

# Spatial Linear Models

Helene Wagner, University of Toronto

## Goal:

- Which connectivity model (IBD or IBR) better explains genetic diversity?
- Can the same pattern be explained by patch size as a proxy for population size?
- Do the same factors explain genetic diversity across the study area?

## Methodological Challenges:

### Video 1: Neighbourhood-level analysis

- Define connectivity hypotheses as distance matrices
- Neighborhood approach: connectivity index SI
- Explain genetic diversity by patch connectivity Si

### Video 2: Autocorrelation in linear models

- Generalized Least Squares regression (GLS)
- Spatial Regression with SAR
- Spatial Filtering with MEM
- Model selection (see also Week 12)
- Geographically-weighted Regression with SVC



Dianthus carthusianorum  
([www.schmetterlingswiesen.de](http://www.schmetterlingswiesen.de))

Refreshers: regression (Week 4), spatial statistics (Week 5), linear mixed models (Week 6)

# Dianthus Dataset

Rico et al. 2013, Molecular Ecology. See dataset video by Yessica Rico!

## Basic Dataset

Patches: 106  
Populations: 65  
Individuals: 1602  
Microsatellites: 11

Population size:  
< 4, < 40, < 100, 100+

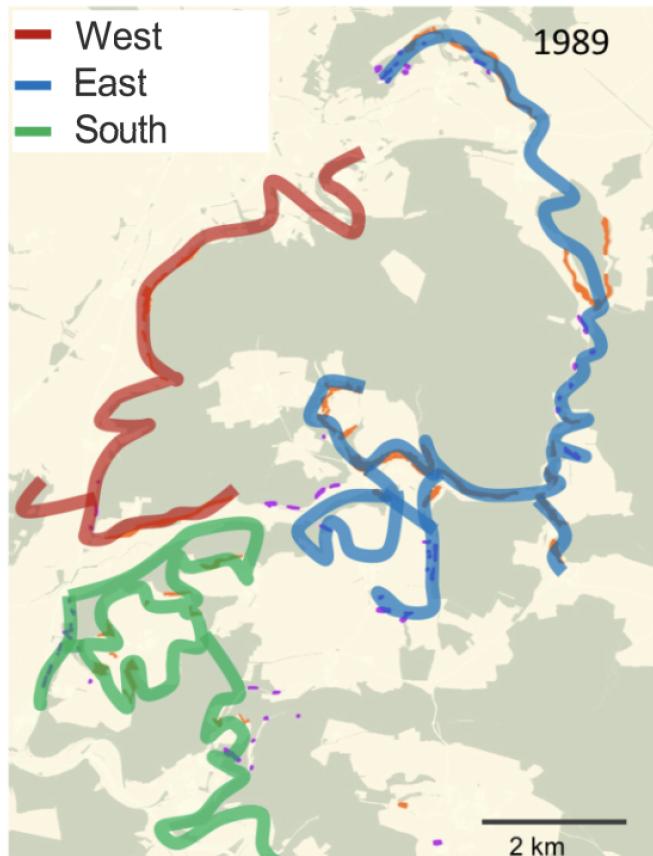
Patch size: ha

Grazing intensity:  
- ungrazed  
- intermittently  
- consistently



[www.neuburg-schrobenhausen.de](http://www.neuburg-schrobenhausen.de)

## Connectivity by Shepherding



Figures and map: Y. Rico

## Alternative Hyptheses

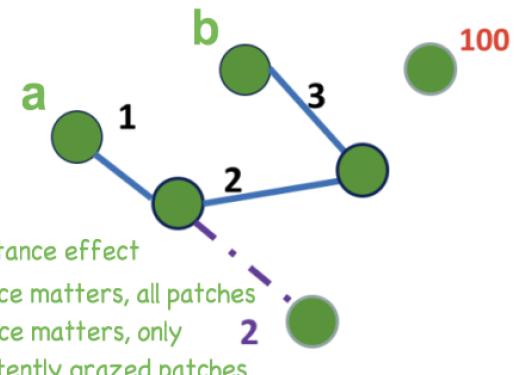
Geographic distance (IBD)



Forest as barrier (IBB)



Connectivity by shepherding (IBR)



- No distance effect
- Distance matters, all patches
- Distance matters, only consistently grazed patches

# Connectivity Index Si

See Week 7 Bonus Material

Connectivity Si  
of focal patch i

Sum over all neighbours j

$$Si = \sum_{j \neq i} \exp(-\alpha d_{ij}) p_j$$

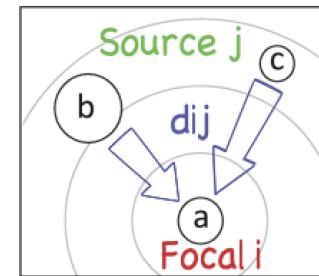
Source patch:  
 $p_j, A_j, N_j$

Connectivity model:  

- ecological distance  $d_{ij}$
- dispersal parameter alpha

$$Y \sim \sum_j X_j$$

C Neighborhood level

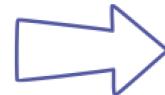


Wagner and Fortin 2013

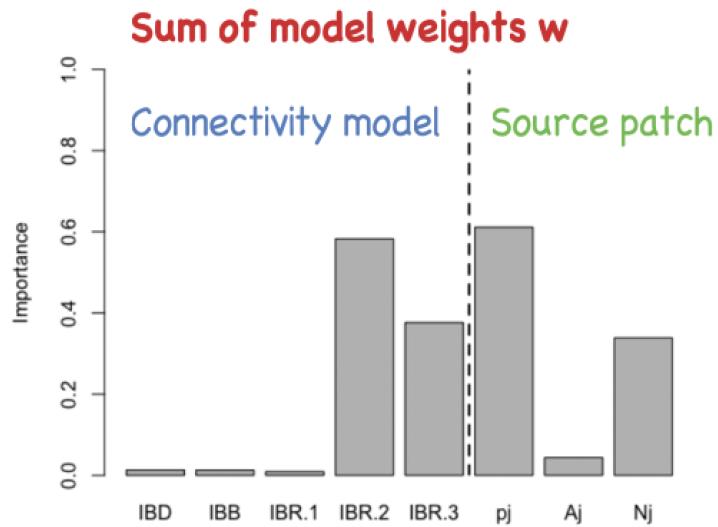
## Correlation with Allelic Richness

	$p_j$	$A_j$	$N_j$
IBD	0.02	0.21	0.10
IBB	0.14	0.17	0.14
IBR.1	0.13	-0.05	0.11
IBR.2	0.40	0.20	0.37
IBR.3	0.37	0.27	0.37

MuMIn:  
Multi-Model Inference

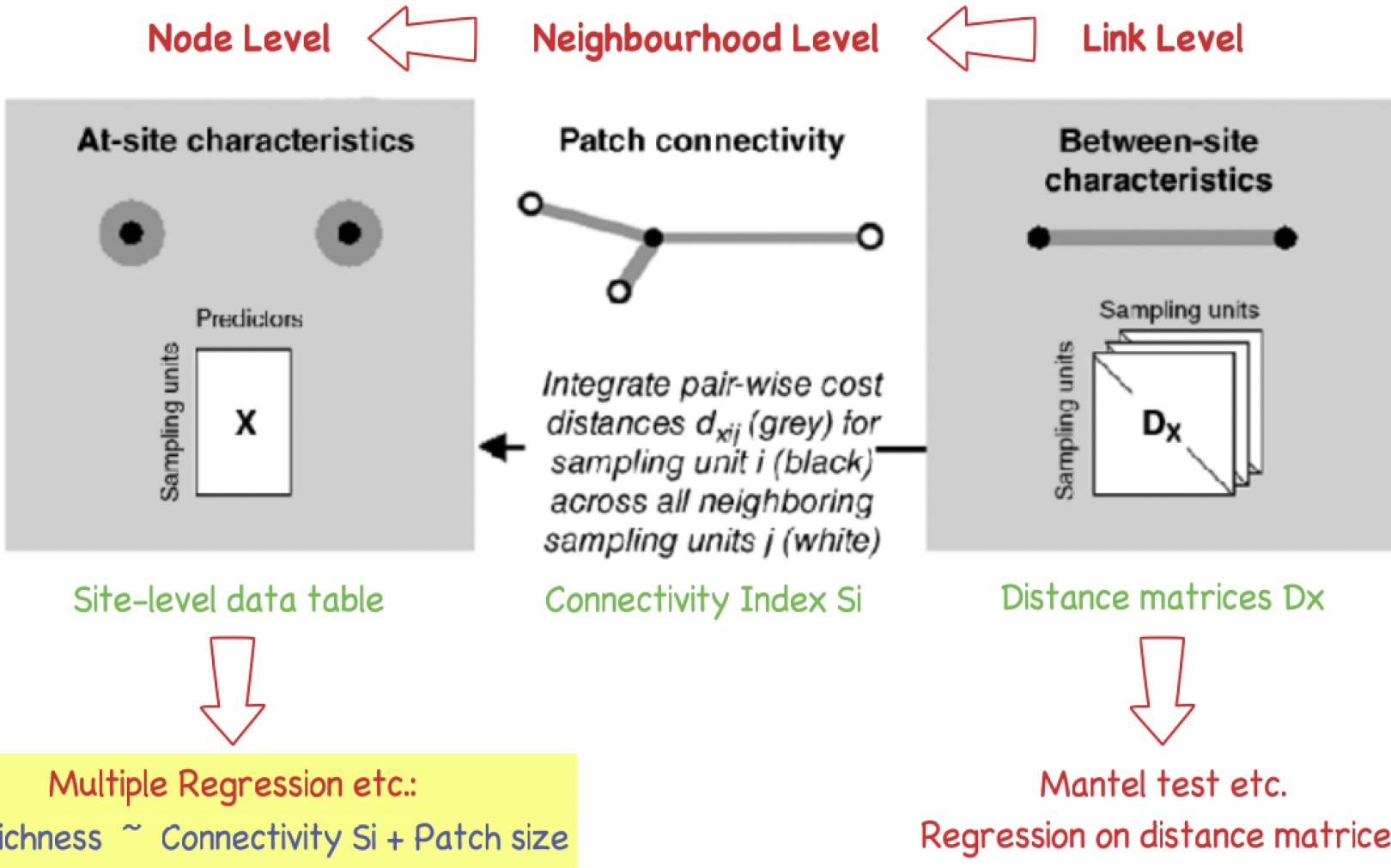


Calculate AIC  
Compare models



# Back to Regression

B: Landscape data



# Generalized Least Squares (GLS)

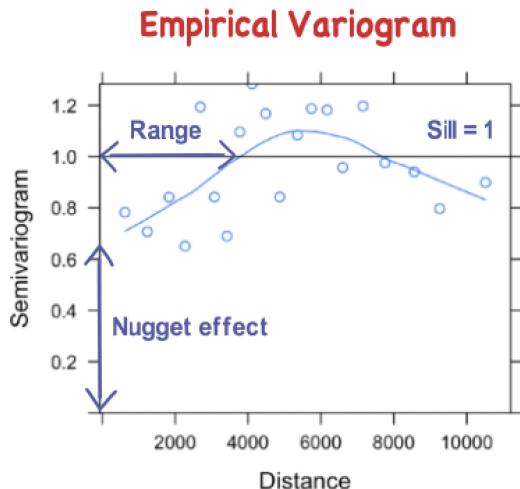
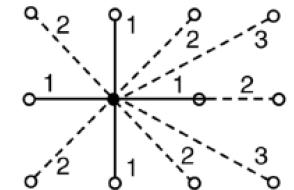
$$Y = \text{intercept} + b * X + \text{residuals}$$

```
model.lm = lm ( Y ~ X, data)
```

```
lm.morantest ( residuals ( model.lm ) )
```

Direction: Sign of Moran's I?  
Size: Value of Moran's I?  
Significance: P-value of Moran's I?

Geostatistical model



Fit Theoretical Variogram

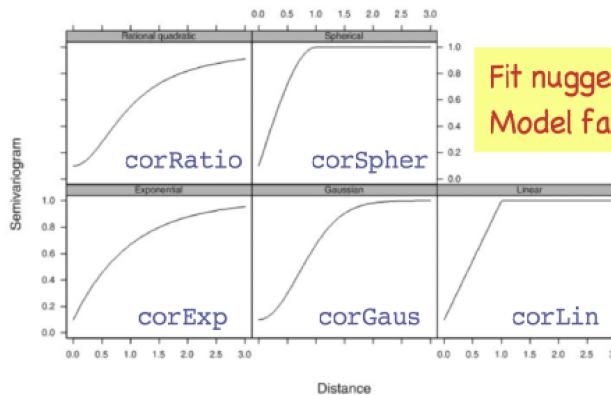


Figure: Pinheiro and Bates 2004, p. 233

```
Variogram( model.lm, form = ~ x + y )
```

```
gls( formula, data, correlation = corExp( form = ~ x + y, nugget=T ) )
```

```
model.sel ( model.lm, mod.corRatio, mod.corSper, mod.corExp, mod.corGaus, mod.corLin )
```

# Which Model Fits Best?

See also Week 12

Ranked models  
(best to worst)

Models penalized  
for df used

Criterion  
used: AICc

delta =  
AICc - AICc(best)

mod.corExp  
mod.corRatio  
model.lm  
mod.corGaus  
mod.corSpher

	(Intrc)	IBR	PtchS	correlation	df	logLik
mod.corExp	4.044	0.08468	0.04269	n:::cE(x+y,T)	6	6.283
mod.corRatio	4.032	0.09704	0.04129	n:::cR(x+y,T)	6	6.229
model.lm	4.052	0.11340	0.04266		4	3.702
mod.corGaus	4.047	0.11840	0.04247	n:::cG(x+y,T)	6	3.763
mod.corSpher	4.048	0.11710	0.04242	n:::cS(x+y,T)	6	3.760

AICc	delta	weight
1.1	0.00	0.336
1.2	0.11	0.319
1.3	0.29	0.291
6.1	5.04	0.027
6.1	5.04	0.027

Do estimates vary  
between models?

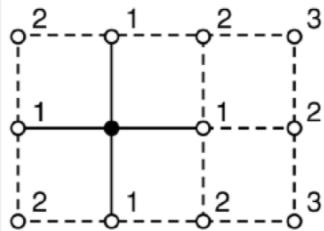
Model weight:  
Support, given  
all models in set

# Spatial Regression with SAR

Simultaneous  
Auto-  
Regressive

SAR

Weight  $w_{ij}$



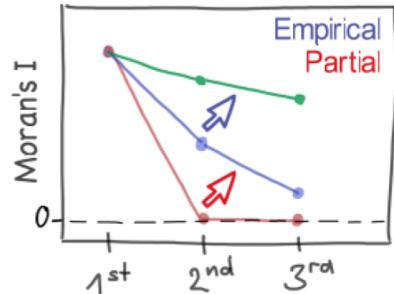
Error covariance =  $f(w_{ij})$

	Ind 1	Ind 2	Ind 3	Ind 4
Ind 1	1	0.7	0.2	0
Ind 2	0.7	1	0.3	0.1
Ind 3	0.2	0.3	1	0.8
Ind 4	0	0.1	0.8	1

Autocorrelation Function (ACF)



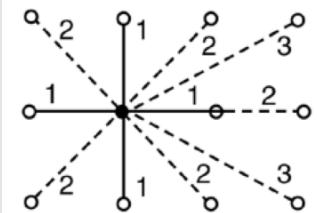
Weights reflect  
partial ACF



Variogram models  
empirical ACF

GLS

Distance  $d_{ij}$

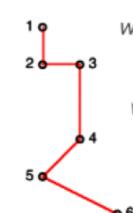


Error covariance =  $f(d_{ij})$

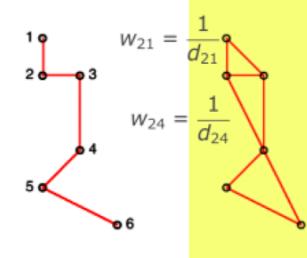
	Ind 1	Ind 2	Ind 3	Ind 4
Ind 1	1	0.7	0.2	0
Ind 2	0.7	1	0.3	0.1
Ind 3	0.2	0.3	1	0.8
Ind 4	0	0.1	0.8	1

How to Choose Neighbours and Weights?

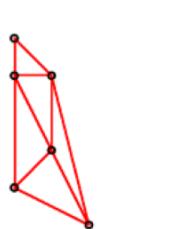
Minimum  
Spanning Tree



Gabriel Graph



Delauney  
Triangulation



All Pairs



$$\sum_j w_{ij} = 1$$

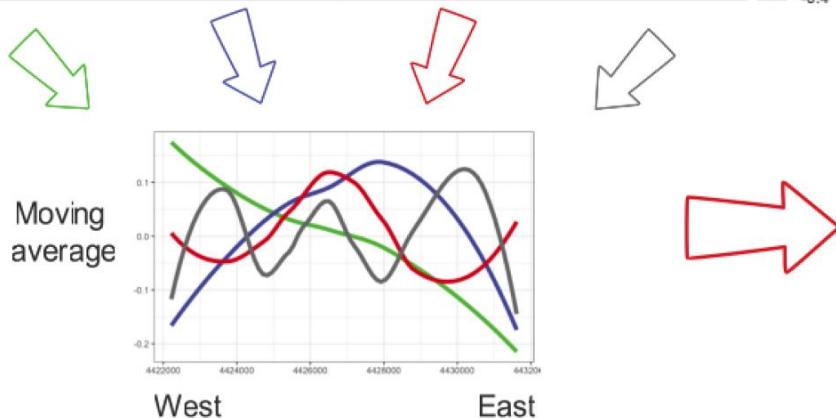
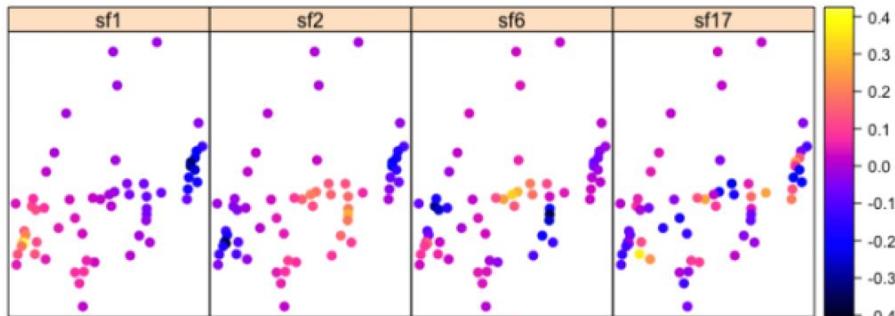
```
errorsarlm( formula, data, listw )
model.sel( model.lm, sar.Bin, sar.Inv.d1, sar.Inv.d2 )
```

# Spatial Filtering with MEM

Moran  
Eigenvector  
Maps

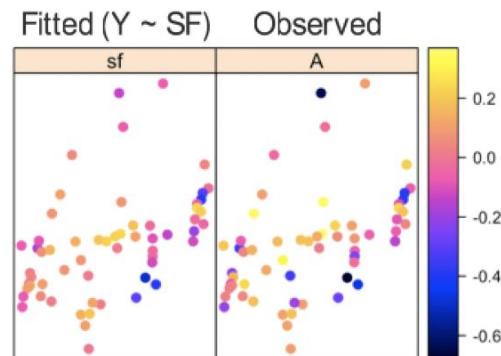
Goal: control for spatial variation when assessing ' $\text{Im} (Y \sim X)$ '

$$Y = \text{intercept} + b.x * X + b.sf * SF + \text{residuals}$$



```
meig <- spmoran :: meigen ( coords, cmat )
spmoran :: esf ( y, x, meig, fn = "r2" )
spmoran :: resf ( y, x, meig, fn = "r2" )
```

1. Define spatial weights matrix
2. Extract spatial eigenvectors (MEM)
3. Stepwise selection of significant MEM
4. Use as additional predictors SF
5. Effect of predictors X assessed from remaining non-spatial variation in Y



cmat: 'connectivity matrix' = spatial weights  
meig: spatial eigenvectors, eigenvalues  
fn: method for stepwise selection  
'esf': fixed effects, 'resf': random effects (REML)

# Which Method to Choose?

What to build into the null model?

Isolation by Distance  
(IBD)

Any spatial structure  
(IBD, IBB, IBR)

Gene flow model?

Spatially continuous

Generalized Least  
Squares GLS

Stepping stone model

Spatial Regression  
with SAR

E.g., individual sampling  
in gradient landscape

E.g., metapopulation

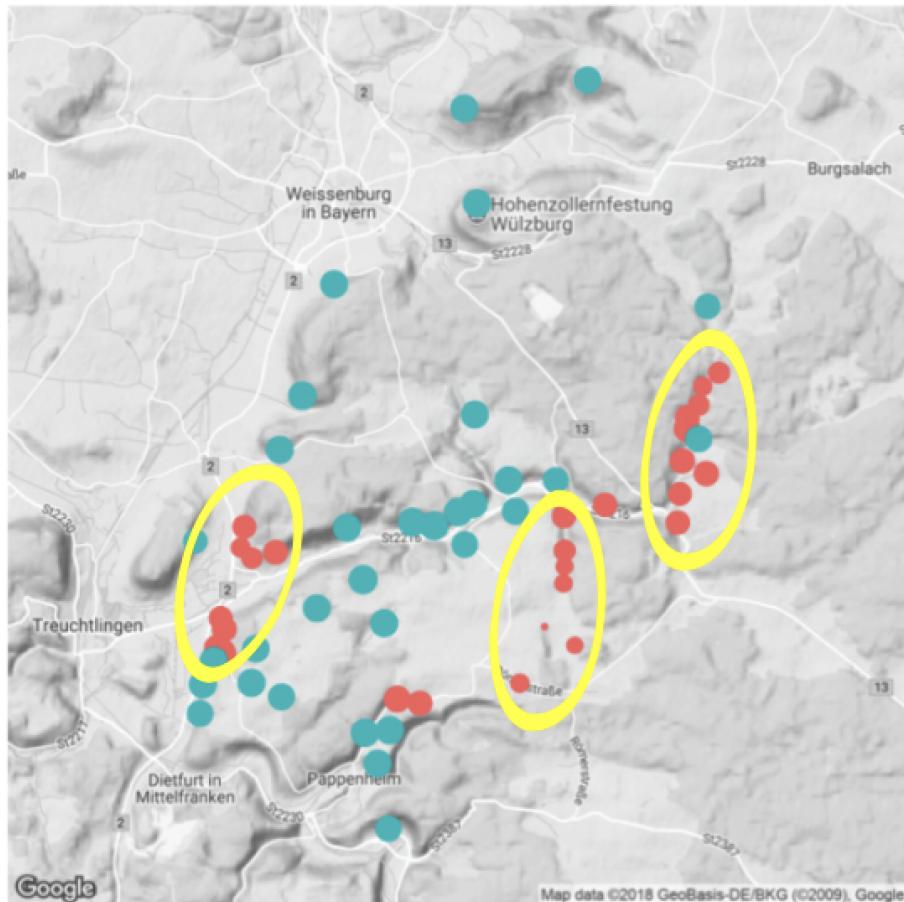
Spatial Filtering  
with MEM

E.g., testing genotype -  
environment association

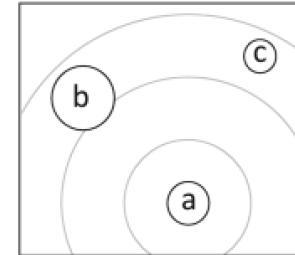
# Spatially Varying Coefficients

Similar to Geographically Weighted Regression (GWR)

Does relationship hold across study area?



C Neighborhood level



$$Y \sim \sum_j X_j$$

Slope estimate for Si.IBR varies across study area

Statistical significance varies across study area

(Intercept)	V1 = Si.IBR
Min. :3.801	Min. :-0.2020
1st Qu.:3.919	1st Qu.: 0.1545
Median :3.941	Median : 0.2549
Mean :3.941	Mean : 0.2101
3rd Qu.:3.969	3rd Qu.: 0.3113
Max. :4.060	Max. : 0.4180

```
spmoran :: resf_vc( y, x, xconst = NULL, meig, method = "reml" )
```