# Spatial modelling

## David Orme

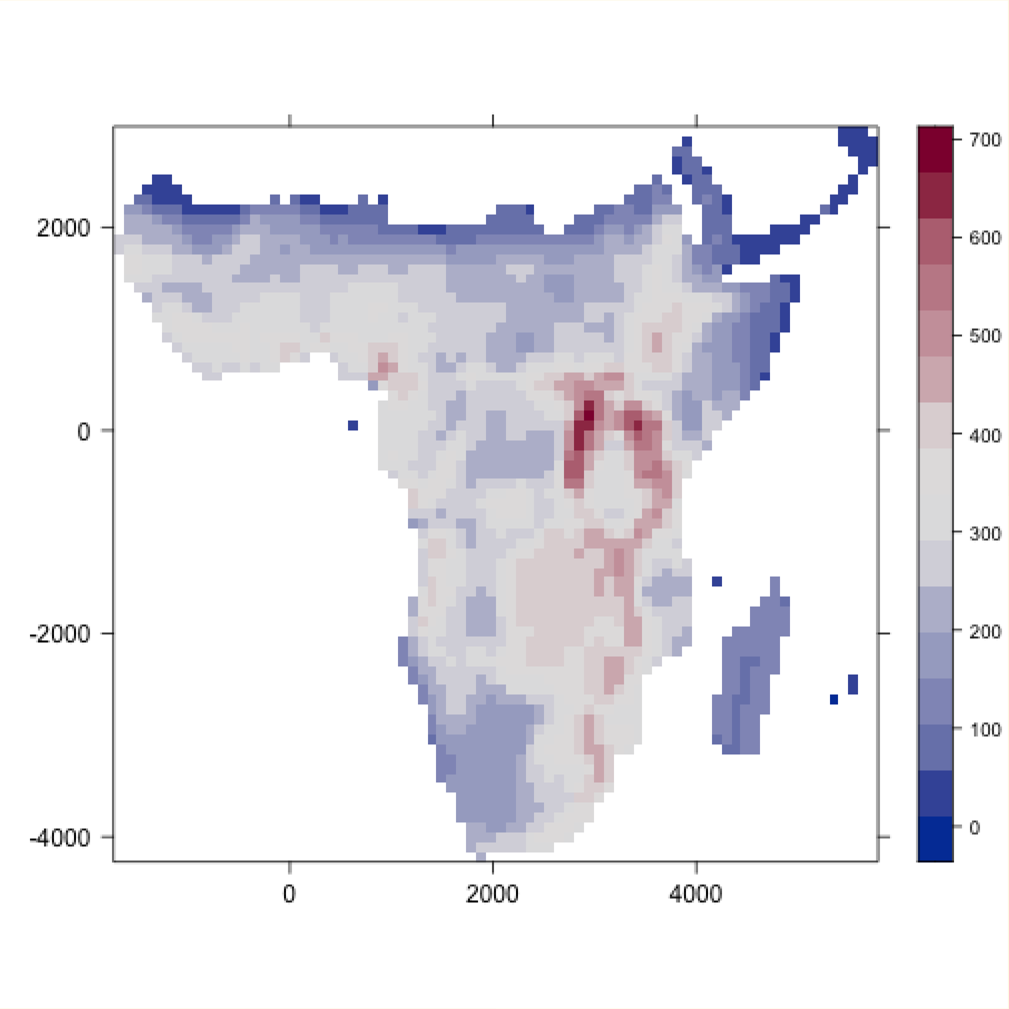
# Spatial modelling tools

- The examples presented here use R
- Another excellent program with a nice GUI interface:
  - Spatial Analysis in Macroecology
  - http://www.ecoevol.ufg.br/sam/

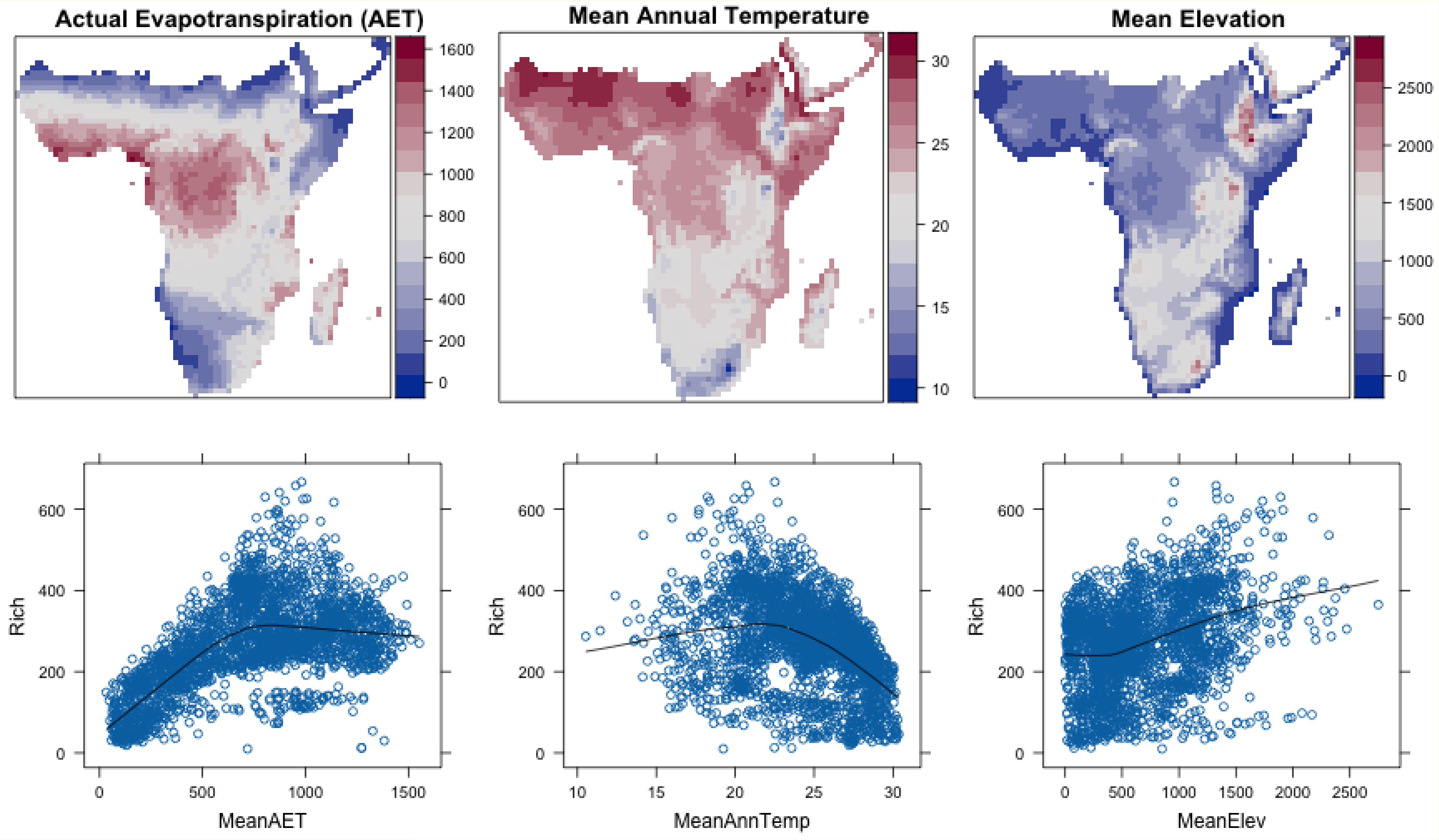
# Overview

- Example data: Afrotropical bird diversity
- Naive models
- Describing spatial autocorrelation
- Accounting for spatial autocorrelation

# Afrotropical bird species richness



# Explanatory variables



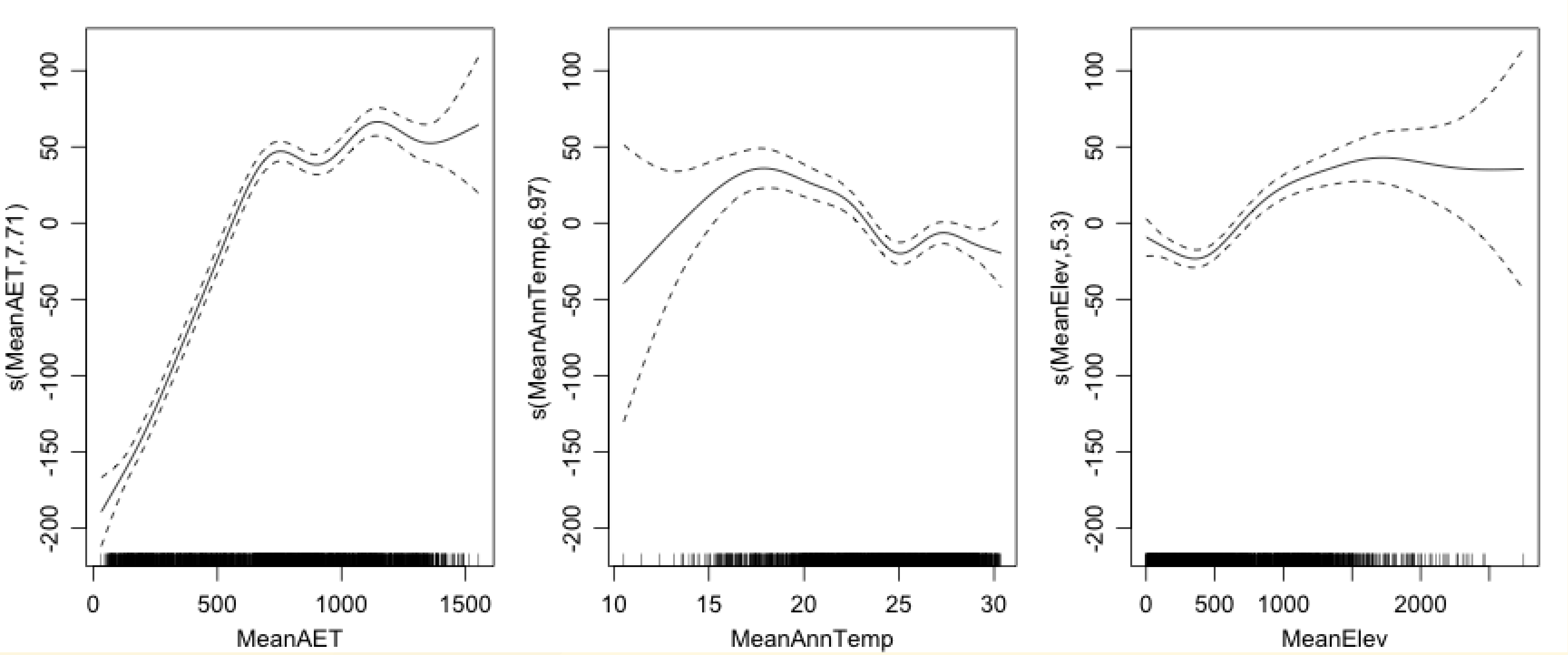
# A simple linear model

**Richness ~ AET + Temperature + Elevation**

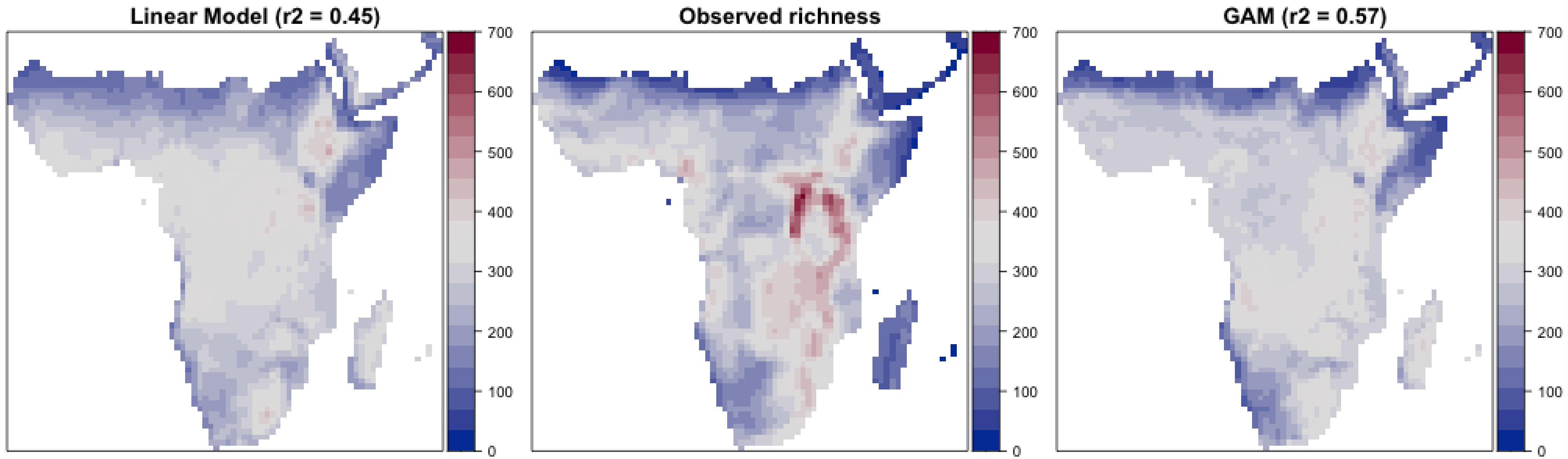
| | Est | SE | t | p |
|---|---|---|---|---|
| (Intercept) | 189.45 | 21.33 | 8.88 | < 0.001 |
| MeanAET | 0.18 | 0.00 | 37.34 | < 0.001 |
| MeanAnnTemp | -4.18 | 0.72 | -5.79 | < 0.001 |
| MeanElev | 0.08 | 0.01 | 13.85 | < 0.001 |

# A simple GAM

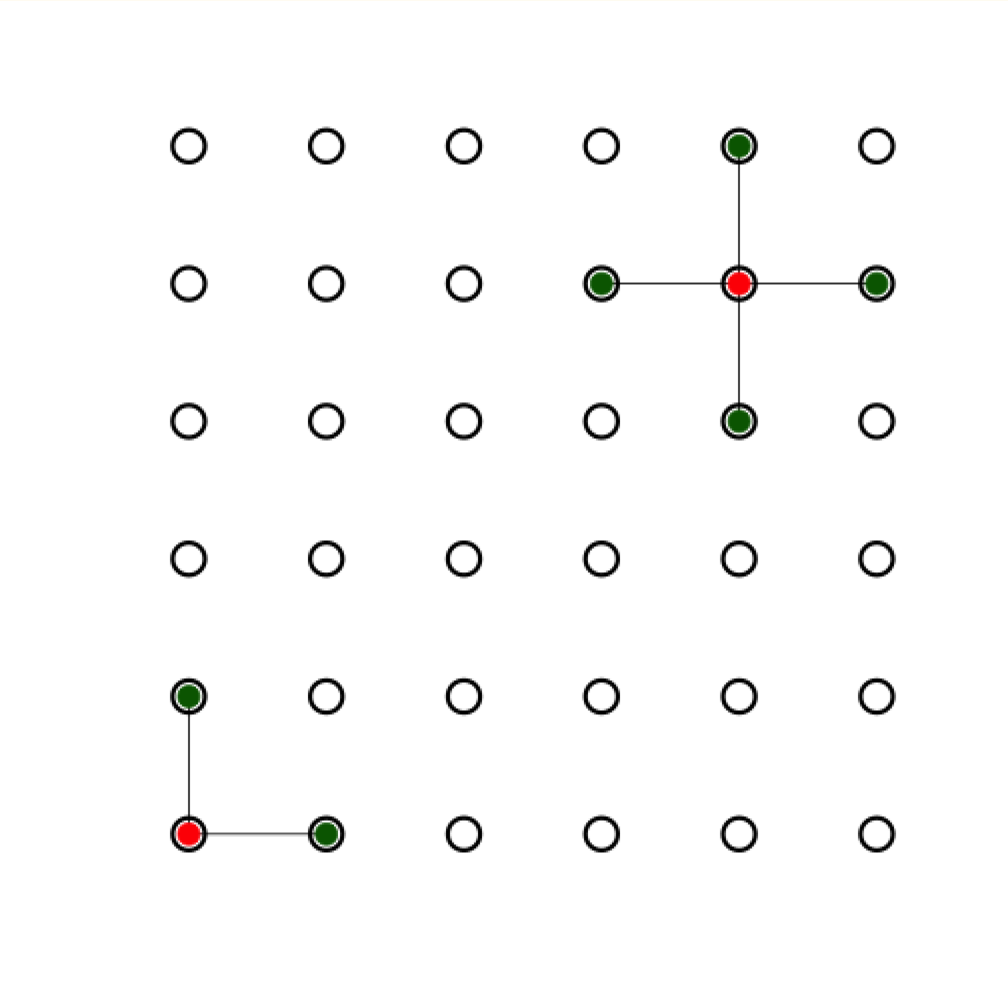
## Richness ~ s(AET) +s( Temperature) + s(Elevation)



# Model predictions



# Neighbourhoods

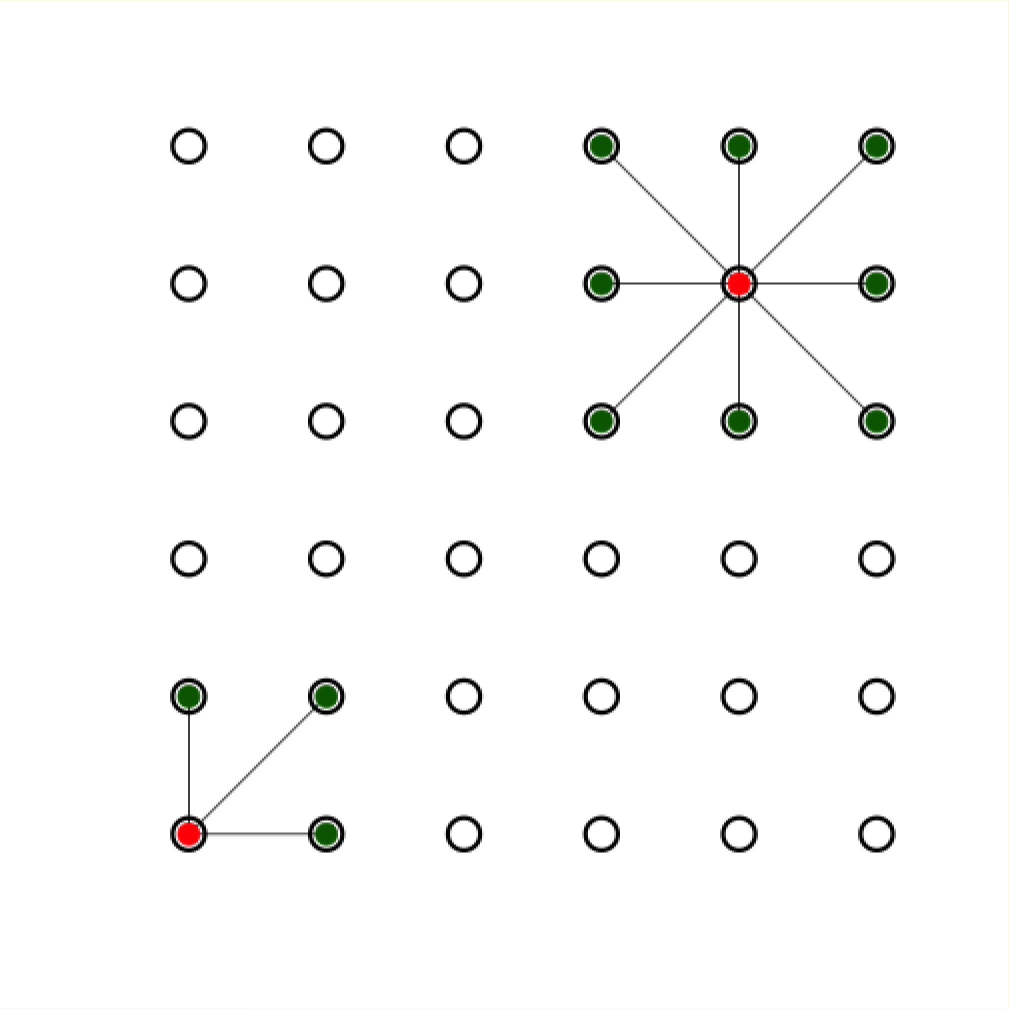


**Rooks move**

All cells within one step:

- vertically or
- horizontally

# Neighbourhoods

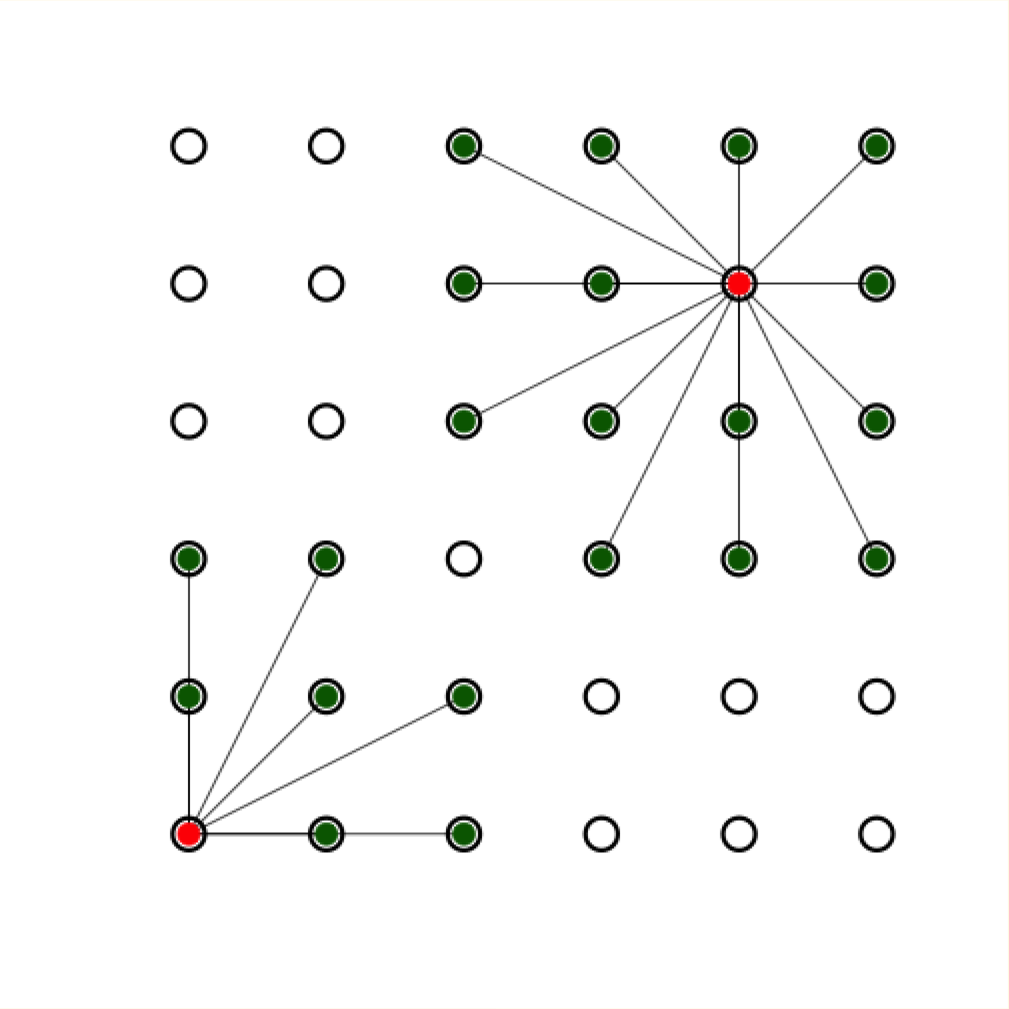


**Queens move**

All cells within one step:

- vertically,
- horizontally or
- diagonally

# Neighbourhoods

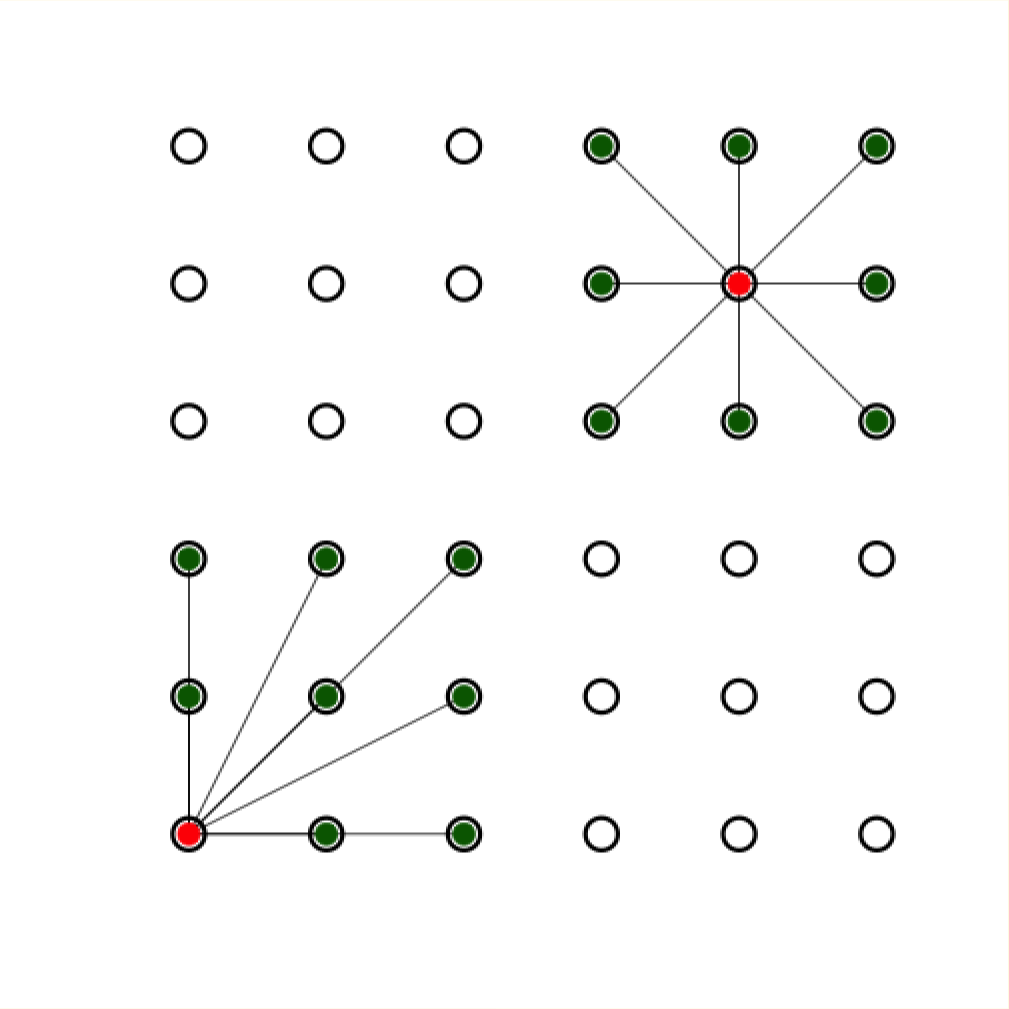


**Distance based**

All cells within:

- 2.4 units

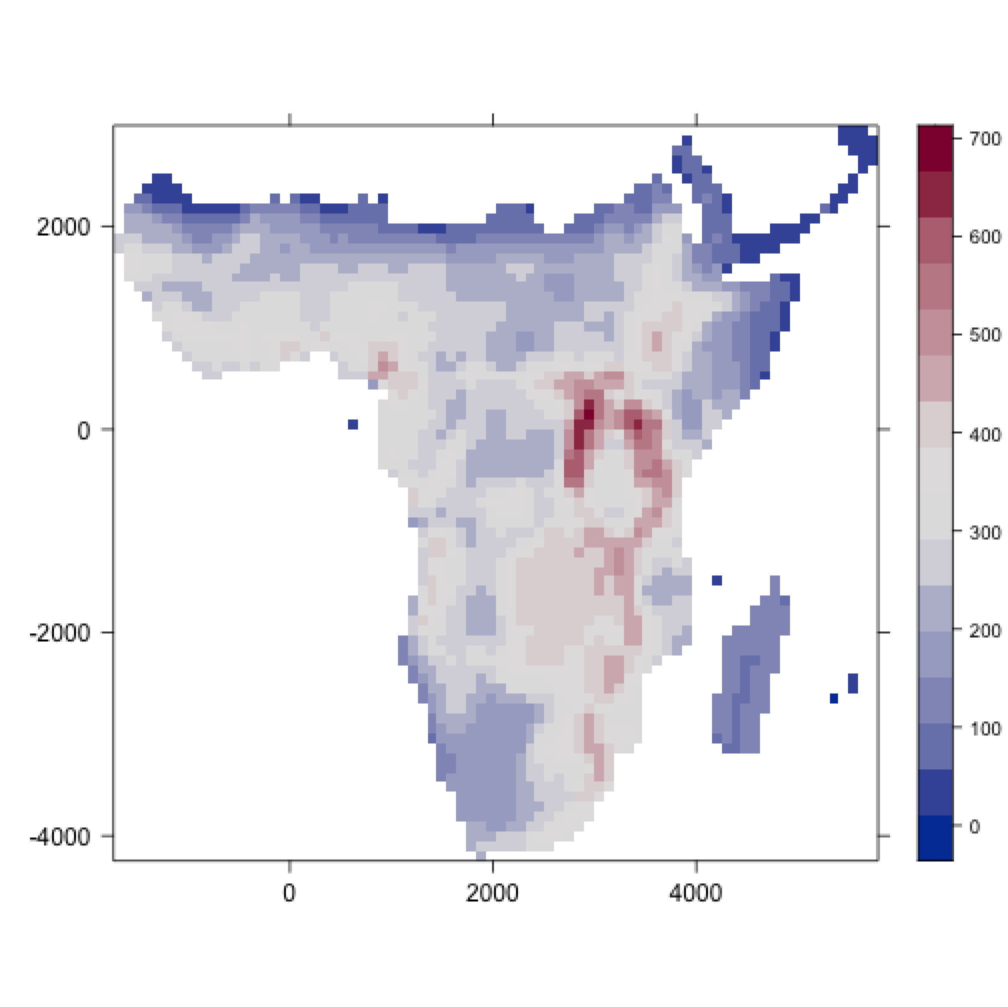
# Neighbourhoods



***k* nearest**

The closest $k$ cells

# Spatial autocorrelation



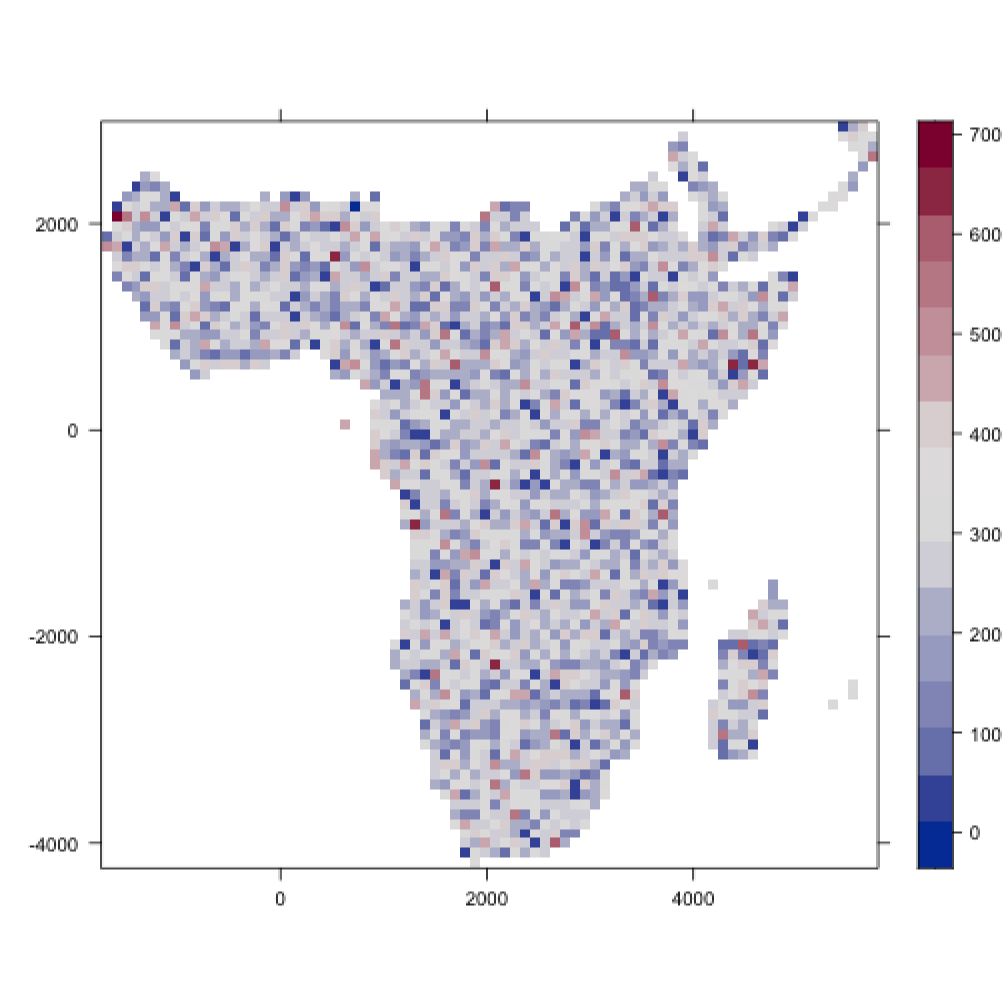

Global Moran's I

- I = 0.922
- p << 0.001

Global Geary's C

- C = 0.070
- p << 0.001

# Spatial autocorrelation




Global Moran's I

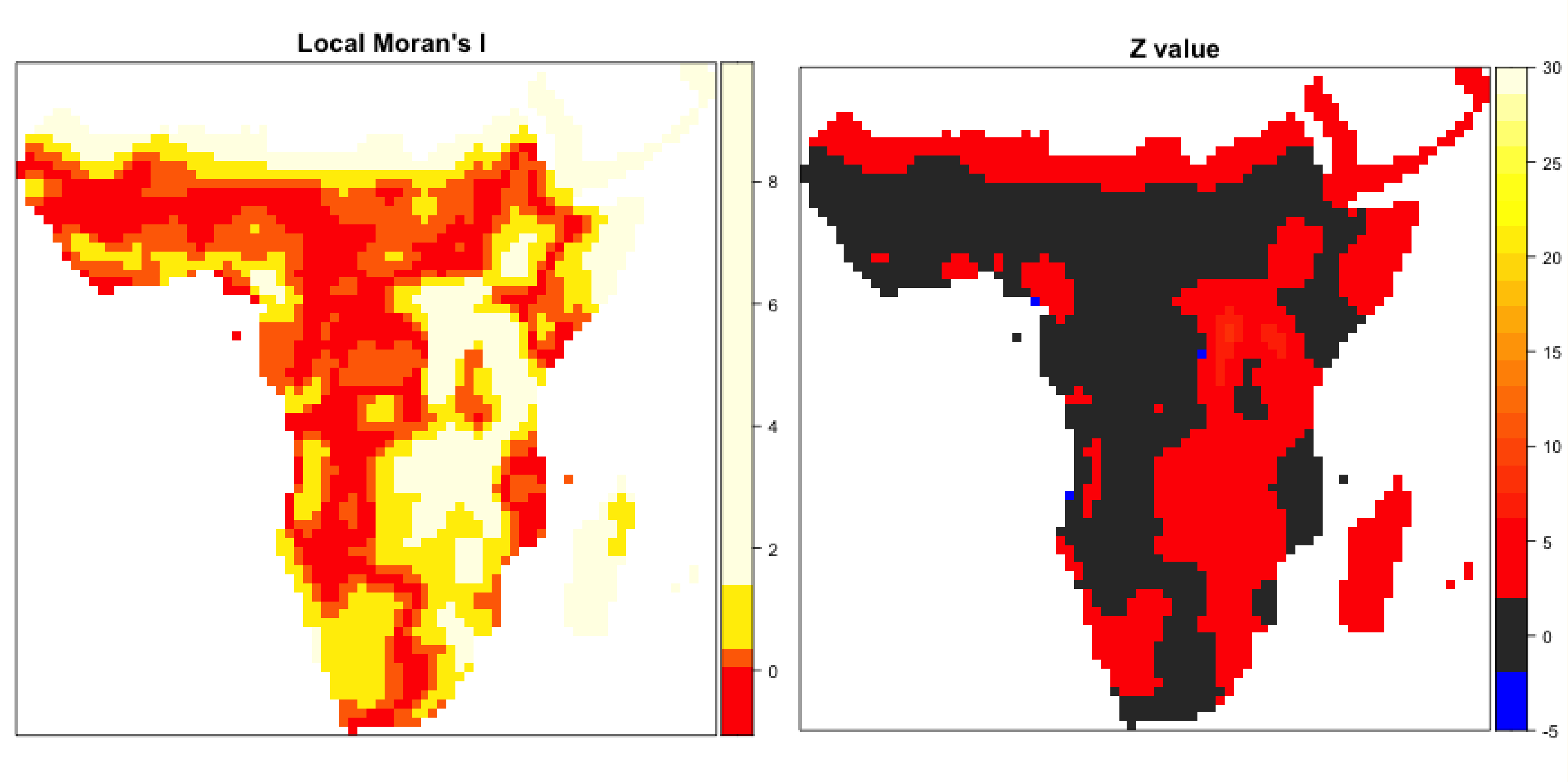
- I = -0.001
- p = 0.522

Global Geary's C

- C = 0.999
- p = 0.460

# Local autocorrelation

Local indicators of spatial autocorrelation (LISA)



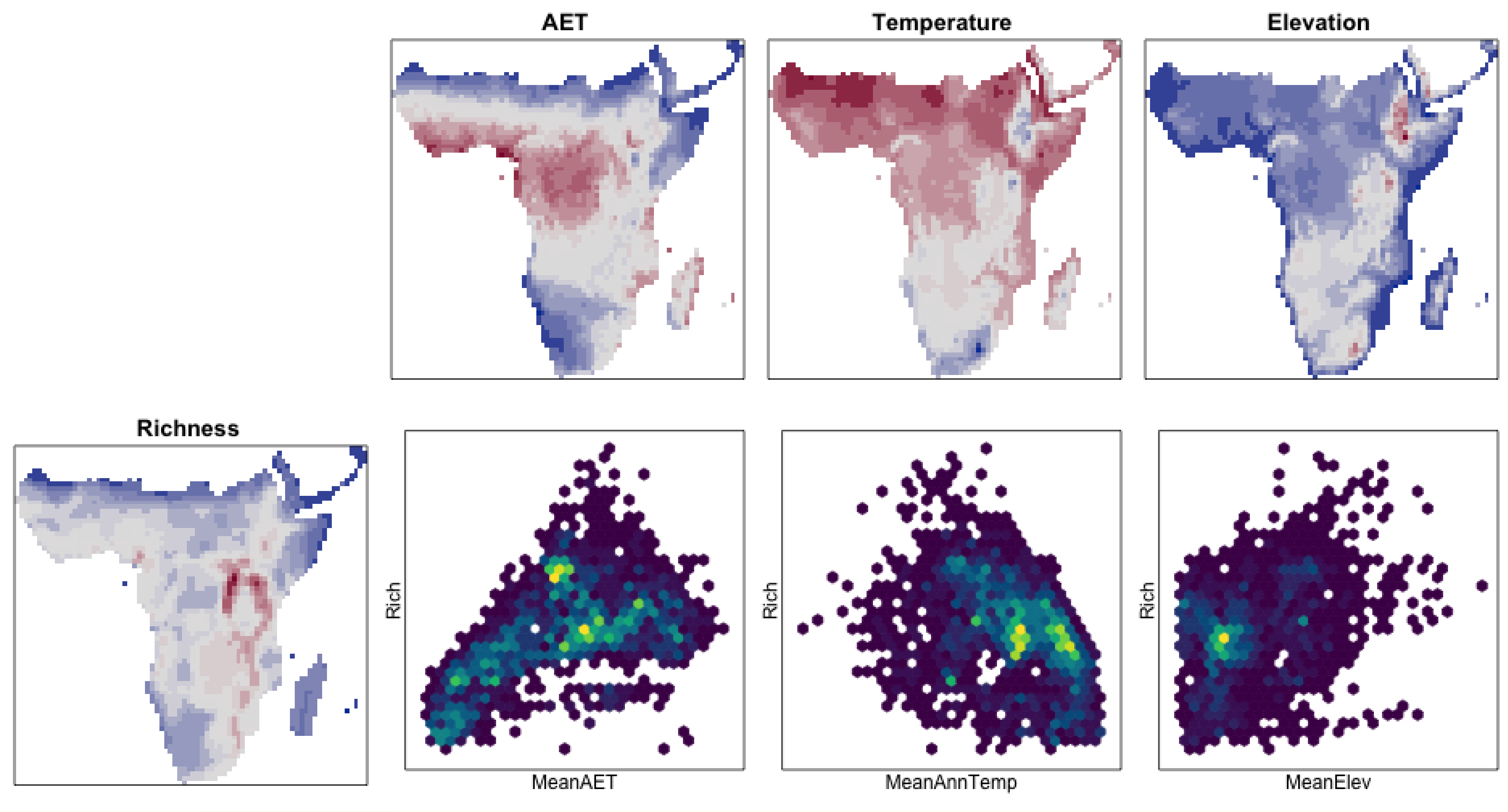
# Effects of Spatial Autocorrelation

- Data points **not independent**
- Degrees of freedom reduced:
  - **standard errors and significance testing affected**
- Not equally weighted :
  - **parameter estimation affected**

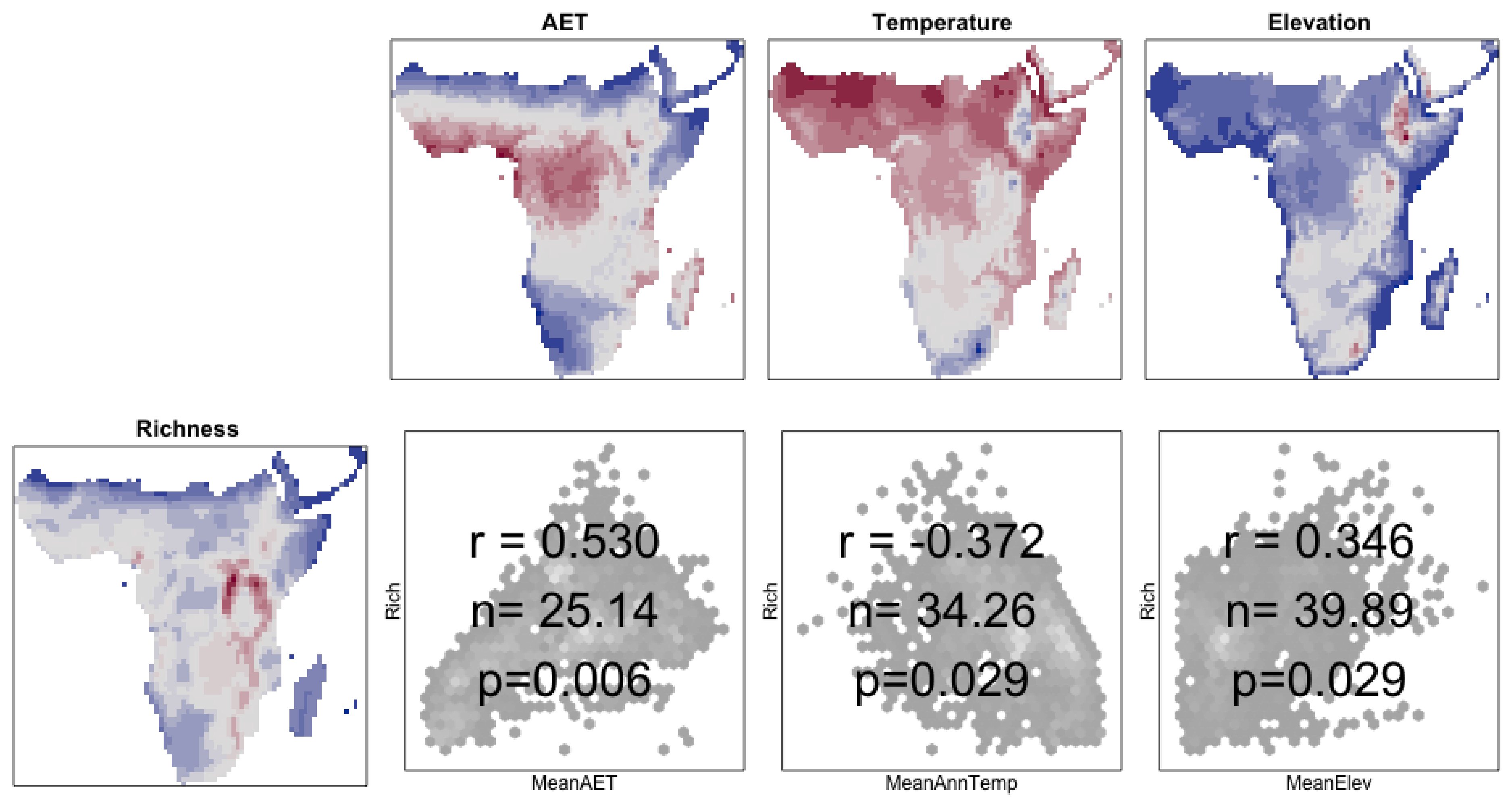
# Dealing with Spatial Autocorrelation

- 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

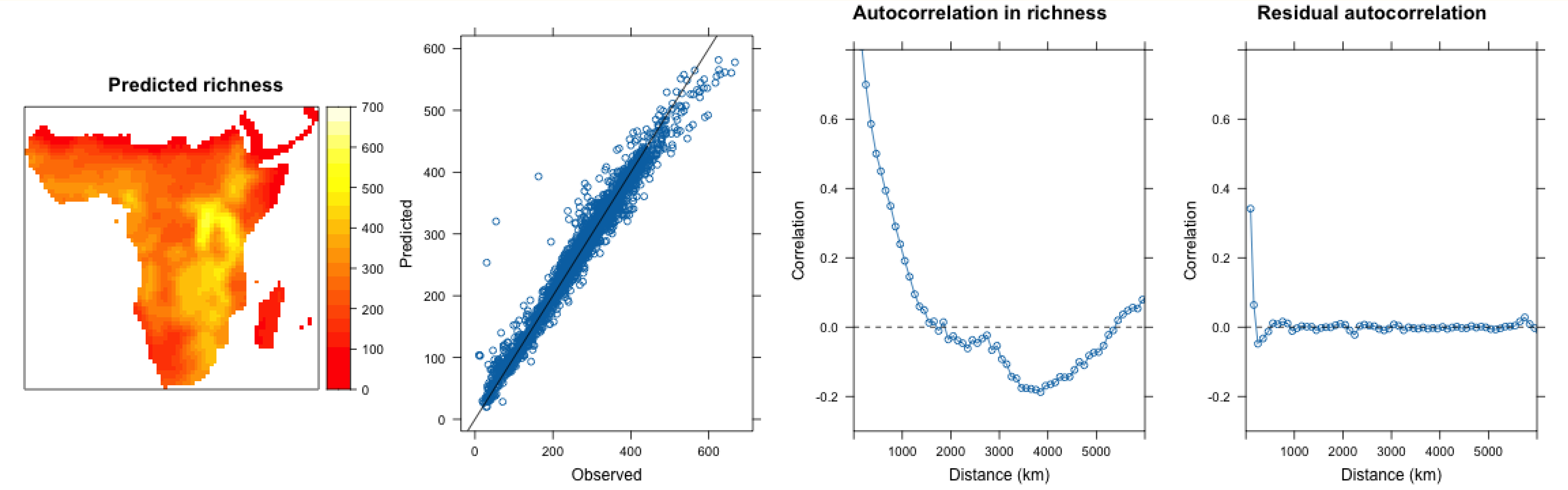


# Spatial Autoregression

| | | | | |
|---|---|---|---|---|
| | | | | $bx_1 + \frac{1}{2}bx_2$ |
| | | | | $\frac{1}{2}bx_1 + bx_2 + \frac{1}{2}bx_3$ |
| | | | | $\frac{1}{2}bx_2 + bx_3 + \frac{1}{2}bx_4$ |
| | | | | $\frac{1}{2}bx_3 + \frac{1}{2}bx_4$ |
| $x_1$ | $x_1$ | $x_3$ | $x_4$ | |

The best fit for coefficient $b$ is influenced by the neighbour and weightings schemes.

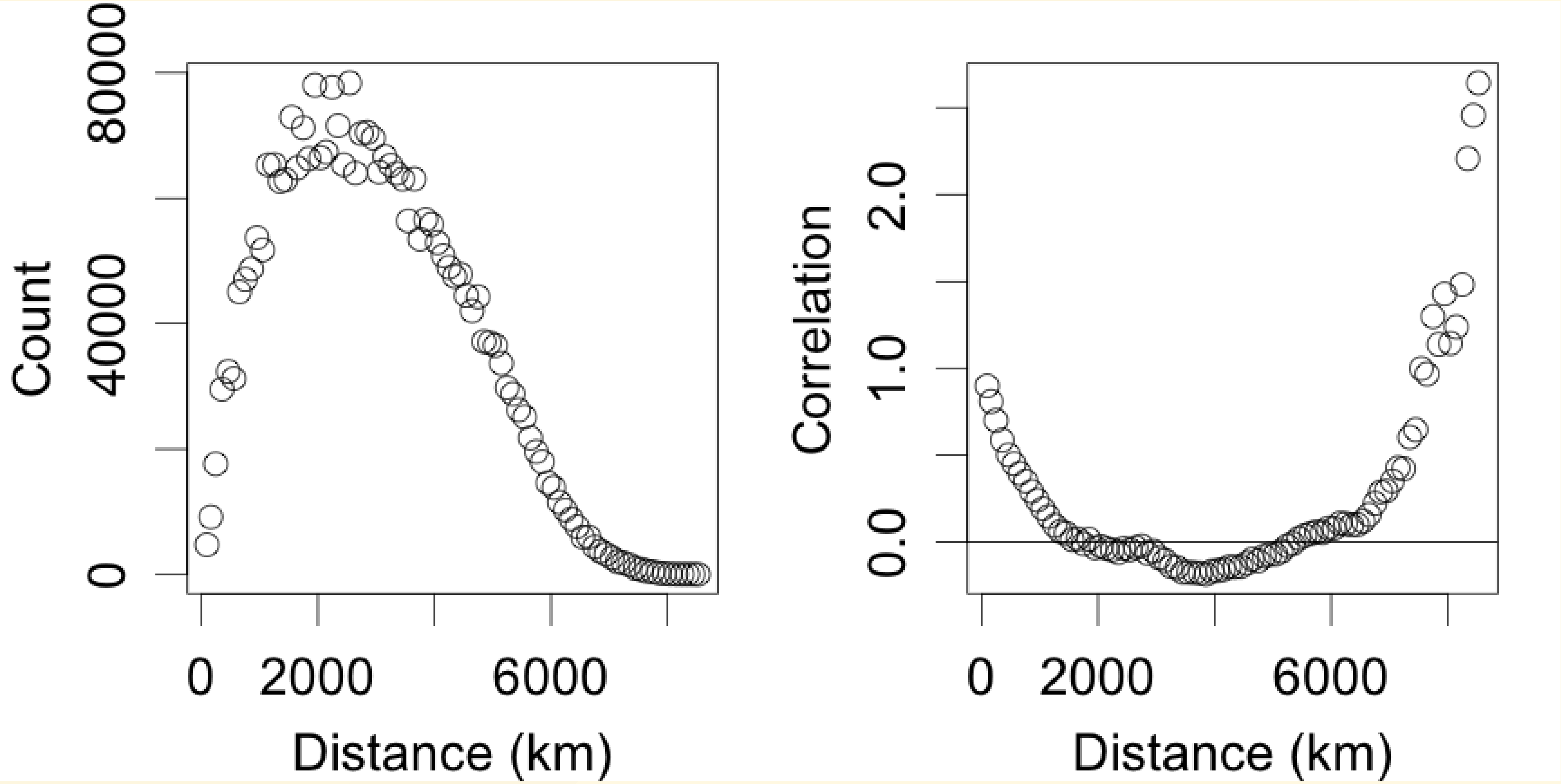
# Spatial Autoregresssion



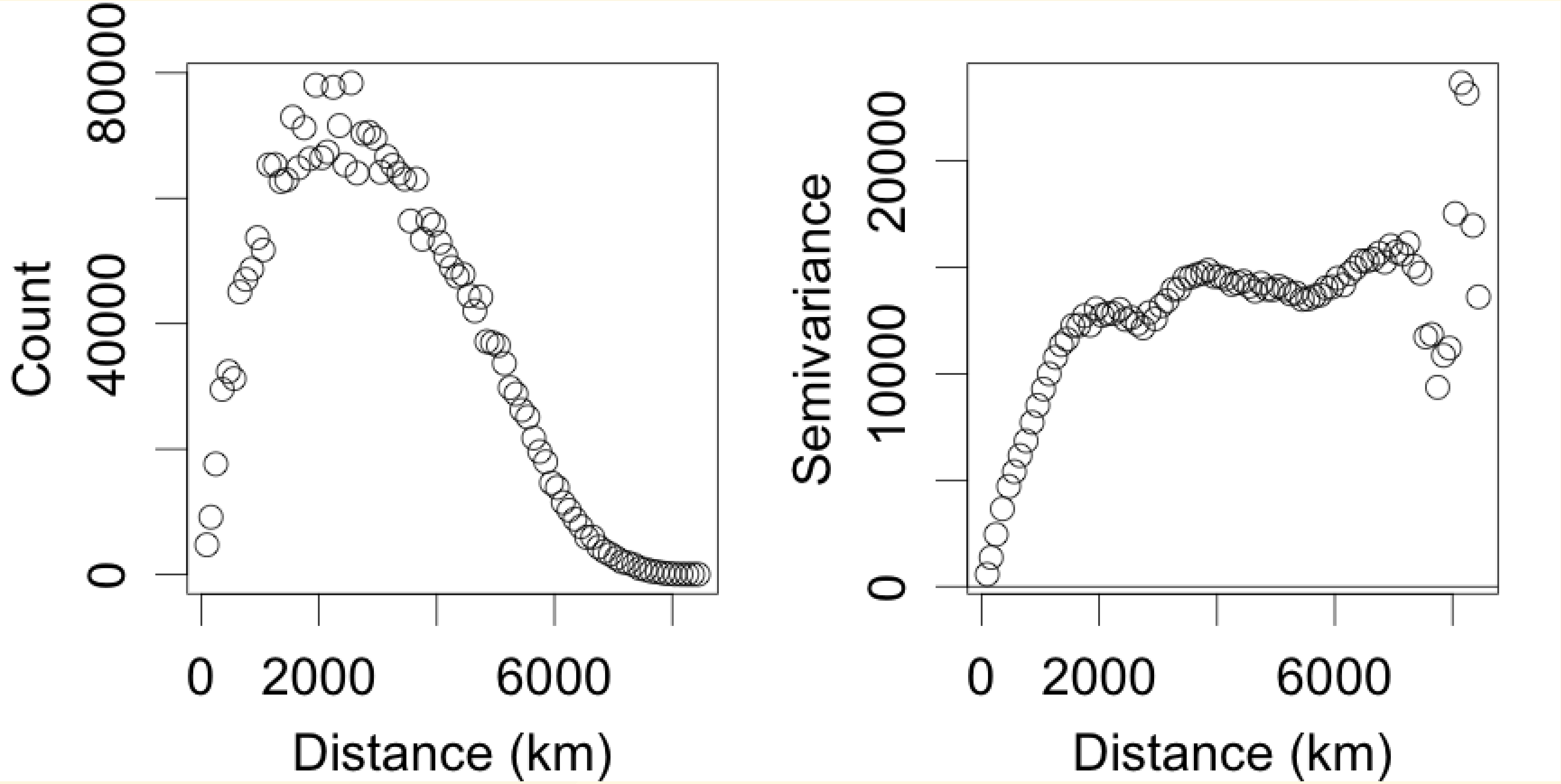
# Spatial Autoregresssion

```
Call:lagsarlm(formula = Rich ~ MeanAETScaled + MeanAnnTempScaled +
    MeanElevScaled, data = figDat@data, listw = figDat.lw, type = "lag")

Residuals:
        Min          1Q      Median          3Q         Max
-266.301618   -9.585387   -0.050679    9.997008  105.697309

Type: lag
Coefficients: (asymptotic standard errors)
                  Estimate Std. Error z value  Pr(>|z|)
(Intercept)        3.03299    0.70394  4.3086 1.643e-05
MeanAETScaled      2.84545    0.48197  5.9037 3.553e-09
MeanAnnTempScaled  1.34439    0.65707  2.0460   0.04075
MeanElevScaled     5.77947    0.69995  8.2570 2.220e-16

Rho: 0.98408, LR test value: 6043.7, p-value: < 2.22e-16
Asymptotic standard error: 0.0025867
    z-value: 380.44, p-value: < 2.22e-16
Wald statistic: 144730, p-value: < 2.22e-16

Log likelihood: -11459.21 for lag model
ML residual variance (sigma squared): 462.9, (sigma: 21.515)
Number of observations: 2484
Number of parameters estimated: 6
AIC: 22930, (AIC for lm: 28972)
LM test for residual autocorrelation
test value: 497.74, p-value: < 2.22e-16
```

# Correlogram

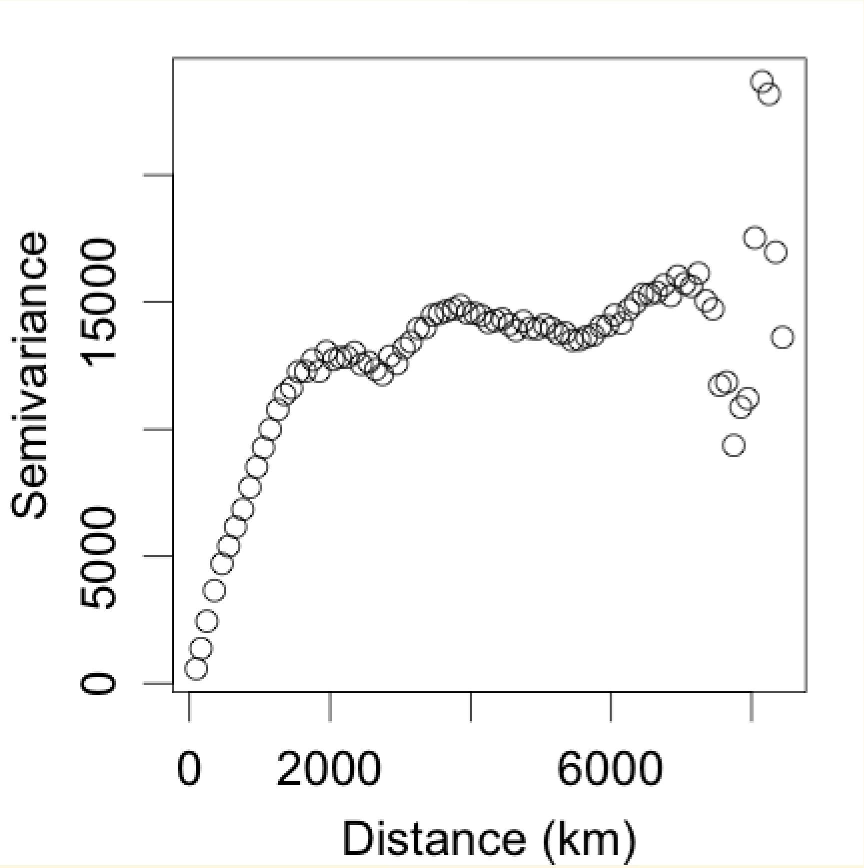


# Variogram



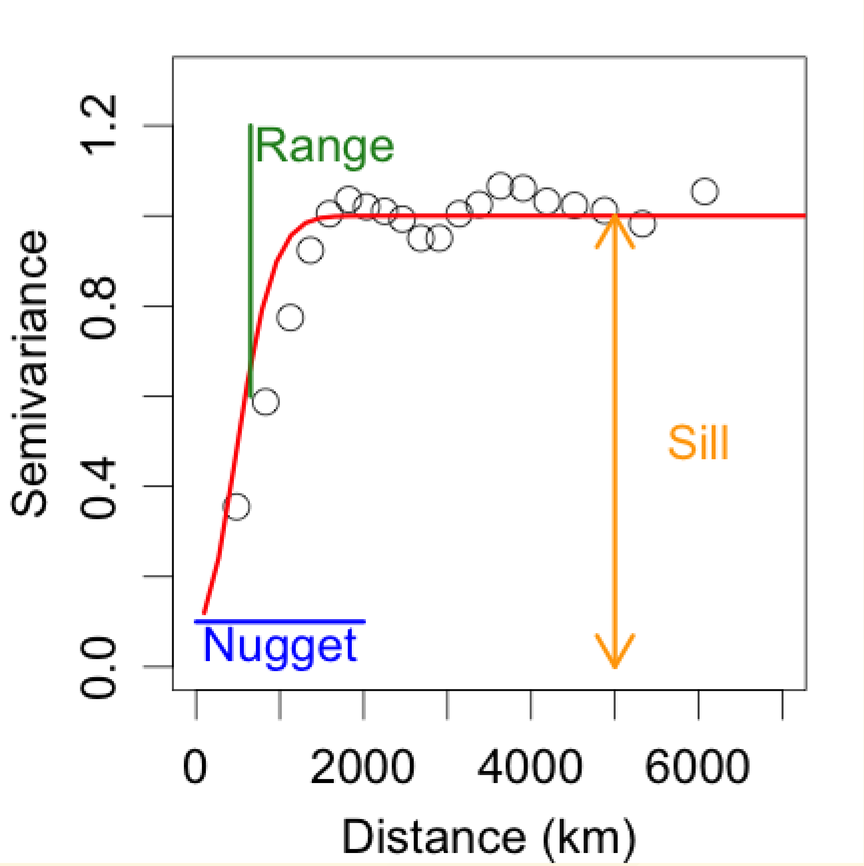
Count
0
40000
80000
0
2000
6000
Distance (km)
Semivariance
0
10000
20000
0
2000
6000
Distance (km)

# Generalised Least Squares




- Model variance as a function of **distance**
- Generate a covariance **matrix**

# Generalised Least Squares




- Different shapes:
    - Exponential
    - Spherical
    - Linear
- Parameters

# Generalised Least Squares

```
Generalized least squares fit by REML
  Model: Rich ~ MeanAETScaled + MeanAnnTempScaled + MeanElevScaled
  Data: figDat
       AIC      BIC    logLik
  24676.89 24705.97 -12333.44

Correlation Structure: Gaussian spatial correlation
 Formula: ~e_centre_behr + n_centre_behr
 Parameter estimate(s):
 range nugget
 650.0    0.1

Coefficients:
                      Value Std.Error   t-value p-value
(Intercept)       199.67323 16.755430 11.916927  0.0000
MeanAETScaled      17.65050  3.038337  5.809265  0.0000
MeanAnnTempScaled -27.53775  5.663121 -4.862645  0.0000
MeanElevScaled      3.59893  4.321932  0.832712  0.4051

 Correlation:
                  (Intr) MnAETS MnAnTS
MeanAETScaled     0.048
MeanAnnTempScaled 0.141  0.047
MeanElevScaled    0.156  0.079  0.936

Standardized residuals:
        Min          Q1         Med          Q3         Max
-2.35080108  0.09376345  0.73813223  1.30659392  4.48588628

Residual standard error: 97.85917
Degrees of freedom: 2484 total; 2480 residual
```

# Stationarity and isotropy

Is the same process happening in:

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

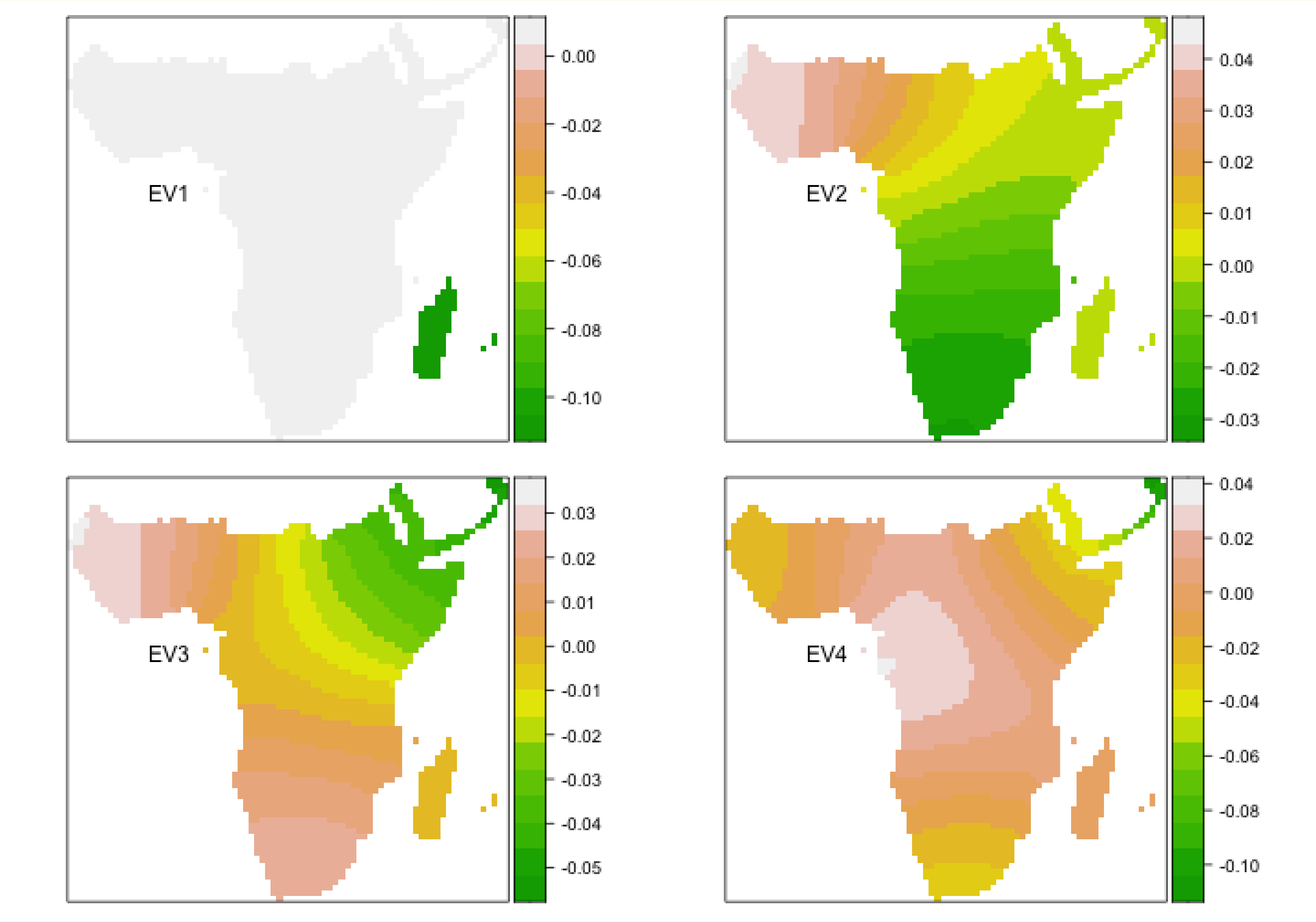
Is the problem in:

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

# Eigenvector filtering

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

# Eigenvector filtering




- First four eigenvector filters
- Independent components of spatial patterning

# Eigenvector filtering

```
lm(Rich ~ MeanAET + MeanAnnTemp + MeanElev
```

| | Est | SE | t | p |
|---|---|---|---|---|
| (Intercept) | 189.45 | 21.33 | 8.88 | < 0.001 |
| MeanAET | 0.18 | 0.00 | 37.34 | < 0.001 |
| MeanAnnTemp | -4.18 | 0.72 | -5.79 | < 0.001 |
| MeanElev | 0.08 | 0.01 | 13.85 | < 0.001 |

# Eigenvector filtering

```
lm(Rich ~ ... + Re(spEV1) + Re(spEV2) + Re(spEV3) + Re(spEV4)
```

| | Est | SE | t | p |
|---|---|---|---|---|
| (Intercept) | 79.95 | 33.00 | 2.42 | 1.5e-02 |
| MeanAET | 0.18 | 0.01 | 31.45 | < 0.001 |
| MeanAnnTemp | 0.11 | 1.14 | 0.09 | 9.3e-01 |
| MeanElev | 0.08 | 0.01 | 12.71 | < 0.001 |
| Re(spEV1) | 1617.43 | 77.65 | 20.83 | < 0.001 |
| Re(spEV2) | -964.81 | 129.21 | -7.47 | < 0.001 |
| Re(spEV3) | 813.41 | 95.80 | 8.49 | < 0.001 |
| Re(spEV4) | 147.25 | 100.29 | 1.47 | 1.4e-01 |

# Eigenvector filtering

```
lm(Rich ~ ... + Re(spEV1) + Re(spEV2) + Re(spEV3)
```

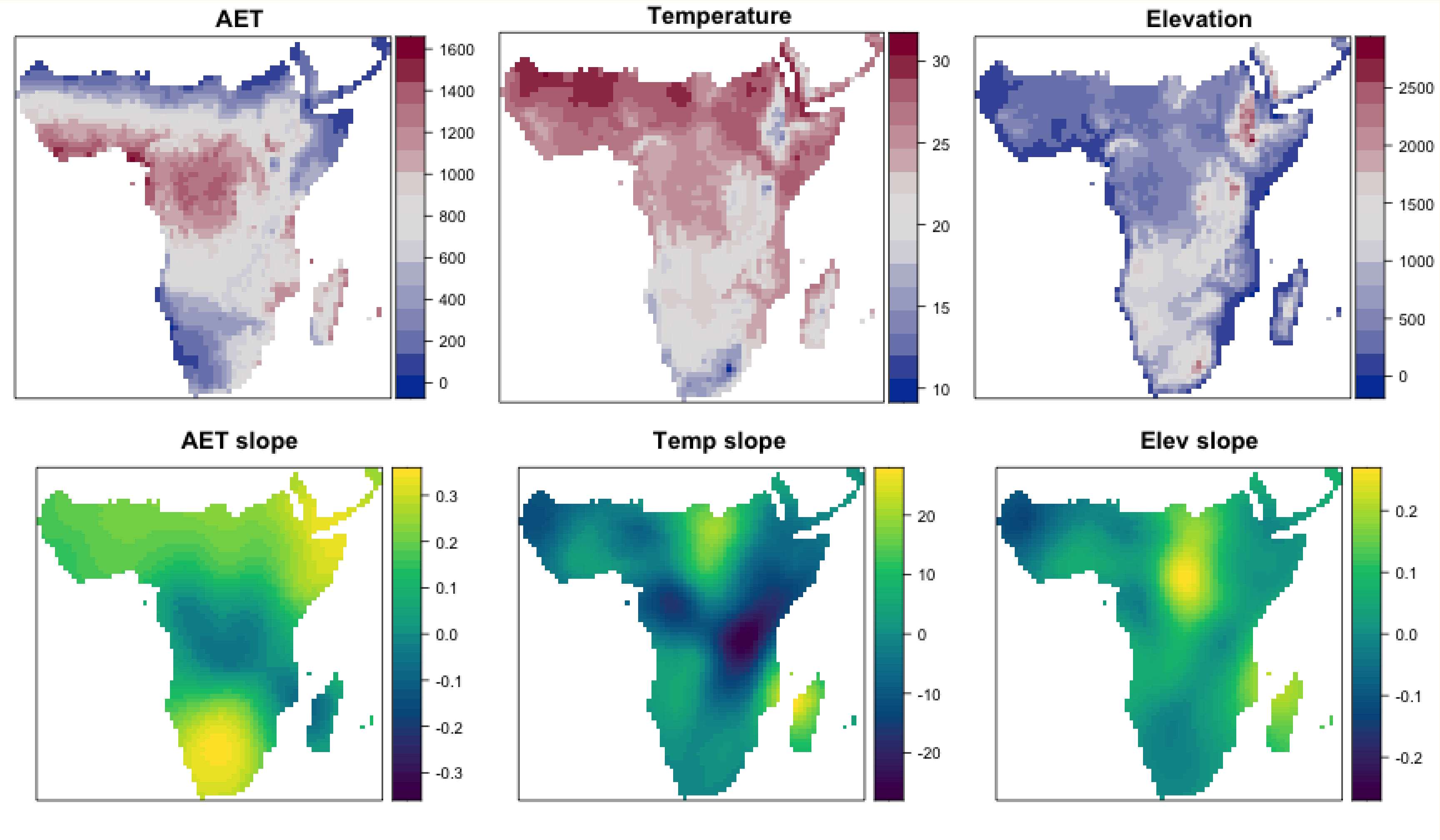
| | Est | SE | t | p |
|---|---|---|---|---|
| (Intercept) | 58.55 | 29.62 | 1.98 | 4.8e-02 |
| MeanAET | 0.19 | 0.00 | 43.67 | < 0.001 |
| MeanAnnTemp | 0.74 | 1.06 | 0.70 | 4.8e-01 |
| MeanElev | 0.08 | 0.01 | 13.78 | < 0.001 |
| Re(spEV1) | 1610.70 | 77.53 | 20.78 | < 0.001 |
| Re(spEV2) | -1031.06 | 121.10 | -8.51 | < 0.001 |
| Re(spEV3) | 847.10 | 93.03 | 9.11 | < 0.001 |

# Geographically weighted regression

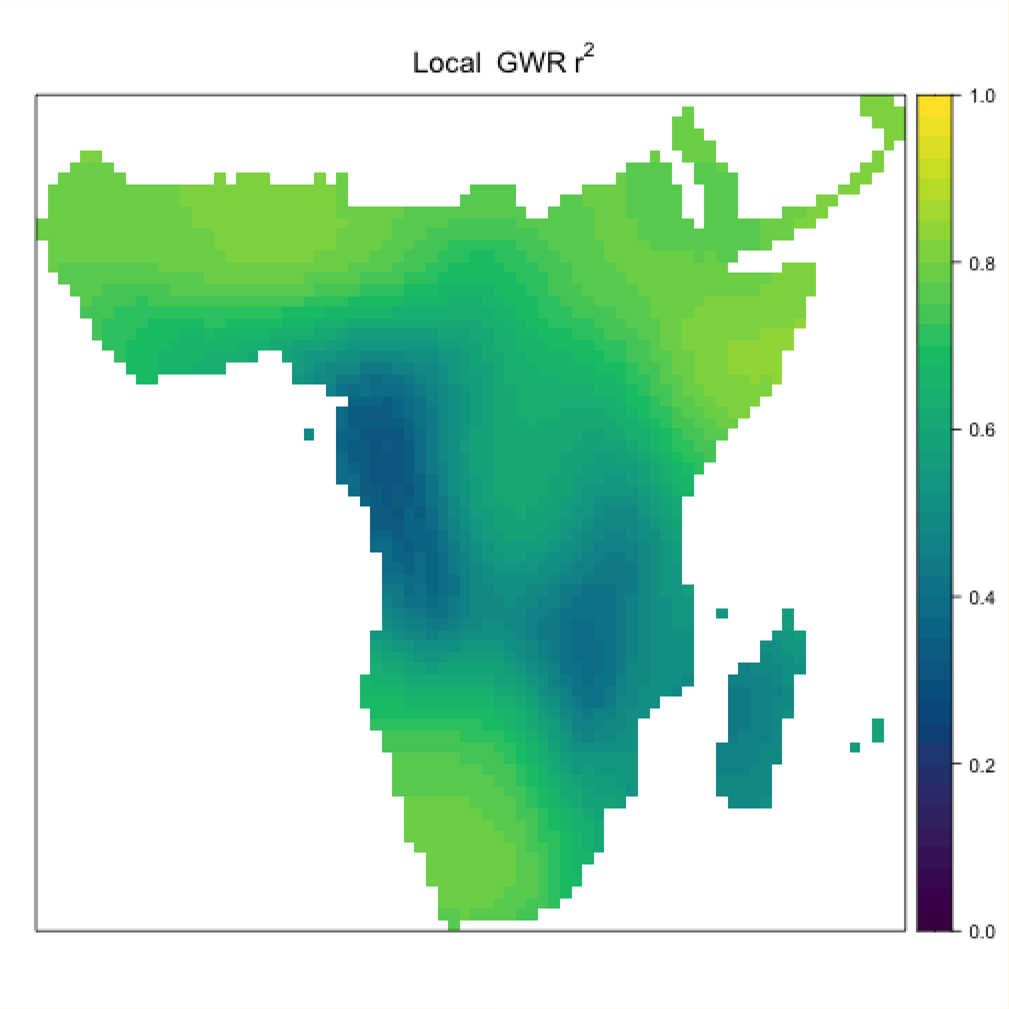
Fit a model for **every cell**:

- Define a local **region size** and a **weighting function**
- Fit a weighted regression for each cell using the weights
- Look at how coefficients **vary in space**
- Possibly serious statistical issues!

# Geographically weighted regression



# Geographically weighted regression




- Local $r^2$ values
- Variation in the explanatory power of the model

# Problems

- Profusion of packages: sf, sp, spdep, mgcv, ncf, gstat, nlme, spgwr
- Different data structures
- Sometimes poor documentation
- Speed of calculation (= size of dataset)
- Memory hungry
- Too many options