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?

Dianthus carthusianorum
(www.schmetterlingswiesen.de)

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

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

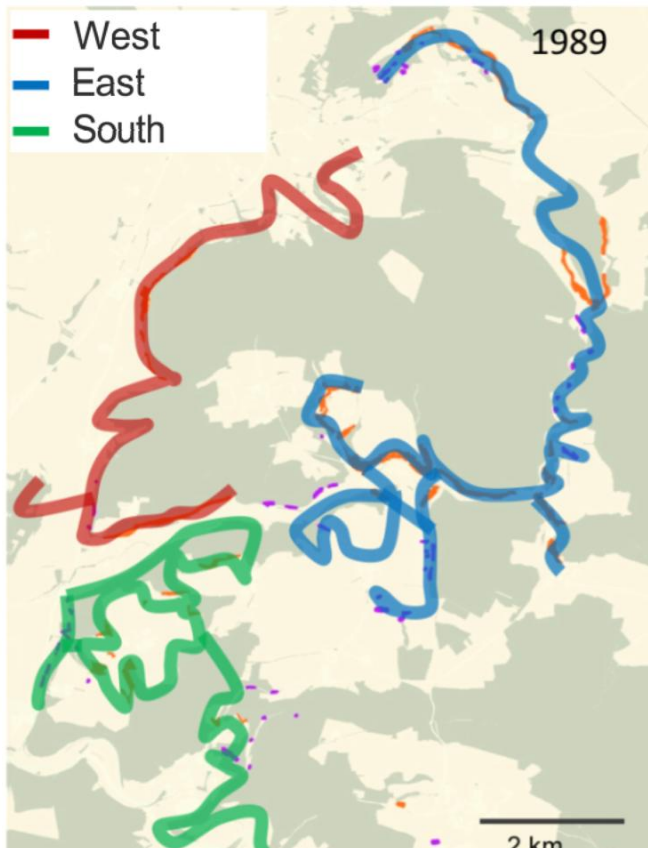
Population size:
< 4, < 40, < 100, 100+

Patch size: ha

Grazing intensity:
- ungrazed
- intermittently
- consistently

www.neuburg-schrobenhausen.de

Connectivity by Shepherding



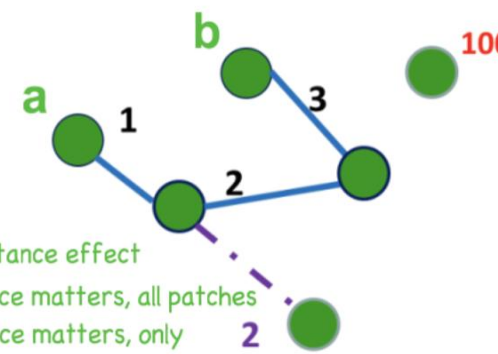
Figures and map: Y. Rico

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

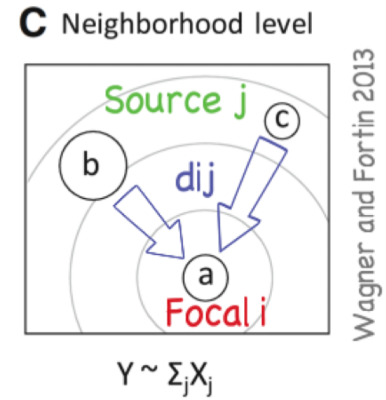
Sum over all neighbours j

Connectivity Si of focal patch i

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

Source patch: pj, Aj, Nj

Connectivity model:
- ecological distance dij
- dispersal parameter alpha



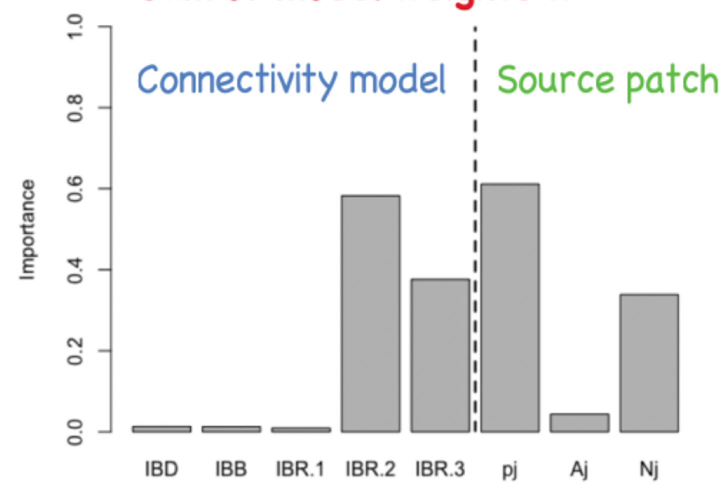
Correlation wit Allelic Richness

	pj	Aj	Nj
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

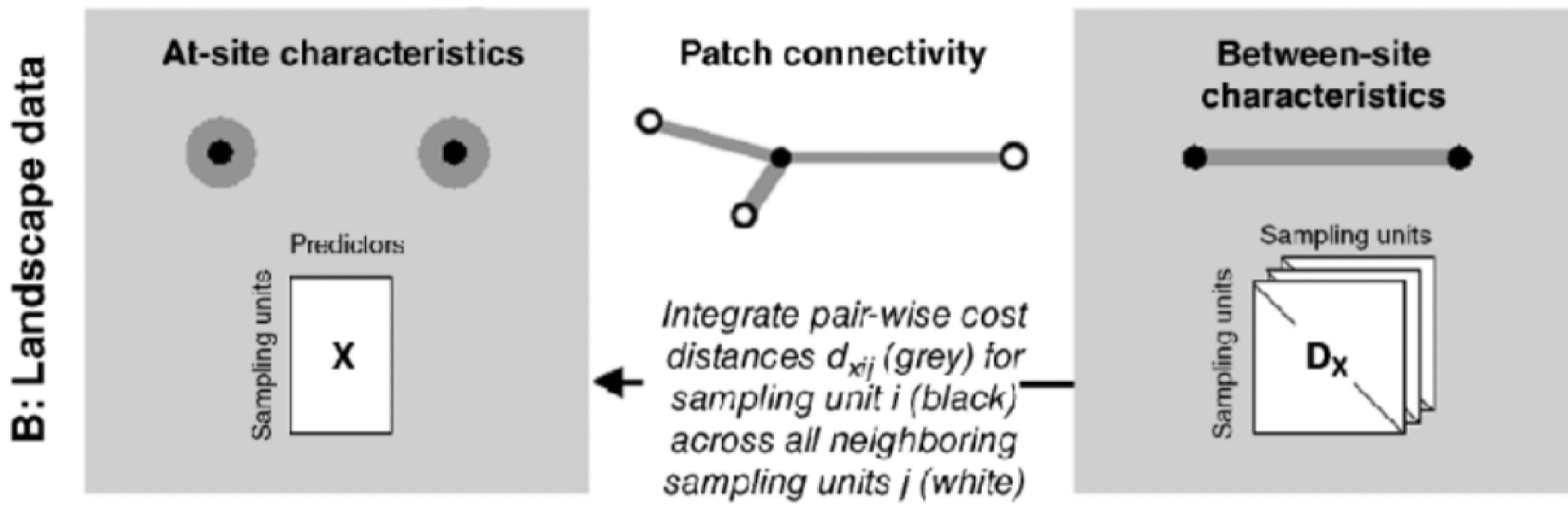
Calculate AIC
Compare models

Sum of model weights w



Back to Regression

Node Level ⇐ **Neighbourhood Level** ⇐ **Link Level**



Wagner and Fortin 2015

Site-level data table | Connectivity Index Si | Distance matrices Dx

⇓ Multiple Regression etc.:
Allelic richness ~ Connectivity Si + Patch size

⇓ Mantel test etc.
Regression on distance matrices

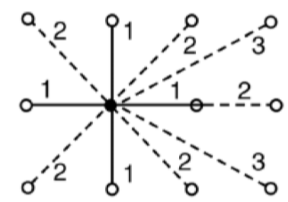
Generalized Least Squares (GLS)

Y = intercept + b * X + residuals

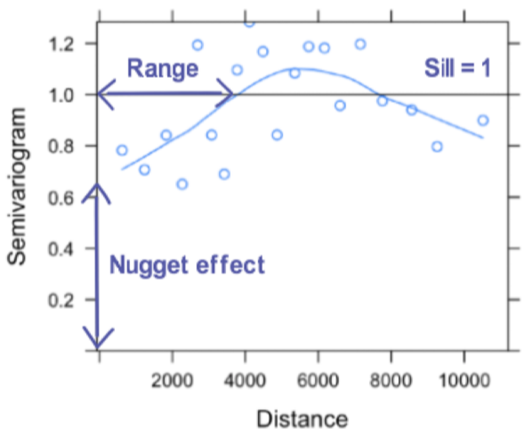
```
model.lm = lm ( Y ~ X, data)
lm.morantest ( residuals ( model.lm ))
```

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

Geostatistical model



Empirical Variogram



Fit Theoretical Variogram

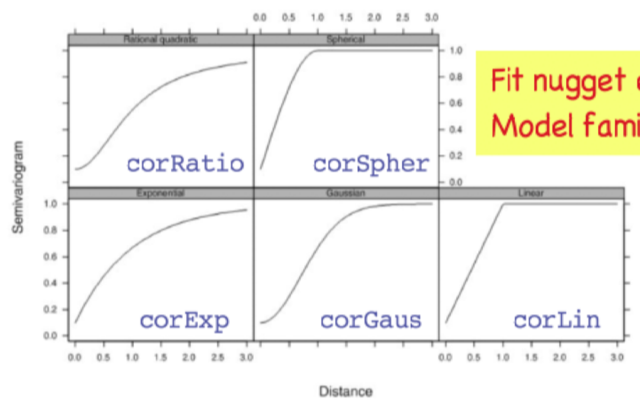


Figure: Pinheiro and Bates 2004, p. 233

Fit nugget effect?
Model family?

```
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)

	(Intrc)	IBR	PtchS	correlation	df	logLik	AICc	delta	weight
mod.corExp	4.044	0.08468	0.04269	n::cE(x+y,T)	6	6.283	1.1	0.00	0.336
mod.corRatio	4.032	0.09704	0.04129	n::cR(x+y,T)	6	6.229	1.2	0.11	0.319
model.lm	4.052	0.11340	0.04266		4	3.702	1.3	0.29	0.291
mod.corGaus	4.047	0.11840	0.04247	n::cG(x+y,T)	6	3.763	6.1	5.04	0.027
mod.corSpher	4.048	0.11710	0.04242	n::cS(x+y,T)	6	3.760	6.1	5.04	0.027

Do estimates vary between models?

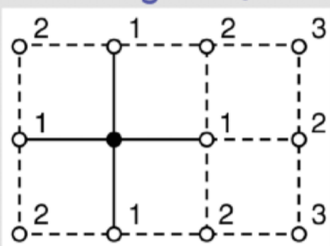
Model weight: Support, given all models in set

Spatial Regression with SAR

Simultaneous
Auto-
Regressive

SAR

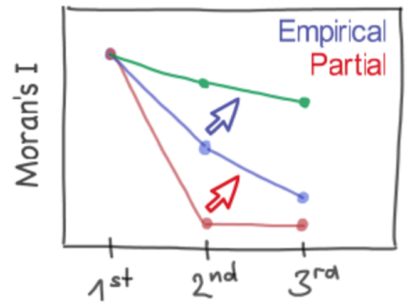
Weight wij



Error covariance = f(wij)

	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)

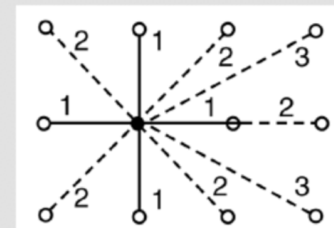


Weights reflect partial ACF

Variogram models empirical ACF

GLS

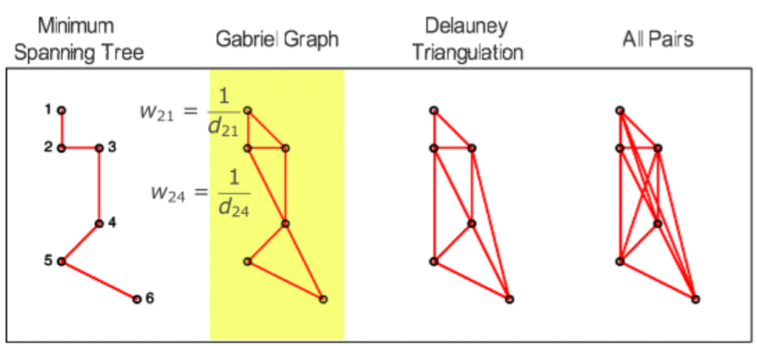
Distance dij



Error covariance = f(dij)

	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?



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

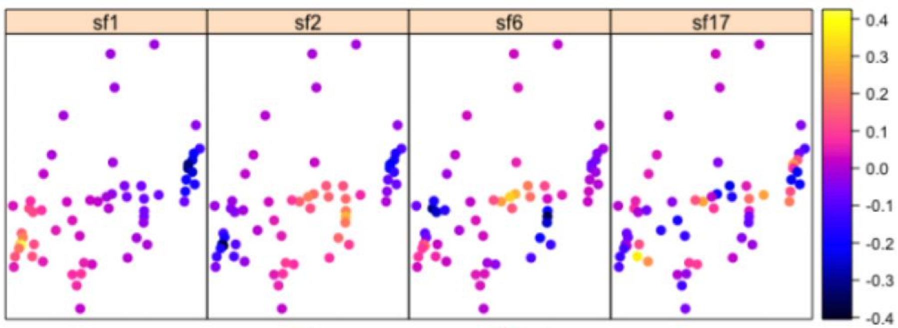
```
errorsarlm ( formula, data, listw ))
model.sel ( model.lm, sar.Bin, sar.Inv.d1, sar.Inv.d2 )
```

Spatial Filtering with MEM

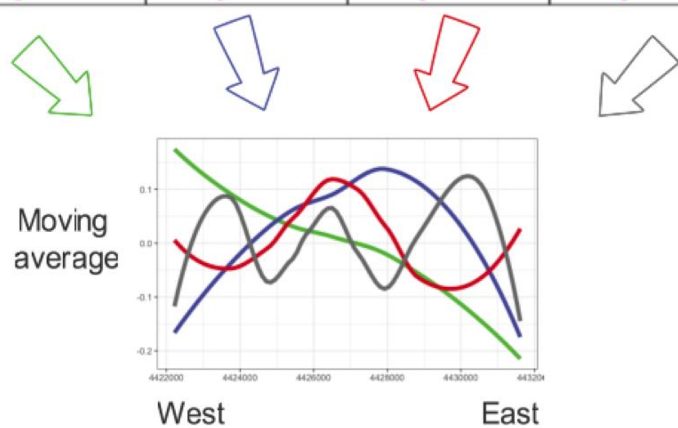
Moran
Eigenvector
Maps

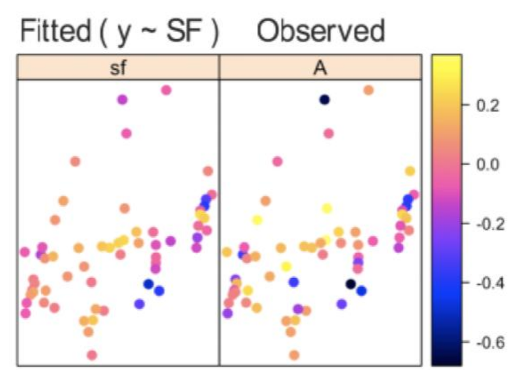
Goal: control for spatial variation when assessing 'lm (y ~ X)'

$$y = intercept + b_X * X + \boxed{b_{SF} * SF} + residual$$



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






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

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?

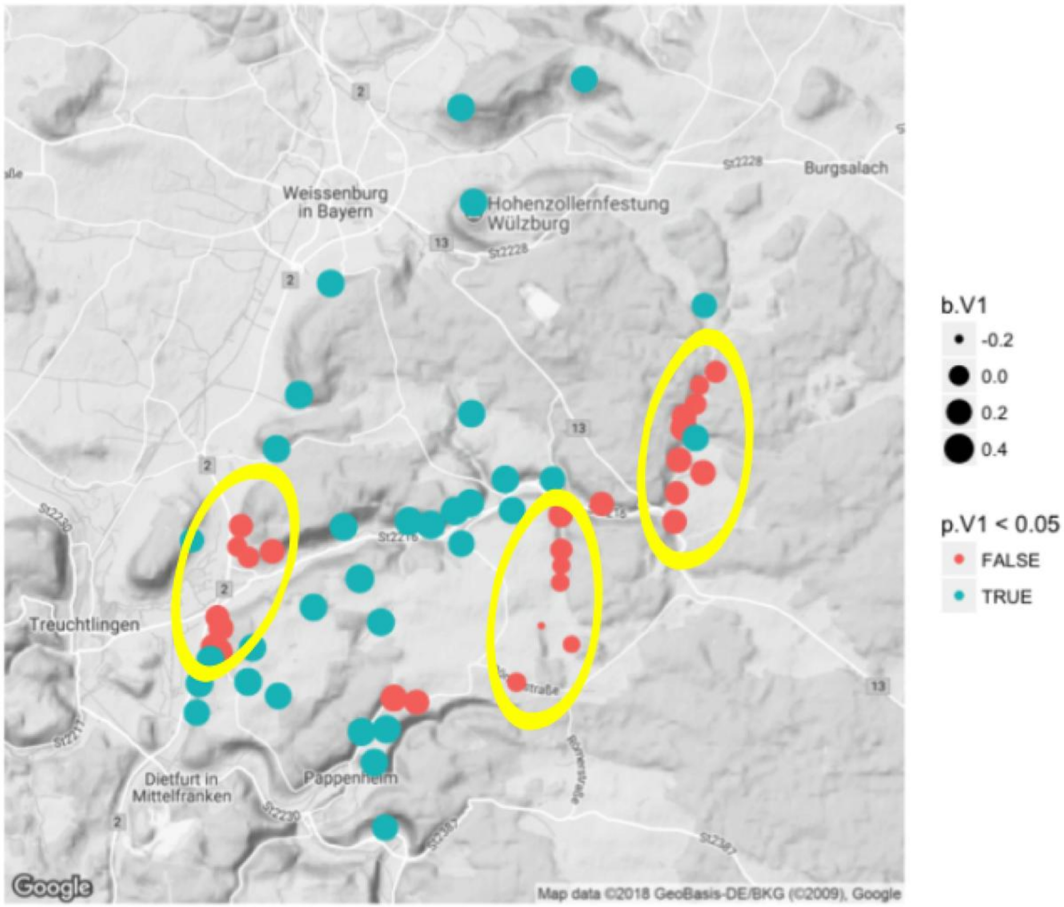
What to build into the null model?

- **Isolation by Distance (IBD)**
  - Gene flow model?
    - Spatially continuous
      - **Generalized Least Squares GLS**
      - E.g., individual sampling in gradient landscape
    - Stepping stone model
      - **Spatial Regression with SAR**
      - E.g., metapopulation
- **Any spatial structure (IBD, IBB, IBR)**
  - **Spatial Filtering with MEM**
  - E.g., testing genotype – environment association

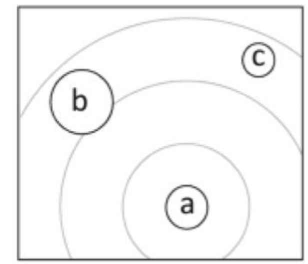
Spatially Varying Coefficients

Similar to Geographically Weighted Regression (GWR)

Does relationship hold across study area?



C Neighborhood level



$Y \sim \Sigma_j X_j$

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" )
```