

Selecting a model

Spatial Statistics

Madlene Nussbaum

m.nussbaum@uu.nl

Utrecht University, Department of Physical Geography

1 Inference and model building

Data analysis often leads to a set of equally plausible candidate models that use different set of covariates and different variograms

- compare fit of candidate models by hypothesis tests taking auto-correlation properly into account
- use established goodness-of-fit criteria (AIC, BIC) to select a “best” model, again taking auto-correlation into account
- use cross-validation to compare the power of candidate models to *predict new data* → chapter 25

Testing hypotheses about trend coefficients

- Likelihood ratio (LRT) test can only be used to test hypotheses and build confidence regions for θ
- LRT for regression for β in general biased (too small p -values)
- Use conditional F -tests for testing hypotheses about β :
 1. Fit covariance parameters of “largest” regression model
 $\Rightarrow \hat{\theta}$
 2. Compute covariance matrix $\Rightarrow \mathbf{\Gamma}_{\hat{\theta}}$
 3. Orthogonalize response vector and design matrix (using Cholesky decomposition)
 4. Conventional F -test with orthogonalized items $\tilde{\mathbf{Y}}$ and $\tilde{\mathbf{X}}$

Fit quadratic trend surface model for Wolfcamp

```
1 r.georob.full <- update(r.georob.reml, .~.+I(x^2)+I(y^2)+x:y)
2 summary(r.georob.full)
```

```
Call:georob(formula = pressure ~ x + y + I(x^2) + I(y^2) + x:y, data = d.w,
  locations = ~x + y, variogram.model = "RMspheric", param = c(variance = 3000,
    nugget = 1000, scale = 100), tuning.psi = 1000)
```

Tuning constant: 1000

Convergence in 10 function and 8 Jacobian/gradient evaluations

Estimating equations (gradient)

	eta	scale
Gradient	: 3.590344e-04	-4.553394e-03

Maximized restricted log-likelihood: -470.3894

Predicted latent variable (B):

Min	1Q	Median	3Q	Max
-89.22	-46.81	-11.06	20.80	94.07

Residuals (epsilon):

Min	1Q	Median	3Q	Max
-59.664	-18.086	6.783	16.245	49.986

Standardized residuals:

Min	1Q	Median	3Q	Max
-2.4096	-0.7213	0.2500	0.6830	3.1042

Conditional F-test on interaction and higher-order polynomials

```
1 waldtest(r.georob.full, .~.-x:y, test="F")
```

Wald test

Model 1: pressure ~ x + y + I(x^2) + I(y^2) + x:y

Model 2: pressure ~ x + y + I(x^2) + I(y^2)

	Res.Df	Df	F	Pr(>F)
1	79			
2	80	-1	1.1032	0.2968

```
1 waldtest(r.georob.full, .~.-I(x^2)-I(y^2)-x:y, test="F")
```

Wald test

Model 1: pressure ~ x + y + I(x^2) + I(y^2) + x:y

Model 2: pressure ~ x + y

	Res.Df	Df	F	Pr(>F)
1	79			
2	82	-3	1.6284	0.1895

2 Model selection with stepwise

Given estimates of covariance parameters $\hat{\theta}$ and keeping them fixed, the usual stepwise procedures (`step()`) for selecting covariates can be used.

`georob` implements selecting models based on AIC and BIC.

AIC and BIC

Both criteria balance **model fit** against **model complexity**:

$$\text{AIC} = -2 \log L + 2k$$

$$\text{BIC} = -2 \log L + k \log(n)$$

where L = likelihood, k = number of estimated parameters, and n = sample size.

- Better fit $\rightarrow$ lower criterion
- More parameters $\rightarrow$ higher penalty
- BIC penalizes complexity more strongly than AIC

Stepwise model selection

`step()` performs **stepwise model selection** by repeatedly adding or removing predictors according to AIC/BIC.

Direction	
forward	Start with a small model → add variables
backward	Start with the full model → remove variables
both	Add or remove variables at each step

→ `step()` is a heuristic search through the model space, not a test of whether individual predictors are “significant”.

Stepwise selection with AIC

(defaults to both directions)

```
1 step(r.georob.full)
```

Start: AIC=922.16

pressure ~ x + y + I(x^2) + I(y^2) + x:y

	Df	AIC	Converged
- I(x^2)	1	922.05	1
- I(y^2)	1	922.13	1
<none>		922.16	
- x:y	1	922.49	1

Step: AIC=922.05

pressure ~ x + y + I(y^2) + x:y

	Df	AIC	Converged
<none>		922.05	
+ I(x^2)	1	922.16	1
- I(y^2)	1	922.54	1
- x:y	1	924.61	1

Tuning constant: 1000

Fixed effects coefficients:

Stepwise selection with BIC

(defaults to both directions)

```
1 step(r.georob.full, k=log(nrow(d.w)))
```

Start: AIC=936.81

pressure ~ x + y + I(x^2) + I(y^2) + x:y

	Df	AIC	Converged
- I(x^2)	1	934.27	1
- I(y^2)	1	934.34	1
- x:y	1	934.70	1
<none>		936.81	

Step: AIC=934.27

pressure ~ x + y + I(y^2) + x:y

	Df	AIC	Converged
- I(y^2)	1	932.31	1
<none>		934.27	
- x:y	1	934.38	1
+ I(x^2)	1	936.81	1

Step: AIC=932.31

pressure ~ x + y + x:y



**Utrecht
University**

Sharing science,
shaping tomorrow