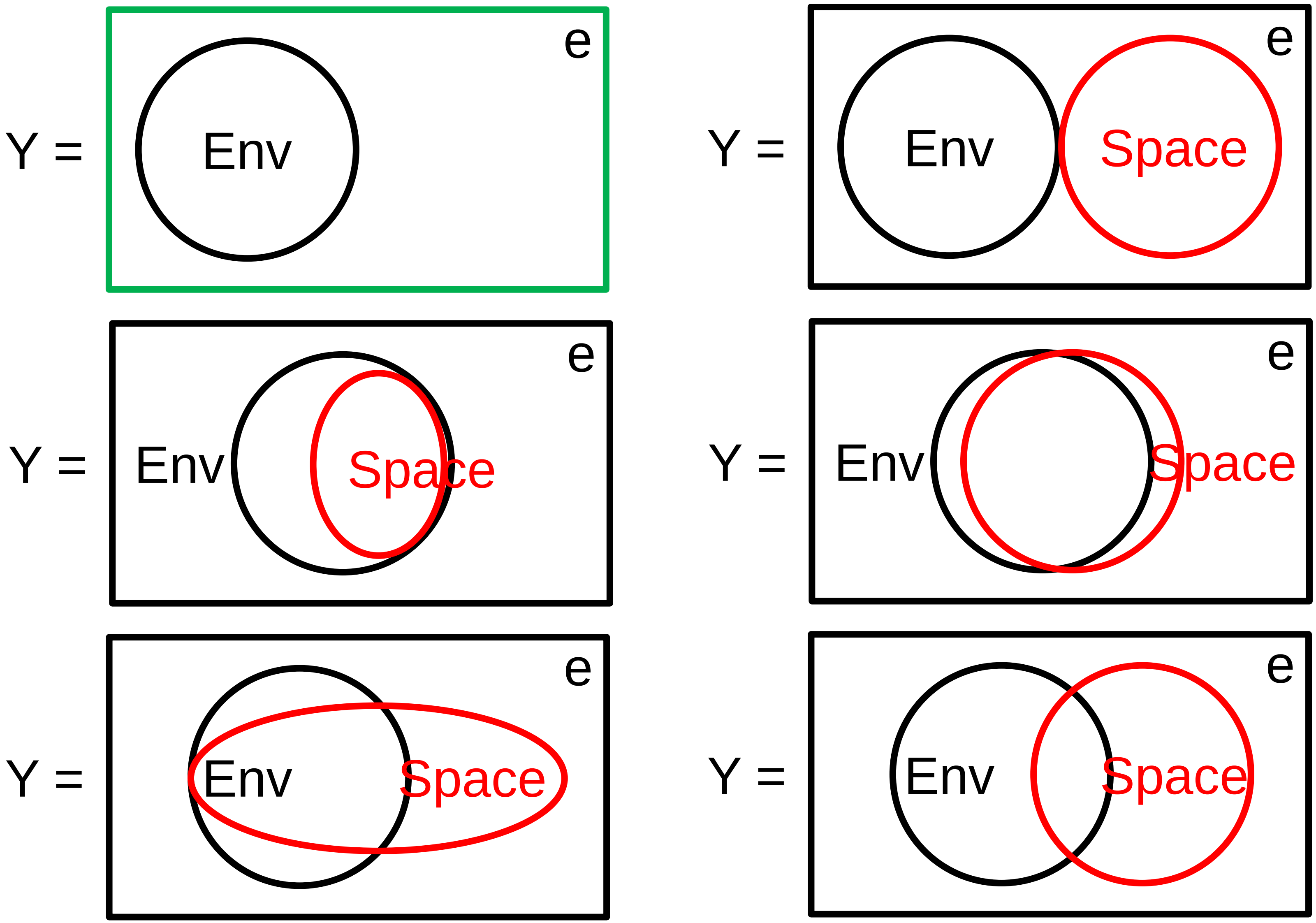
Without Autocorrelation

With Autocorrelation

Wagner & Fortin 2016

Spatial Dependence + Spatial Autocorrelation

Environmental and Spatial Contributions



Spatial Regression

Spatial Regression (GLMM, Error, Lag)

Spatial process creates spatial autocorrelation

→Need to account for spatial autocorrelation

Spatial Regression (Spatial Filtering)

Multiple processes at different spatial scales

→Need to partial out spatial variations at all scales

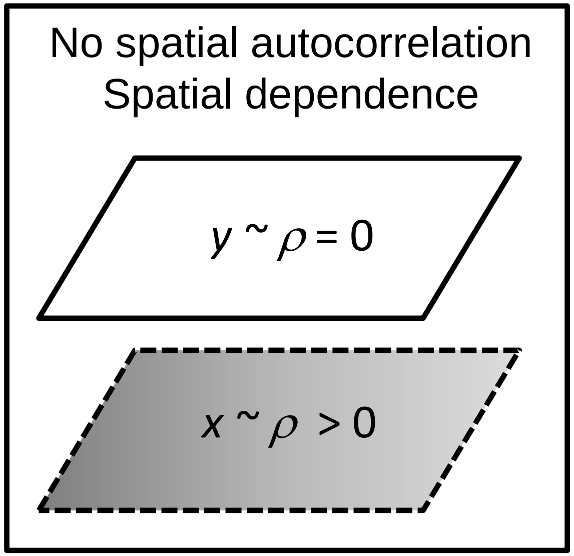
Spatial Regression: Trend Surface Analysis

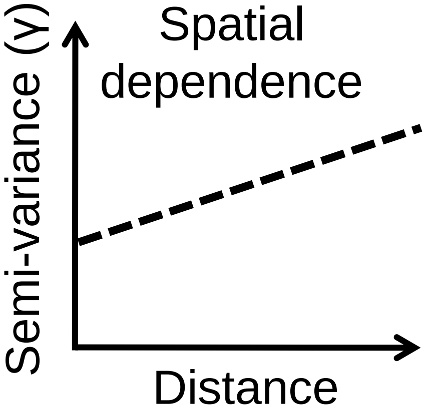
Linear regression (e.g., OLS):

$\{x_1(s), \dots x_m(s)\} \longrightarrow y(s) \longleftarrow \varepsilon$

Linear regression adding spatial coordinates as predictors using OLS:

$\{x_1(s), \dots x_m(s), \text{lat}, \text{long}\} \longrightarrow y(s) \longleftarrow \varepsilon$





Dale & Fortin 2014

Spatial Regression: Trend Surface Analysis

Regression including *x-y* coordinates
Polynomial regression

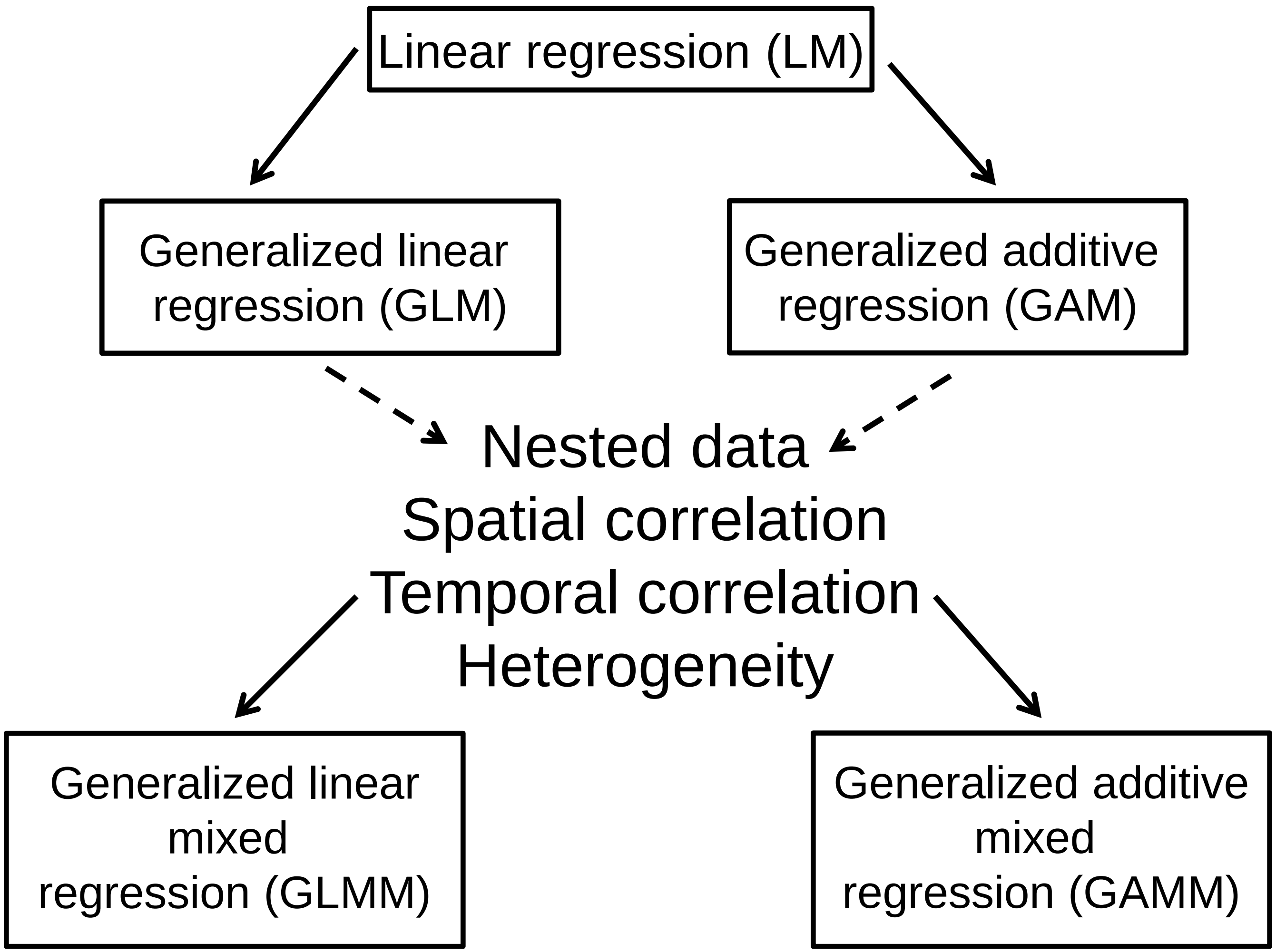
→ Only explanatory variables are powers and cross products of the *x-y* coordinates

→ Power of the polynomial
→Linear, quadratic, cubic, etc.

$$Y = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \ldots + \beta_m x_i^m + \varepsilon$$

Dale & Fortin 2014

Regression



Adapted from Zuur et al (2009) [Zuur & Ieno: January 9, 2023]

Linear Regression

Linear regression (e.g., OLS):

$\{x_1(s), \ldots x_m(s)\} \longrightarrow y(s) \longleftarrow \varepsilon$

Linear regression adding spatial coordinates as predictors using OLS:

$\{x_1(s), \ldots x_m(s), \text{lat}, \text{long}\} \longrightarrow y(s) \longleftarrow \varepsilon$

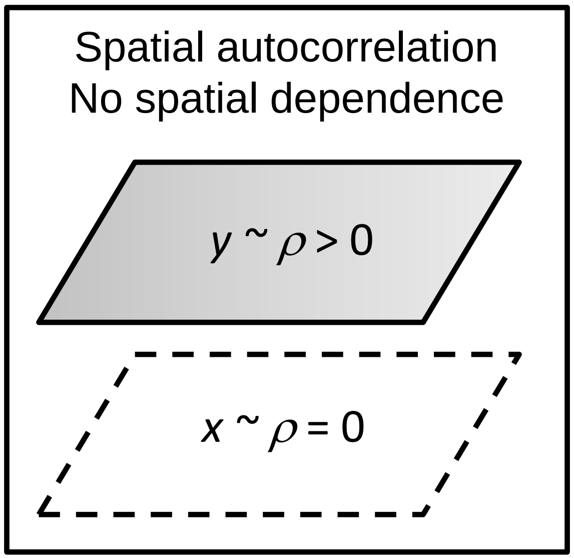
Linear regression adding spatial coordinates as random effects using Generalised Linear Mixed Model (GLMM):

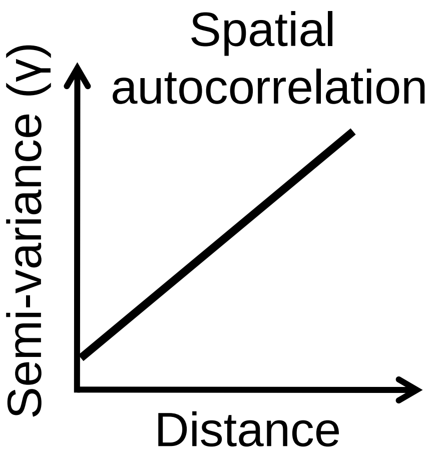
$\{x_1(s_i), \ldots x_m(s_i), z_i\} \longrightarrow y(s_i) \longleftarrow \varepsilon$

Dale & Fortin 2014

Spatial Regression

Name	Model
Independent Errors (no spatial dependence)	$Y = X\beta + \varepsilon$
Autoregressive (AR)	$Y = X\beta + \rho WY + \varepsilon$





Dale & Fortin 2014

Spatial Regression

Name	Model
Independent Errors (no spatial dependence)	$Y = X\beta + \varepsilon$
Autoregressive (AR)	$Y = X\beta + \rho WY + \varepsilon$
Simultaneous Autoregressive (SAR) Predict from error at neighboring locations	$Y = X\beta + \rho W(Y - X\beta) + \varepsilon$
Conditional Autoregressive (CAR) Predict from response at neighboring locations	$Y = X\beta + \rho C(Y - X\beta) + \varepsilon$

Spatial filtering regression: add lagged covariates using conditional autoregressive (CAR) or simultaneous autoregressive (SAR):

$\{x_1(s), \ldots x_m(s), x_1(s\pm h), \ldots x_m(s\pm h)\} \longrightarrow y(s) \longleftarrow \varepsilon$

where $(s\pm h)$ is the neibourhood of s

Dale & Fortin 2014

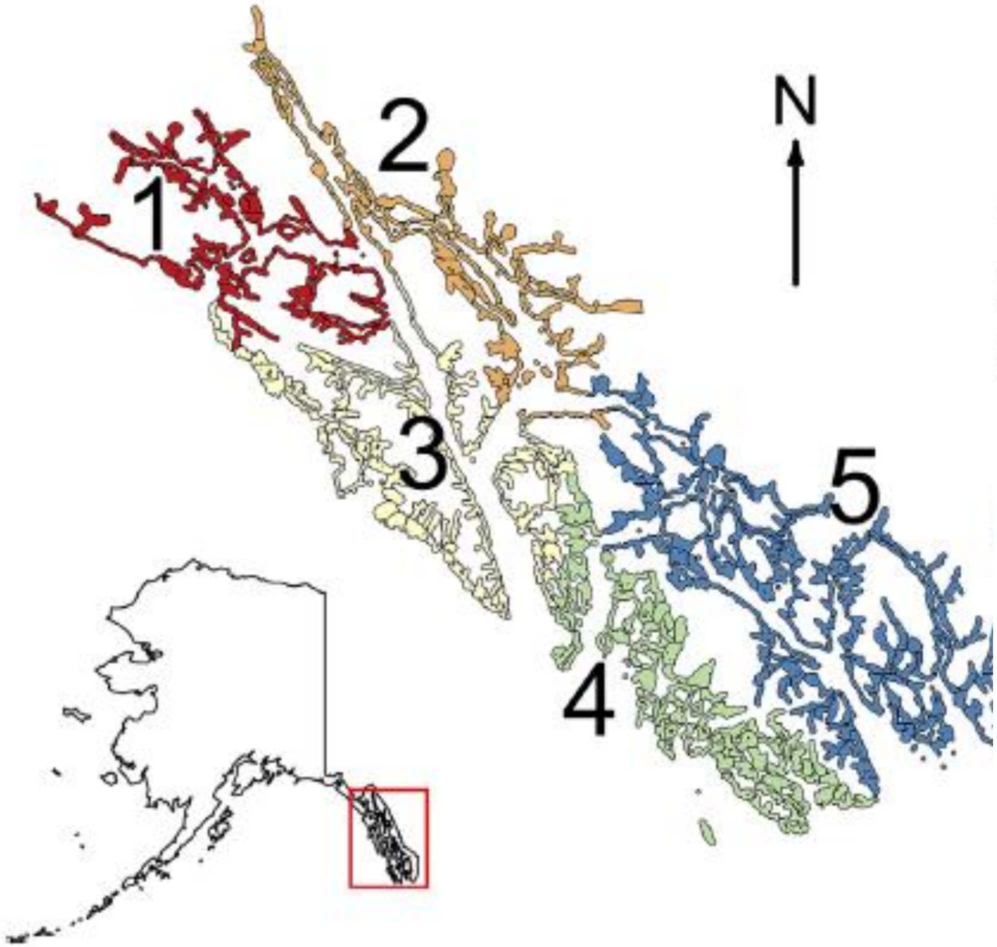
Common Objectives of Spatial Autoregressive Models

Model Comparison & Selection	CAR and SAR models are often part of a spatial (generalized) linear model. Prior to inference, one may compare models, and then choose one.
Regression	Estimate the spatial regression coefficients that quantify how an explanatory variable affects the response variable.
Auto-correlation	Estimate the strength of autocorrelation (or spatial connectivity) that quantifies how similarly sites change in the residual errors, after accounting for regression effects.
Connectivity	Estimate covariate effects on connectivity (neighborhood) structure.
Prediction	Goal of geostatistics (but rarely used in CAR and SAR). If sites have missing data, prediction is possible.
Smoothing	Create values at spatial sites that smooth over observed data by using values from nearby locations to provide better estimates.

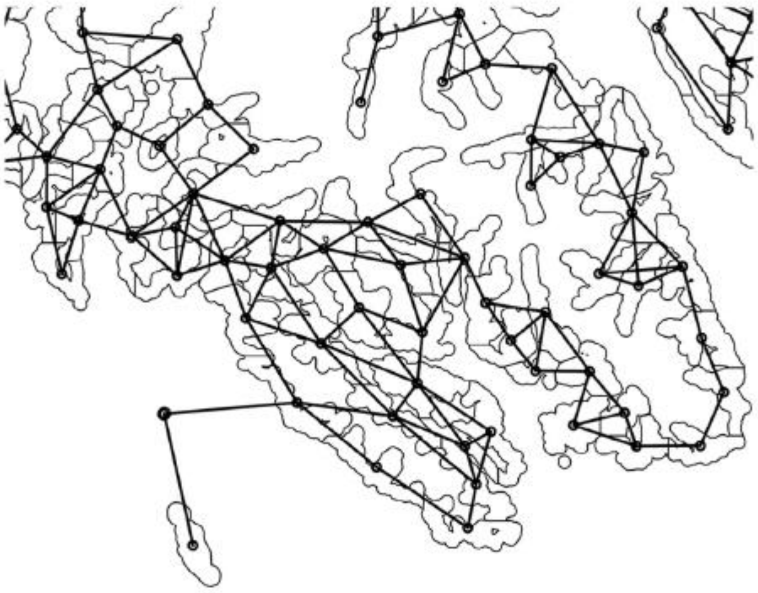
Ver Hoef et al 2018 *Ecological Monographs*

Conditional Autoregressive Models

Harbor seals in Alaska: 463 polygons

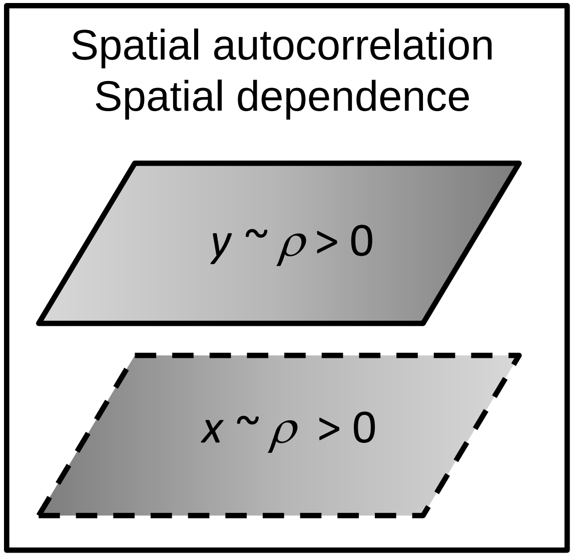


First-Neighbors

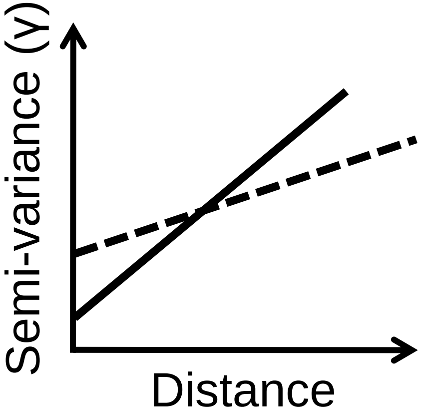


Ver Hoef et al 2018 *Ecological Monographs*

Spatial Regression



Spatial autocorrelation
and Spatial dependence



Dale & Fortin 2014

Generalized Least Square (GLS) Spatial Error Model (SEM)

$$y_i = \beta_0 + \beta_i x_i + \varepsilon_i + \boxed{\gamma\varepsilon_k}$$

Model Spatial Error Covariance

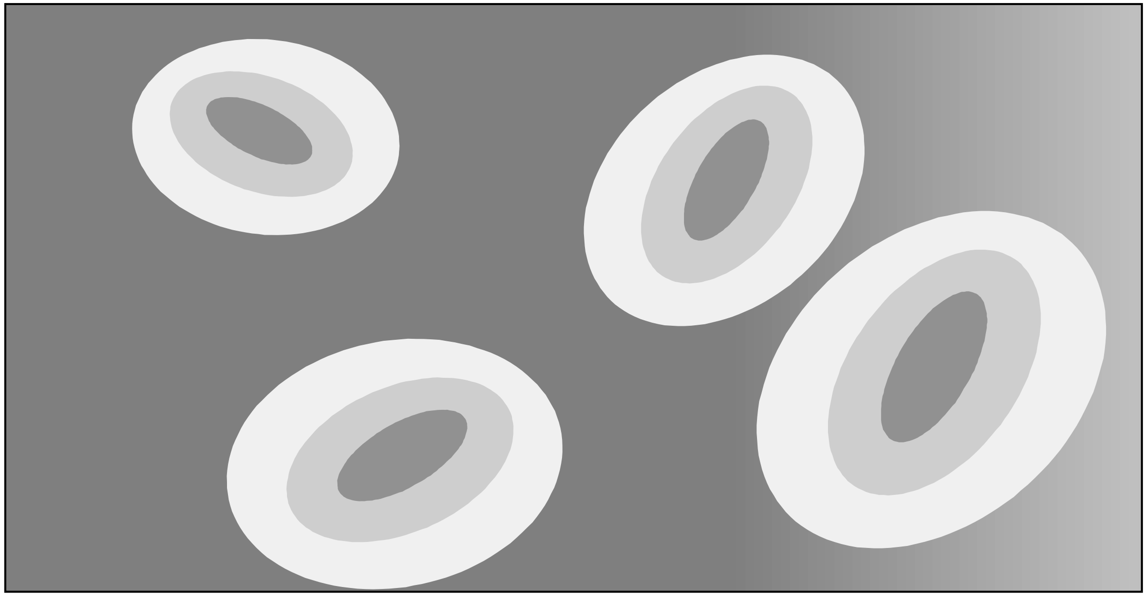
◦ Accounts for spatial autocorrelation that depends on distance only

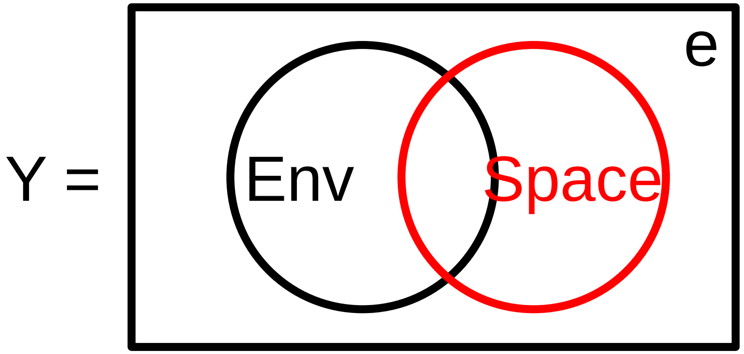
◦ Fits a variogram model to errors (geostatistics)

Estimating GLS residuals is an iterative process that first involves parameter estimation by OLS because before the regression is performed, the residuals are unknown.

Dale & Fortin 2014

Several Processes → Several Scales



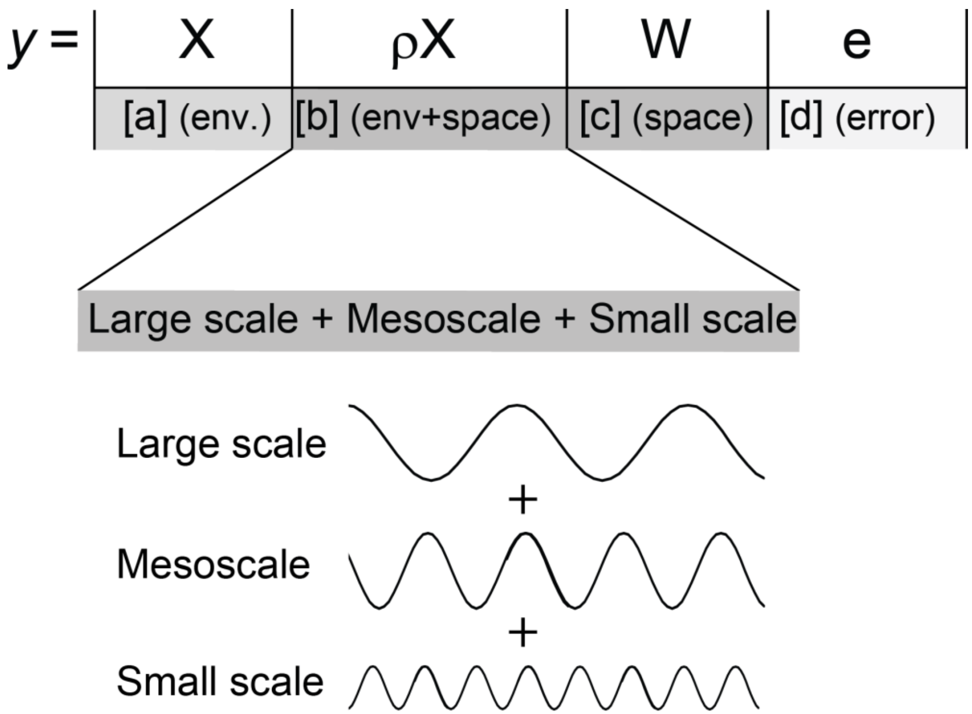


Dale & Fortin 2014

Spatial Filtering Regression

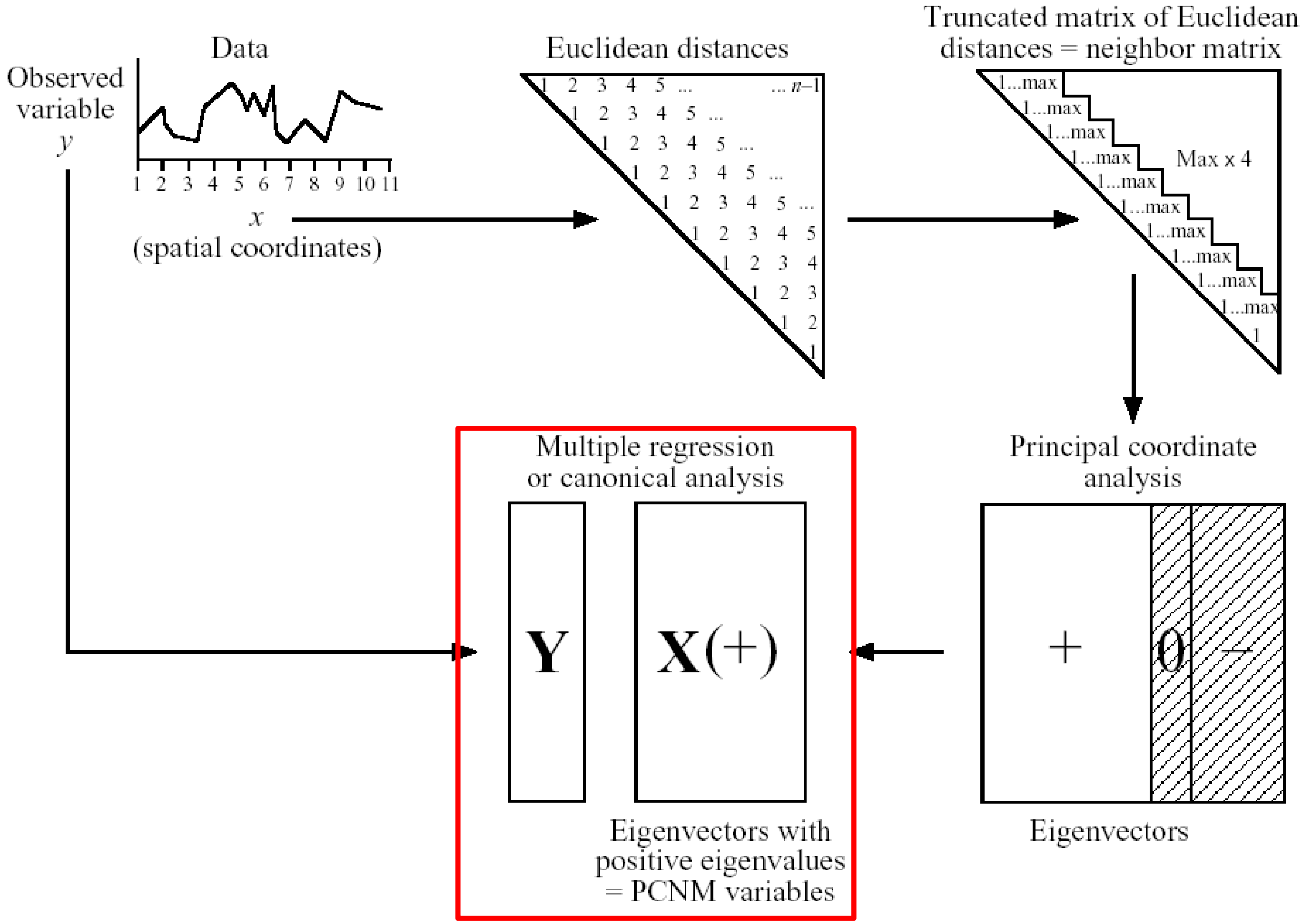
Spatial filteringregression: add spatial predictors (e.g. Moran's Eigenvector Maps):

$\{x_1(s), \ldots x_m(s), MEM_i, MEM_k\} \longrightarrow y(s) \longleftarrow \varepsilon$



Dale & Fortin 2014

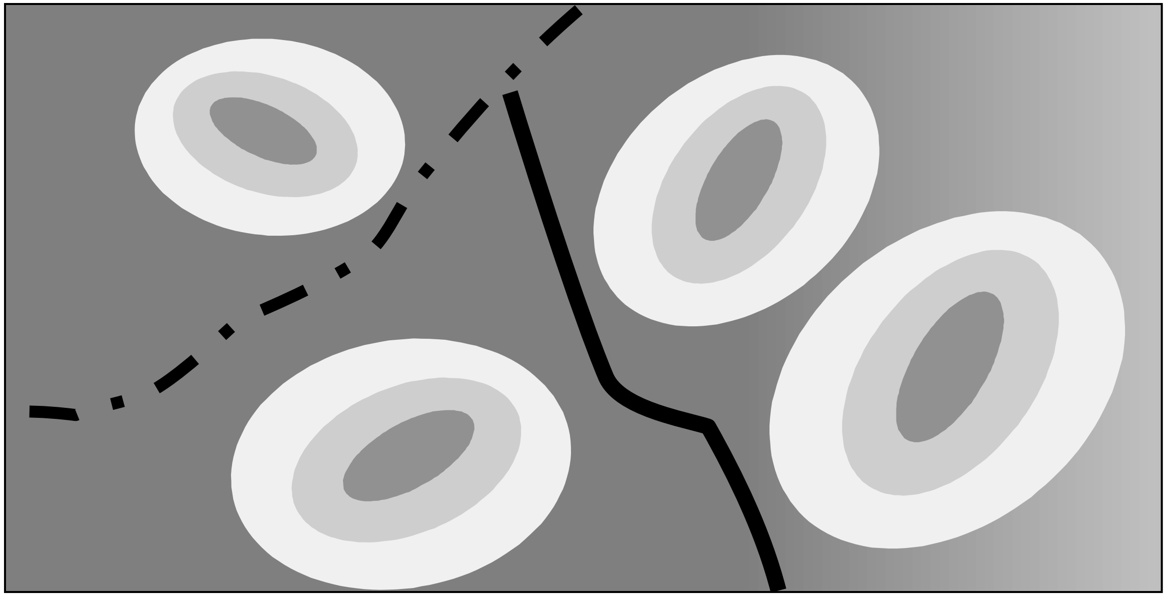
Principal Coordinate Neighbor Matrix (PCNM/dbMEM)



Truncate the matrix of geographic distances
Decompose **D** by Principal Coordinate Analysis (PCoA)
Centre **D** and then compute eigenvalues and eigenvectors

Borcard & Legendre 2002, Dray 2011

Several Processes + Several Regions



Dale & Fortin 2014

Geographically Weighted Regression

Geographically Weighted Regression (GWR) using neighbourhood to compute a linear regression at each sampling location *i*

$\{x_1(s_i), \ldots x_m(s_i)\} \longrightarrow y(s_i) \longleftarrow \varepsilon_i$

Account for localized spatial heterogeneity

$$y_i = \beta_0(i) + \beta_1(i)x_{1i} + \beta_2(i)x_{21i} + \ldots + \beta_n(i)x_{ni} + \varepsilon_i$$

Weights specific to each location *i*
such that observations nearer to *i* are given
greater weight in the regression
than observations further away.

Fotheringham et al 2003

Geographically Weighted Regression

Residuals from GWR are much lower and are less spatially autocorrelated than those of linear regression

GWR gives a much better fits even accounting for increases in the number of parameters
→ Because it is overfitting

Fotheringham et al 2003

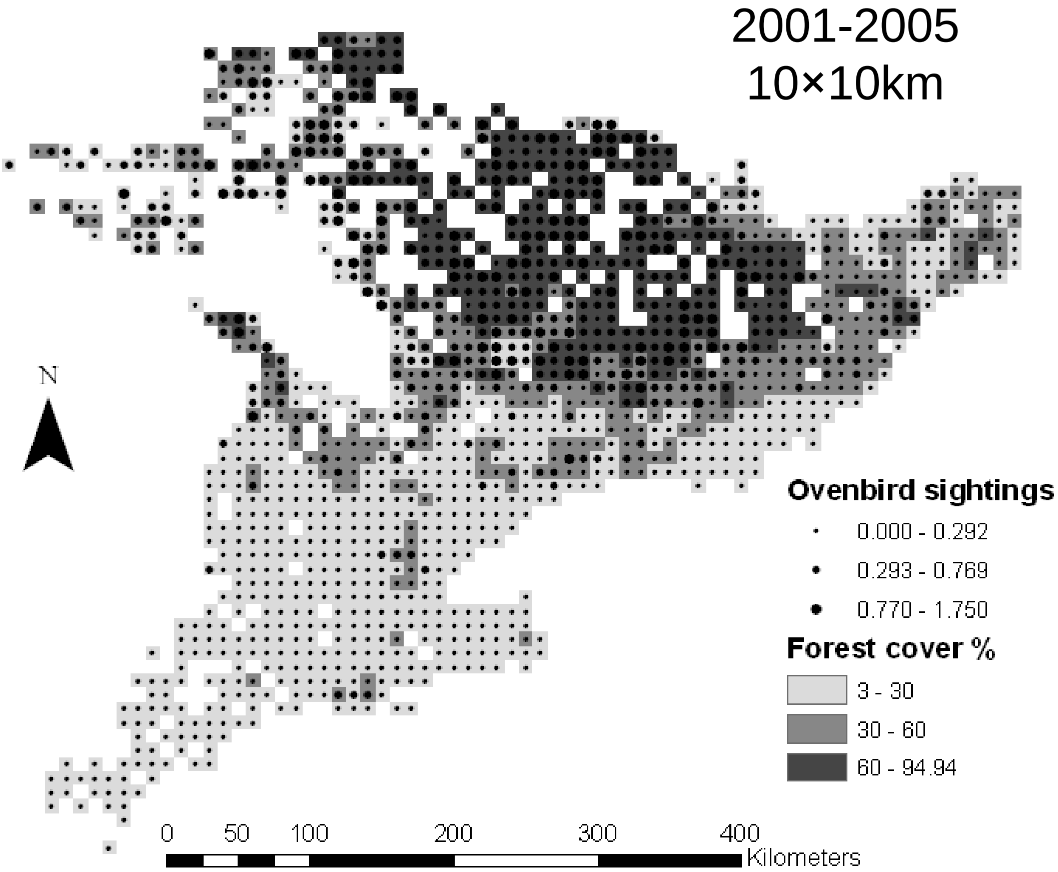
Ovenbird Sightings ~ Forest Cover

Ontario Breeding Bird Atlas

Picture by Dick Freshley

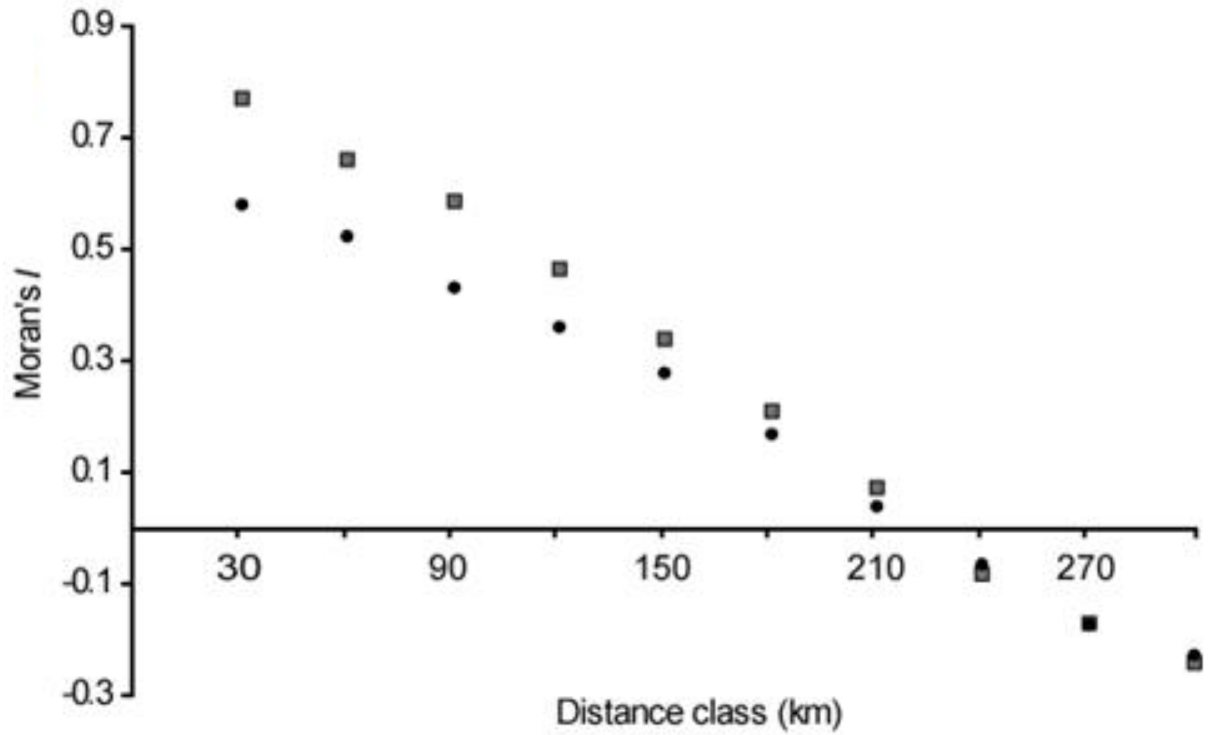
Ovenbird
Seiurus aurocapillus

Size: 6 inches
Diet: Eats insects, worms, spiders while poking under leaves.



Fortin & Melles 2009

Ovenbird Sightings ~ Forest Cover

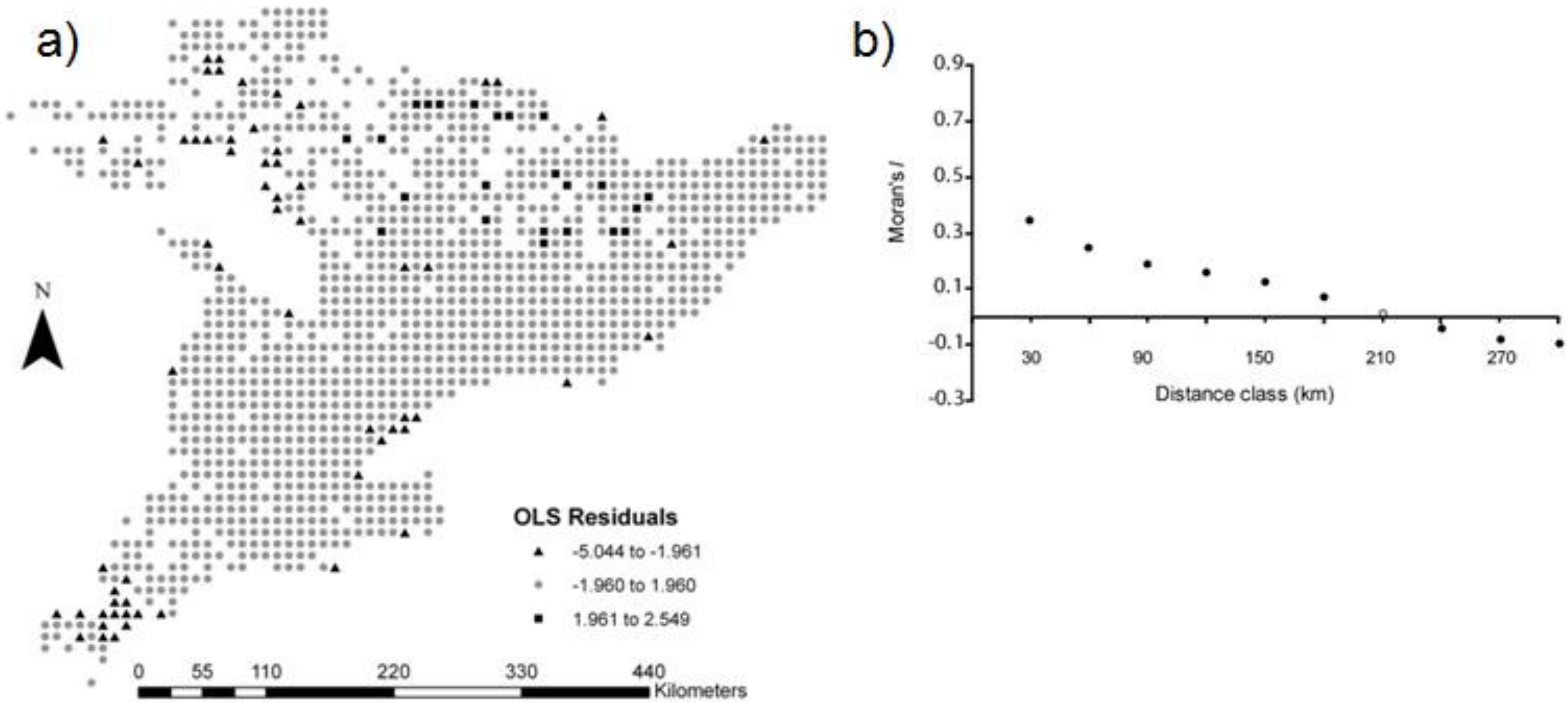


Forest cover (squares), Mean number of ovenbirds (circles) in an Atlas square 10×10 km (*n*=1359).
Solid symbols indicate significant Moran's *I* ($p \leq 0.05$)

Fortin & Melles 2009

Ovenbird Sightings ~ Forest Cover

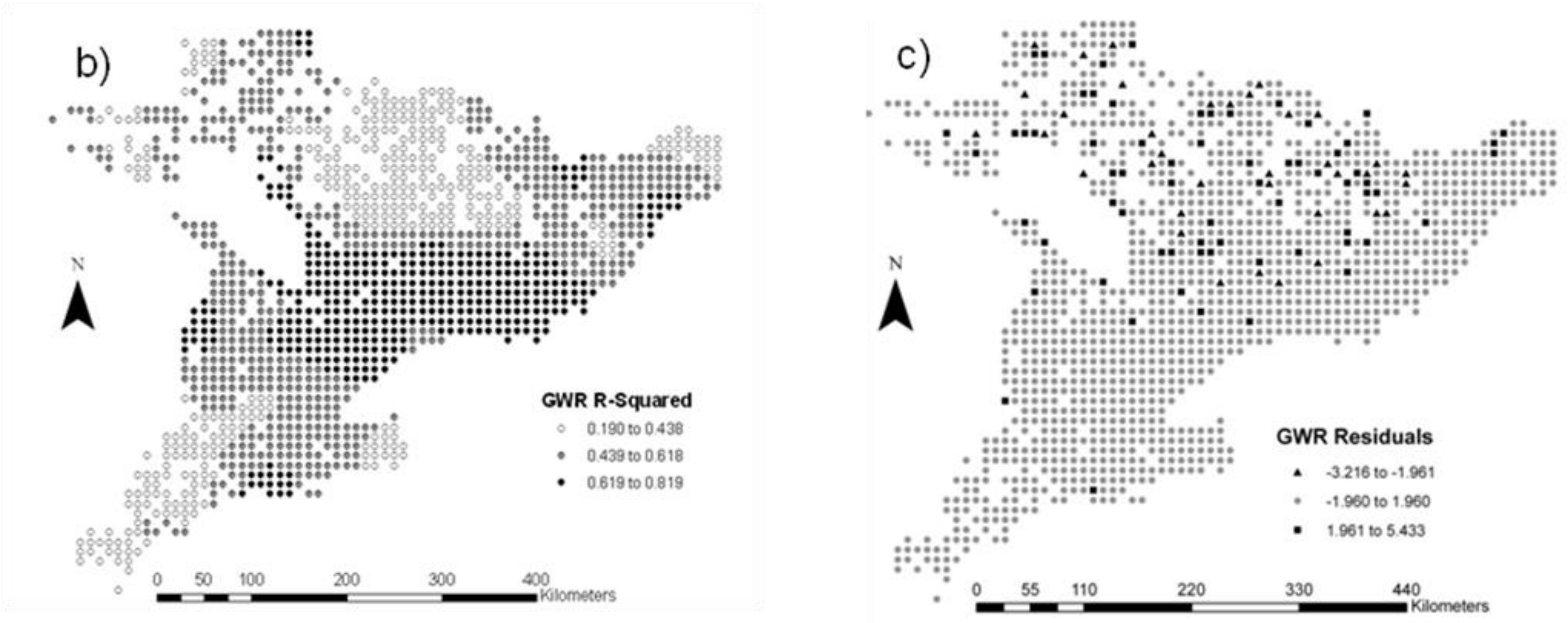
Ordinary Least Square Regression $R^2 = 0.43$



Fortin & Melles 2009

Ovenbird Sightings ~ Forest Cover

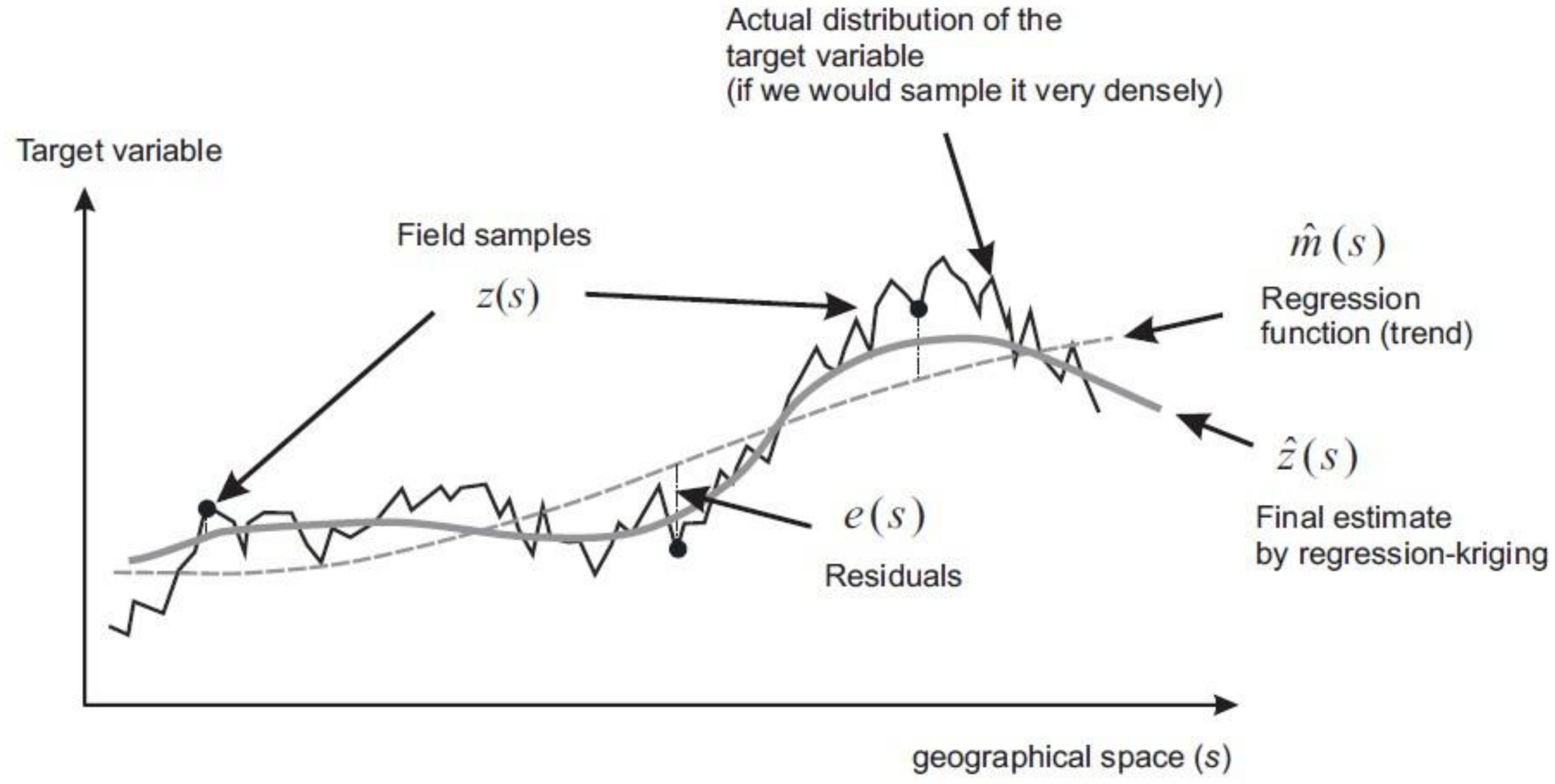
Geographically Weighted Regression $R^2 = 0.19$ to 0.82



Ordinary Least Square Regression $R^2 = 0.43$

Fortin & Melles 2009

Regression Kriging ≈ GLS



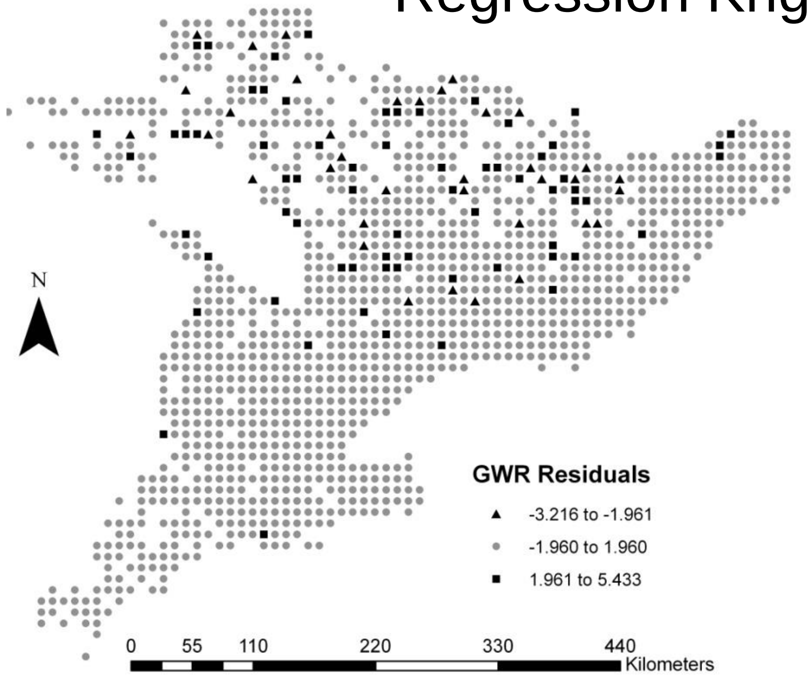
Regression kriging on residuals of linear regression

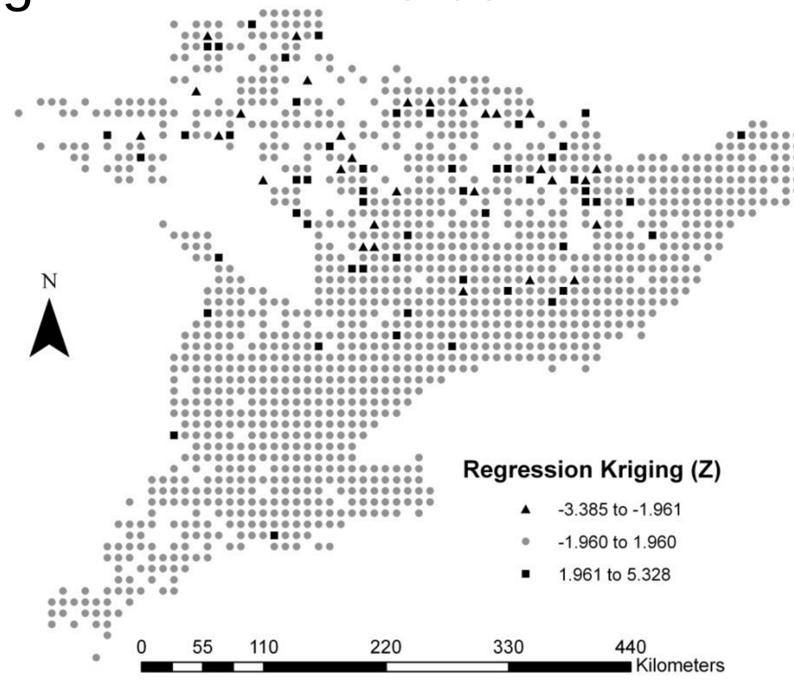
Fortin & Melles 2009

Ovenbird Sightings ~ Forest Cover

Geographical Weighted Regression $R^2 = 0.19$ to 0.82

Regression Kriging $R^2 = 0.68$



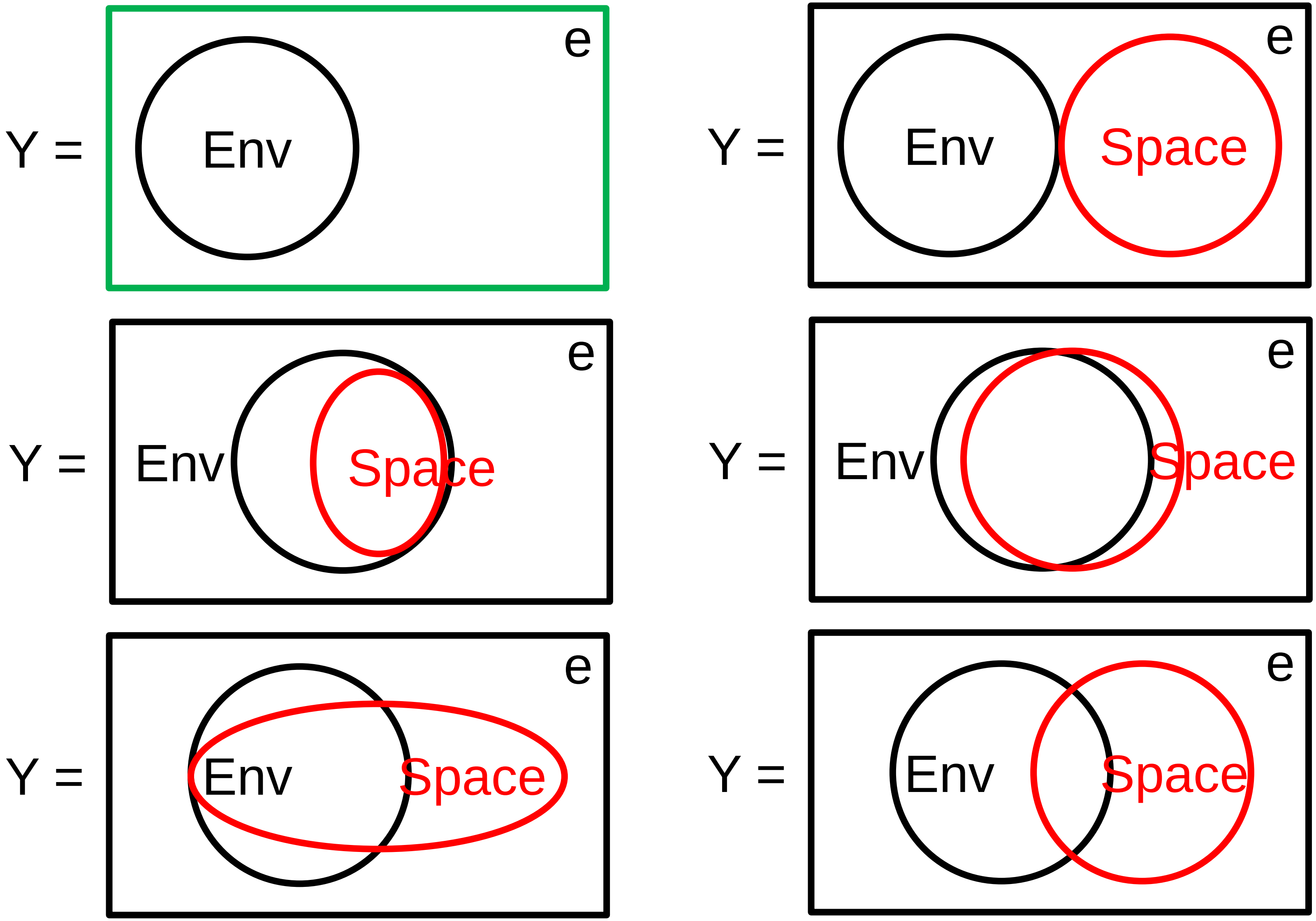


Ordinary Least Square Regression $R^2 = 0.43$

Fortin & Melles 2009

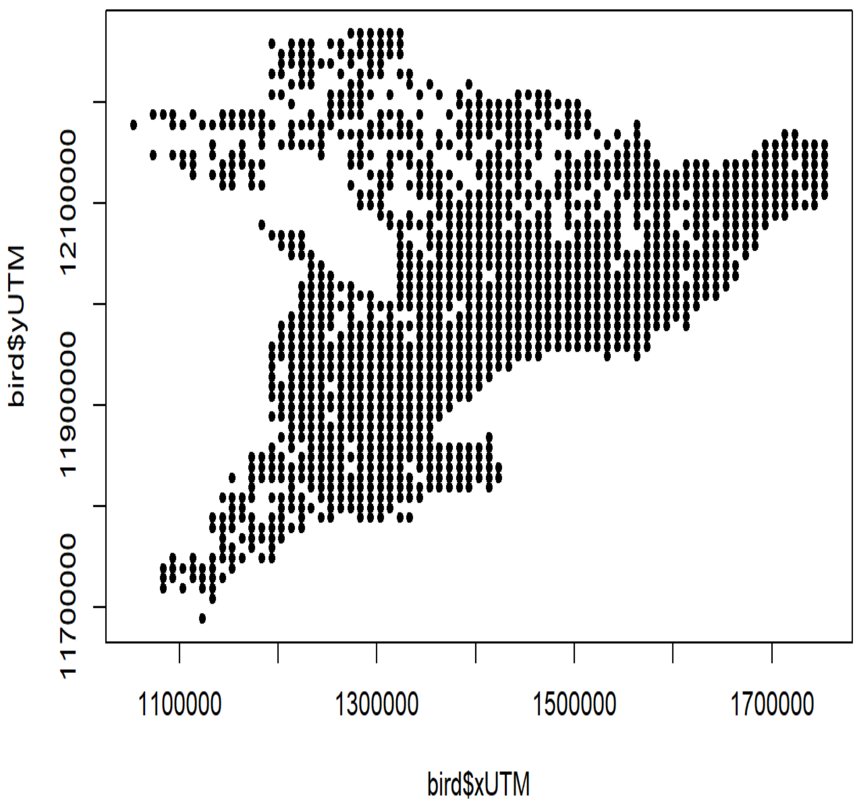
Spatial Dependence + Spatial Autocorrelation

Environmental and Spatial Contributions

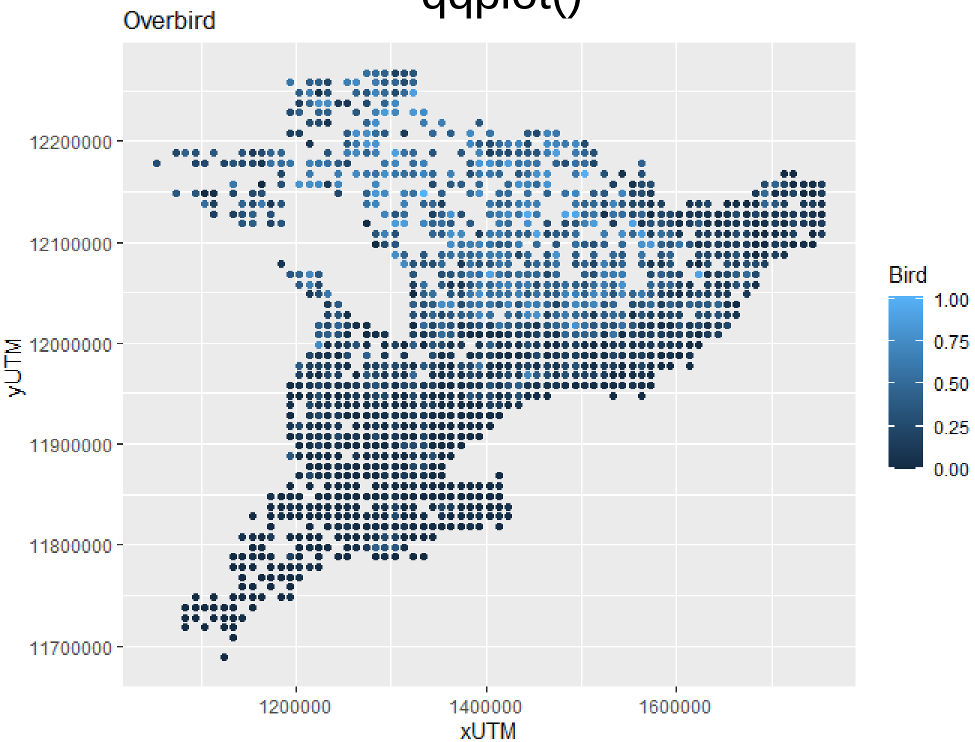


Ovenbird Sightings ~ Forest Cover
bird_forest.csv

plot()



qqplot()



Spatial Regression

library(lme4)   # GLMM
library(MuMIn)   # r.squaredGLMM

library(nlme)   # GLS

library(spdep)   # GLS (SAR-err), SAR-lag
library(spgwr)   # Geographically Weighted Regression
library(vegan)   # spatial filtering using MEM

library(care)   # CAR
library(spatialreg) # spatial filtering using MEM

```
## OLS - Linear Regression
Call: lm(formula = Bird ~ Forest)
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.268e-02  8.054e-03   1.575    0.115
Forest      6.322e-05  1.748e-06  36.168   <2e-16 ***

R-squared: 0.4908, Adjusted R-squared:0.4905
F-statistic: 1308 on 1 and 1357 DF,
p-value: < 2.2e-16
```

```
## OLS - Linear Regression
Call: lm(formula = Bird ~ Forest)
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.268e-02  8.054e-03   1.575    0.115
Forest      6.322e-05  1.748e-06  36.168   <2e-16 ***

R-squared: 0.4908, Adjusted R-squared:0.4905
F-statistic: 1308 on 1 and 1357 DF,
p-value: < 2.2e-16

## Linear regression - Trend Surface (x-y coordinates)
Call: lm(formula = Bird ~ Forest + xUTM + yUTM)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -8.179e+00  4.913e-01  -16.65   <2e-16 ***
Forest       4.968e-05  1.971e-06   25.21   <2e-16 ***
xUTM        -3.672e-07  2.882e-08  -12.74   <2e-16 ***
yUTM         7.282e-07  4.169e-08   17.47   <2e-16 ***
R-squared:  0.607,     Adjusted R-squared:  0.6061
F-statistic: 697.6 on 3 and 1355 DF
p-value: < 2.2e-16
```

Spatial Regression

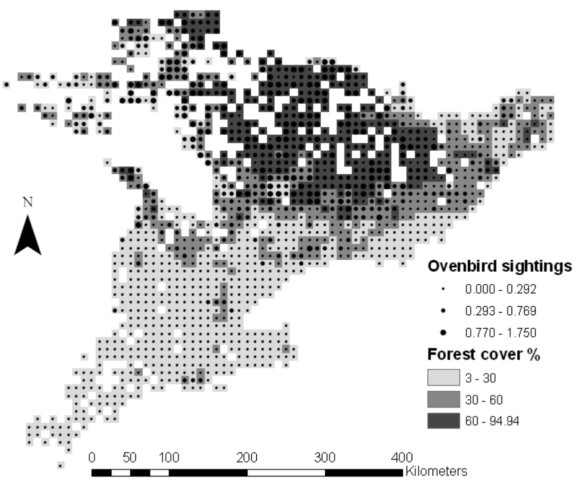
	AIC	AdjR^2	
Bird.lm	-900.3632	0.4908	
Bird.lm.xy	-1248.2828	0.6061	
Bird.random	-836.0008		0.4816
Bird.GLMM	-898.5819	0.4859	0.4880

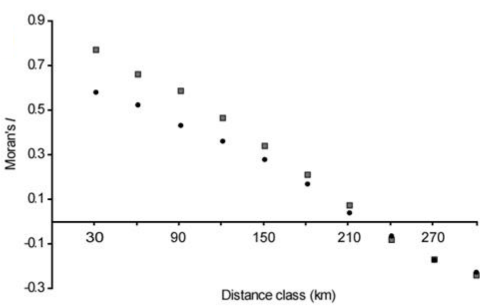
Spatial Regression: GLS (*nlme*)

```
Bird.GLSx = gls(Bird ~ Forest,
   correlation = corAR1(form = ~ 1 | xUTM))

Bird.GLSy = gls(Bird ~ Forest,
   correlation = corAR1(form = ~ 1 | yUTM))

Bird.corLin <- gls(Bird ~ Forest, correlation =
   corLin(form = ~ xUTM + yUTM))

Bird.corSpher <- gls(Bird ~ Forest, correlation =
   corSpher(form = ~ xUTM + yUTM, nugget = TRUE))
```

Spatial Regression

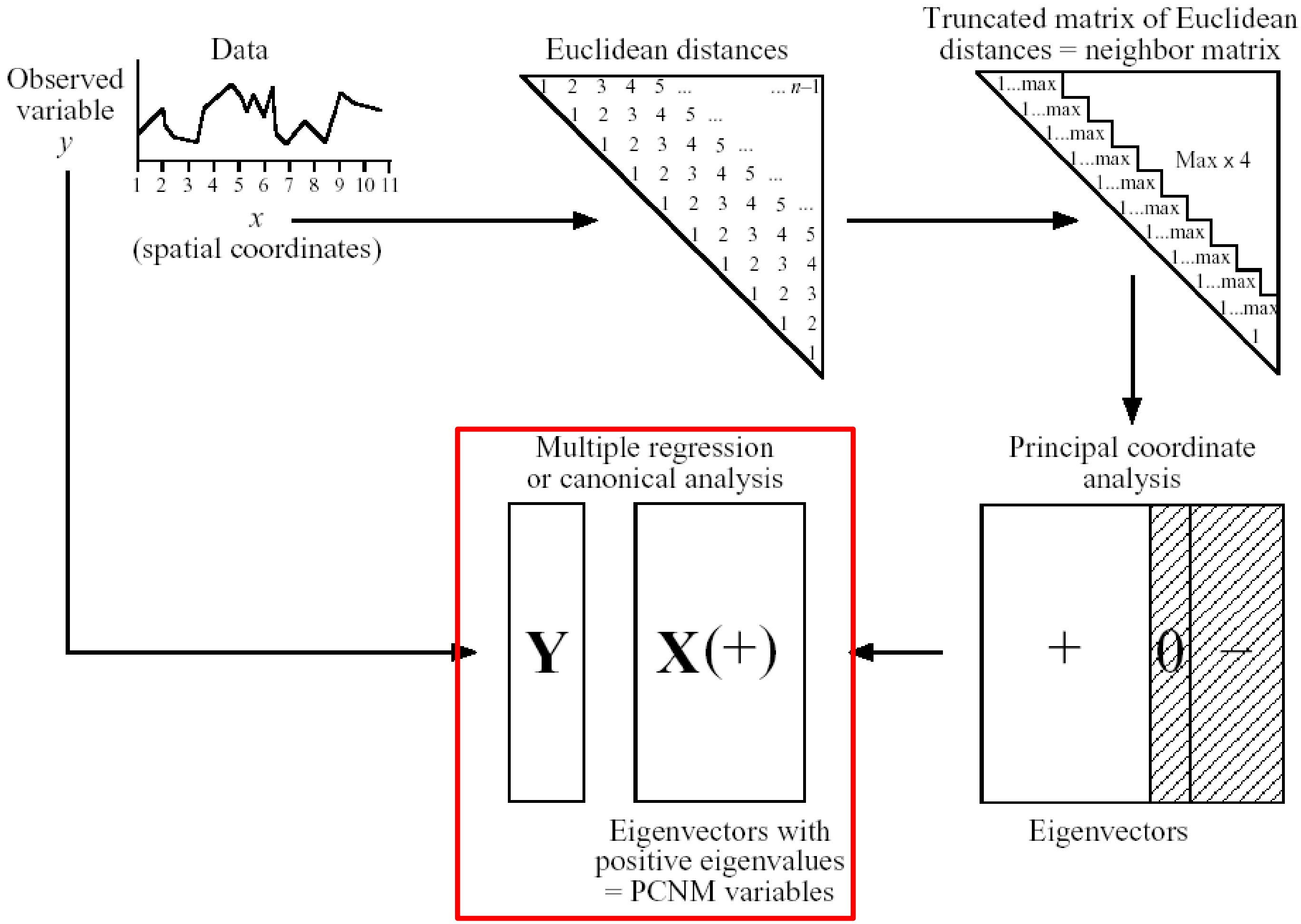
	AIC	AdjR^2	
Bird.lm	-900.3632	0.4908	
Bird.lm.xy	-1248.2828	0.6061	
Bird.random	-836.0008		0.4816
Bird.GLMM	-898.5819	0.4859	0.4880
Bird.GLSx	-885.3516		
Bird.GLSy	-930.9072		
Bird.corLin	-1219.8813		
Bird.corSpher	-1449.1100		





Forest (squares), Ovenbirds (circles)
10×10 km (n=1359)

Principal Coordinate Neighbor Matrix (PCNM/dbMEM)



Truncate the matrix of geographic distances
Decompose **D** by Principal Coordinate Analysis (PCoA)
Centre **D** and then compute eigenvalues and eigenvectors

Borcard & Legendre 2002

Spatial Regression

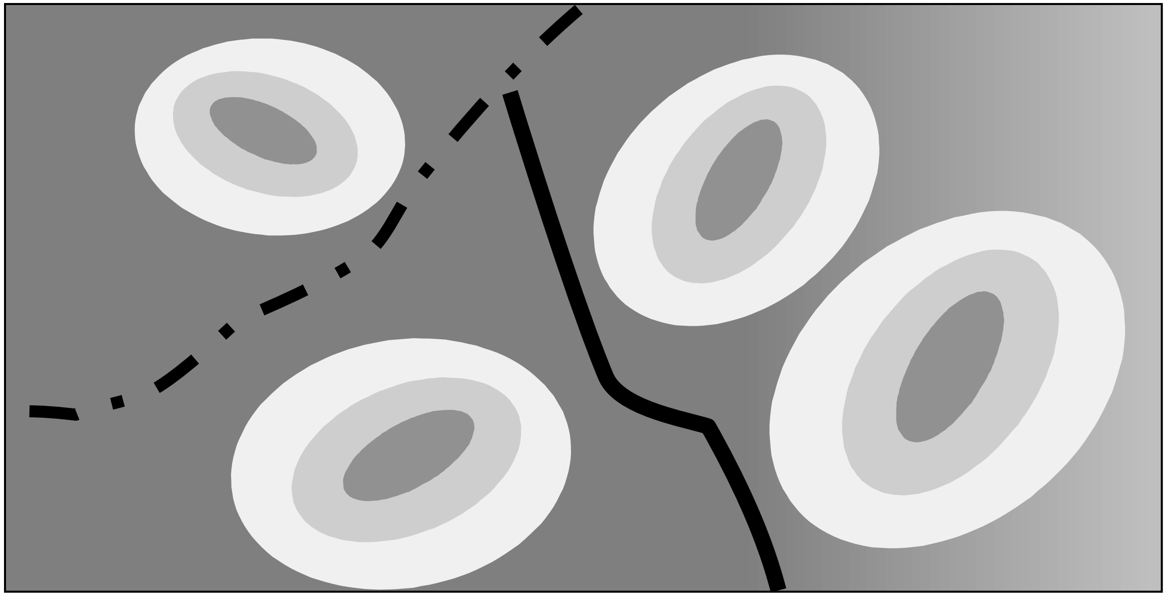
	AIC	AdjR^2	
Bird.lm	-900.3632	0.4908	
Bird.lm.xy	-1248.2828	0.6061	
Bird.random	-836.0008		0.4816
Bird.GLMM	-898.5819	0.4859	0.4880
Bird.GLSx	-885.3516		
Bird.GLSy	-930.9072		
Bird.corLin	-1219.8813		
Bird.corSpher	-1449.1100		

Spatial Filtering (PCNM/dbMEM)

dbMEM711(out of 1359): AdjR^2: 0.7168

dbMEM10 (out of 1359): AdjR^2: 0.5374

Several Processes + Several Regions

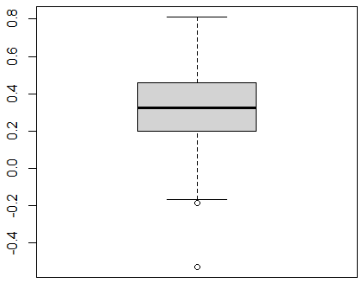


Dale & Fortin 2014

Spatial Regression: GWR

```
> Bird.gwr
Call:
gwr(formula = Bird ~ Forest, data = bird,
    coords = cbind(bird$xUTM,
    bird$yUTM), adapt = BirdGWRbandwidth,
    hatmatrix = TRUE, se.fit = TRUE)
Kernel function: gwr.Gauss
Summary of GWR coefficient estimates at data points:
                    Min.    Median        Max.    Global
X.Intercept. -2.6903e-01 2.1842e-02  1.4355e+00  0.0127
Forest       -1.1257e-04 3.5399e-05  1.2214e-04  0.0001
Number of data points: 1359
Quasi-global R2: 0.7676538
```

```
> summary(bird$grw.R2)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.5295  0.2005  0.3261  0.3394  0.4585  0.8126
```



Spatial Regression

	AIC	AdjR^2	
Bird.lm	-900.3632	0.4908	
Bird.lm.xy	-1248.2828	0.6061	
Bird.random	-836.0008		0.4816
Bird.GLMM	-898.5819	0.4859	0.4880
Bird.GLSx	-885.3516		
Bird.GLSy	-930.9072		
Bird.corLin	-1219.8813		
Bird.corSpher	-1449.1100		

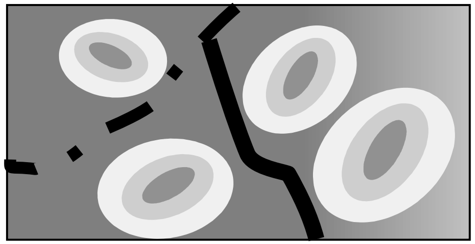
Spatial Filtering (PCNM/dbMEM)

dbMEM711: AdjR^2: 0.7168

dbMEM10: AdjR^2: 0.5374

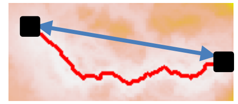
Geographically Weighted Regression R^2

Min.	Median	Mean	Max.
-0.5295	0.3261	0.3394	0.8126

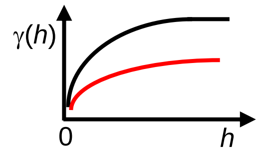


Spatial Aspects

x-y Coordinates
Euclidean Distances
Least-cost Distances

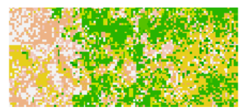


Spatial Autocorrelation
Spatial Dependence

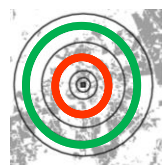


Spatial Relationship

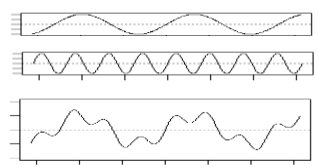
Spatial Legacy
Spatial Contingency



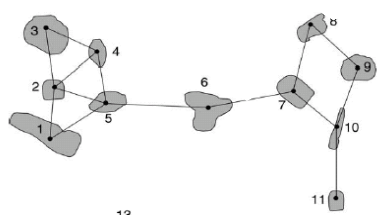
Spatial Perception



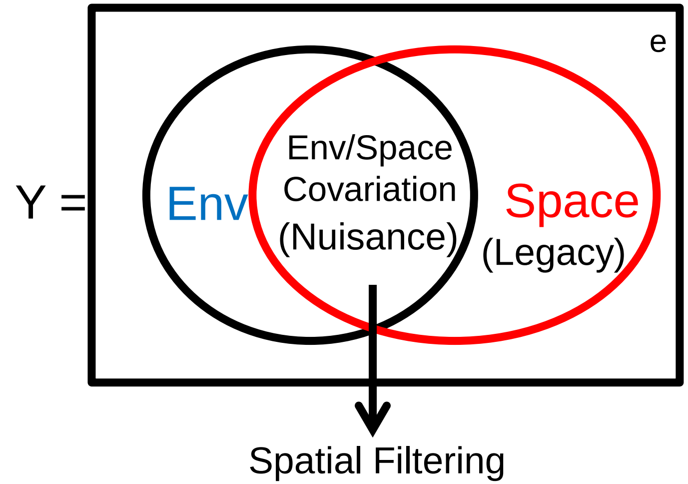
Multiscale Analysis



Metapopulation
Metacommunity
Metaecosystem
Metanetwork



Space does not replace
Environmental Factors



Better Fit: Include Space

Better Prediction/Knowledge:
Include Processes & Factors