David Orme

# Spatial modelling

David Orme

# Spatial modelling

- Start with a plug for another program
  - **S**patial **A**nalysis in **M**acroecology
  - http://www.ecoevol.ufg.br/sam/

# Overview

- Example data: Afrotropical bird diversity
- Naive models
- Describing spatial autocorrelation
- Accounting for spatial autocorrelation (degrees of freedom correction, SAR, GLS, eigenvector filtering, GWR)

# Afrotropical bird species richness

# Predictor variables

# A simple linear model

```
Call:
lm(formula = Rich ~ MeanAET + MeanAnnTemp + MeanElev,
   data = figDat)

Residuals:
      Min        1Q    Median        3Q       Max
-354.0920  -53.2450   -0.9335   52.6389  325.5849

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 189.452774  21.328794   8.882  < 2e-16 ***
MeanAET       0.176412   0.004724  37.342  < 2e-16 ***
MeanAnnTemp  -4.178441   0.722002  -5.787 8.05e-09 ***
MeanElev      0.076027   0.005490  13.849  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 82.4 on 2480 degrees of freedom
Multiple R-squared: 0.4554, Adjusted R-squared: 0.4548
F-statistic: 691.4 on 3 and 2480 DF,  p-value: < 2.2e-16
```

# A simple GAM

```
Approximate significance of smooth terms:
                edf Ref.df       F  p-value
s(MeanAET)    7.707  8.207 222.273  < 2e-16 ***
s(MeanAnnTemp) 6.987  7.487   8.259 1.57e-10 ***
s(MeanElev)   5.291  5.791  13.476 1.17e-14 ***
```

# Model predictions

# Spatial Autocorrelation

Global
Moran's I = 0.92
p << 0.001

# Spatial Autocorrelation

Global
Moran's I = -0.1
p = 0.54

# Spatial Autocorrelation

Global
Moran's I = 0.92
p << 0.001

# Spatial Autocorrelation

Global
Moran's I = 0.92
p << 0.001

# Correlograms

# Variograms

Count
0
20000
40000
60000
80000
0
2000
4000
6000
8000
Distance (km)

Semivariance
0
5000
10000
15000
20000
0
2000
4000
6000
8000
Distance (km)

# Local indicators of spatial autocorrelation (LISA)

# Effects of spatial autocorrelation

Violates assumption of independence between data points

- Degrees of freedom not equal to number of data points: **standard errors and significance testing affected**
- Data points do not contribute equally to determining the relationship between variables: **parameter estimation affected**

# Dealing with spatial autocorrelation

- Modify the degrees of freedom in significance testing
- Account for autocorrelation in models:
  - Simultaneous autoregressive models
  - Generalised least squares
  - Eigenvector filtering
  - Geographically weighted regression

# Degrees of freedom correction

# Degrees of freedom correction

# Simultaneous autoregressive models (SAR)

For each point:

- A neighbourhood
- A weighting scheme

# Simultaneous autoregressive models (SAR)

For each point:

- A neighbourhood
- A weighting scheme

# Simultaneous autoregressive models (SAR)

For each point:

- A neighbourhood
- A weighting scheme

# Simultaneous autoregressive models (SAR)

| | | | | **Simultaneous equations** |
|---|---|---|---|---|
| 1 | 1 | 0 | 0 | $b \,.\, x_1 + 1 \,.\, b \,.\, x_2$ |
| 0.5 | 1 | 0.5 | 0 | $0.5 \,.\, b \,.\, x_1 + b \,.\, x_2 + 0.5 \,.\, b \,.\, x_3$ |
| 0 | 0.5 | 1 | 0.5 | $0.5 \,.\, b \,.\, x_2 + b \,.\, x_3 + 0.5 \,.\, b \,.\, x_4$ |
| 0 | 0 | 1 | 1 | $1 \,.\, b \,.\, x_1 + b \,.\, x_4$ |

# SAR model results

# Generalised Least Squares (GLS)

# Generalised Least Squares (GLS)

# Generalised Least Squares (GLS)

# Generalised Least Squares (GLS)

# Generalised Least Squares (GLS)

# Generalised Least Squares (GLS)

Exponential correlation: $e^{-\text{distance/range}}$

| | | | |
|---|---|---|---|
| 1 | 0.64 | 0.41 | 0.26 |
| 0.64 | 1 | 0.64 | 0.41 |
| 0.41 | 0.64 | 1 | 0.64 |
| 0.26 | 0.41 | 0.64 | 1 |

# Generalised Least Squares (GLS)

Exponential correlation: $e^{-\text{distance/range}}$

| | | | |
|---|---|---|---|
| 1 | 0.64 | 0.41 | 0.26 |
| 0.64 | 1 | 0.64 | 0.41 |
| 0.41 | 0.64 | 1 | 0.64 |
| 0.26 | 0.41 | 0.64 | 1 |

# Generalised Least Squares (GLS)

Exponential correlation: $e^{-\text{distance/range}}$

| | | | |
|---|---|---|---|
| 1 | 0.64 | 0.41 | 0.26 |
| 0.64 | 1 | 0.64 | 0.41 |
| 0.41 | 0.64 | 1 | 0.64 |
| 0.26 | 0.41 | 0.64 | 1 |

# Generalised Least Squares (GLS)

Exponential correlation: $e^{-\text{distance/range}}$

| | | | |
|---|---|---|---|
| 1 | 0.64 | 0.41 | 0.26 |
| 0.64 | 1 | 0.64 | 0.41 |
| 0.41 | 0.64 | 1 | 0.64 |
| 0.26 | 0.41 | 0.64 | 1 |

# Stationarity and isotropy

Is the same process happening in:

- different locations (stationarity)?
- different directions (isotropy)?

Is the problem in:

- the spatial structure of autocorrelation?
- differences in the actual relationship?

# Eigenvector filtering

- Transform a spatial weights model into a series of eigenvectors
- Use eigenvectors as *variables* in the model
- Use a selection process to identify and include only important eigenvectors

# Eigenvector filtering

# Eigenvector filtering

```
Call:
lm(formula = Rich ~ MeanAET + MeanAnnTemp + MeanElev, data =
figDat)

Residuals:
      Min        1Q    Median        3Q       Max
-354.0920  -53.2450   -0.9335   52.6389  325.5849

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 189.452774  21.328794   8.882  < 2e-16 ***
MeanAET       0.176412   0.004724  37.342  < 2e-16 ***
MeanAnnTemp  -4.178441   0.722002  -5.787 8.05e-09 ***
MeanElev      0.076027   0.005490  13.849  < 2e-16 ***
```

# Eigenvector filtering

```
Call:
lm(formula = Rich ~ MeanAET + MeanAnnTemp + MeanElev + Re(spEV1) +
    Re(spEV2) + Re(spEV3) + Re(spEV4), data = figDat@data)

Residuals:
     Min       1Q   Median       3Q      Max
-385.507  -49.879   -1.487   46.563  332.406

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.923e+01  3.306e+01   1.792   0.0733 .
MeanAET      1.928e-01  5.963e-03  32.331  < 2e-16 ***
MeanAnnTemp  6.057e-01  1.141e+00   0.531   0.5957
MeanElev     7.934e-02  6.089e-03  13.029  < 2e-16 ***
Re(spEV1)   -2.231e+02  1.082e+01 -20.613  < 2e-16 ***
Re(spEV2)   -1.505e+02  2.000e+01  -7.526 7.26e-14 ***
Re(spEV3)    1.450e+02  1.760e+01   8.237 2.83e-16 ***
Re(spEV4)    1.108e+01  1.872e+01   0.592   0.5540
```

# Eigenvector filtering

```
Call:
lm(formula = Rich ~ MeanAET + MeanAnnTemp + MeanElev + Re(spEV1) +
    Re(spEV2) + Re(spEV3), data = figDat@data)

Residuals:
     Min       1Q   Median       3Q      Max
-384.409  -50.367   -1.606   46.507  332.070

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.720e+01  3.018e+01   2.227   0.0261 *
MeanAET      1.904e-01  4.361e-03  43.659  < 2e-16 ***
MeanAnnTemp  3.752e-01  1.073e+00   0.350   0.7265
MeanElev     7.841e-02  5.884e-03  13.326  < 2e-16 ***
Re(spEV1)   -2.234e+02  1.081e+01 -20.663  < 2e-16 ***
Re(spEV2)   -1.471e+02  1.913e+01  -7.688 2.14e-14 ***
Re(spEV3)    1.420e+02  1.689e+01   8.408  < 2e-16 ***
```

# Geographically weighted regression (GWR)

# Geographically weighted regression (GWR)

# Geographically weighted regression (GWR)

# Problems

- Profusion of packages: sp, spdep, mgcv, ncf, gstat, nlme, spgwr
- Different package data structures require some programming skill
- Sometimes poor documentation
- Speed of calculation (= size of dataset)
- Memory hungry