Space: the final frontier

FW 891

Click here to view presentation online

Christopher Cahill

30 October 2023

Quantitative Fisheries Center
MICHIGAN STATE UNIVERSITY

Purpose

- Background
- Recognizing dependency in simple models
- How can we deal with dependency problems
    - Introduction to spatial random effects

Why do we care about space

- Most of our problems in ecology and mgmt are inherently spatial

Special Feature

Space: The Final Frontier for Ecological Theory[1]

Most ecologists realize that the interplay of dispersal, disturbance, and spatial mosaics can profoundly alter the outcome of species interactions. Indeed, an appreciation of the spatial dimension and spatial heterogeneity is well established in natural history. Nonetheless, for many ecologists, "spatial complications" are used as a catch-all for explaining away surprising results, or as a condemnation of ecological theory that is accused of oversimplification because of its neglect of spatial variation. What is missing are serious experiments that explicitly test major hypotheses emerging from recent theoretical explorations of spatial effects. The purposes of this Special Feature are to alert ecologists to the general insights emerging from spatial models, and to show how these concepts apply to the natural world.

Tillman et al. 1994

Nearby things tend to be more alike

Photo credit: accuweather.com

Paraphrased from Fisher 1935 and Tobler 1970

However...

- It is really hard to estimate spatial (or spatial-temporal) dynamics in most statistical models

Cressie and Wikle 2011

Spatial dependency poses issues for traditional experimental design

Sir Ronald Fisher

Spatial dependency poses issues for traditional experimental design

- Statistical theory of experimental designs (Fisher 1935)

Paraphrased from Fisher 1935

Spatial dependency poses issues for traditional experimental design

- Statistical theory of experimental designs (Fisher 1935)
  - *After choosing the area we usually have no guidance beyond the widely verified fact that patches in close proximity are commonly more alike…than those which are further apart*

Paraphrased from Fisher 1935

Spatial dependency poses issues for traditional experimental design

- Statistical theory of experimental designs (Fisher 1935)
  - *After choosing the area we usually have no guidance beyond the widely verified fact that patches in close proximity are commonly more alike…than those which are further apart*
- Fisher avoided space with the introduction of randomization into the scientific method

Paraphrased from Fisher 1935

Spatial dependency poses issues for traditional experimental design

- Statistical theory of experimental designs (Fisher 1935)
  - *After choosing the area we usually have no guidance beyond the widely verified fact that patches in close proximity are commonly more alike…than those which are further apart*
- Fisher avoided space with the introduction of randomization into the scientific method
- We have hustled spatial dependence to the back door

Paraphrased from Fisher 1935

Spatial dependency poses issues for traditional experimental design

- Statistical theory of experimental designs (Fisher 1935)
  - *After choosing the area we usually have no guidance beyond the widely verified fact that patches in close proximity are commonly more alike…than those which are further apart*
- Fisher avoided space with the introduction of randomization into the scientific method
- We have hustled spatial dependence to the back door
  - Not rejecting this paradigm of experimental inference

Paraphrased from Fisher 1935

Observational data represent the wilder side of science and statistical inference

Observational data represent the wilder side of science and statistical inference

- Great potential for biased inference

sensu Walters 1986; Cressie and Wikle 2011

Observational data represent the wilder side of science and statistical inference

- Great potential for biased inference
  - Wrong operating model
  - Unmeasured yet influential covariates
  - Lack of replication or randomization

sensu Walters 1986; Cressie and Wikle 2011

Observational data represent the wilder side of science and statistical inference

- Great potential for biased inference
  - Wrong operating model
  - Unmeasured yet influential covariates
  - Lack of replication or randomization
- Causal inference more challenging than in controlled experiments

sensu Walters 1986; Cressie and Wikle 2011

Observational data represent the wilder side of science and statistical inference

- Great potential for biased inference
  - Wrong operating model
  - Unmeasured yet influential covariates
  - Lack of replication or randomization
- Causal inference more challenging than in controlled experiments
- Notions of space and time even more relevant?

sensu Walters 1986; Cressie and Wikle 2011

Observational data represent the wilder side of science and statistical inference

- Great potential for biased inference
  - Wrong operating model
  - Unmeasured yet influential covariates
  - Lack of replication or randomization
- Causal inference more challenging than in controlled experiments
- Notions of space and time even more relevant?
- This is most of our data!

sensu Walters 1986; Cressie and Wikle 2011

Observational data represent the wilder side of science and statistical inference

Hierarchical modeling as a powerful tool for ecological inference

- Unifying field of statistics
- Many technical benefits
- Helps deal with dependency
  - i.e., Hurlburt's pseudoreplication
- So we are going to use these to address the spatial dependency issue

Hurlburt 1984; Royle and Dorazio 2008

Recognizing dependency in statistical models

What do we look for when checking statistical models?

- Zuur and Ieno 2016 offered a 10 step approach

Protocol for presenting regression-type analyses

1. State appropriate questions
2. Visualize the experimental design
3. Conduct data exploration
4. Identify the dependency structure in the data
5. Present a statistical model
6. Fit said model
7. “validate” the model
8. Interpret and present output
9. Model visualizations
10. Simulate from the model

Zuur and Ieno 2016

The take home point from Zuur and Ieno

“*Before starting the analysis carefully consider which observations for your response variable can be dependent. If dependency is present, apply a statistical technique that is able to cope with it*”

Zuur et al. 2017

Thinking about dependency

- Suppose samples are collected by different laboratories
  - Perhaps due to differences in techniques or scientists themselves among laboratories, we likely have among-lab dependency

Zuur et al. 2017

Thinking about dependency

- Suppose related species are analyzed together in one analysis
  - Perhaps due to similarities between the species we have among-species dependency

Zuur et al. 2017

Thinking about dependency

- Suppose we take repeated measurements on individual rats
  - Perhaps due to individual differences among rats, the data scientists collect from one rat is more likely to be similar to other data from collected from that rat

Zuur et al. 2017

What dependency is not

- Two sites have high bird counts, and they weren't sampled by the same person or lab
- However, both sites have a low cover measurement, which is the cause of the high bird counts

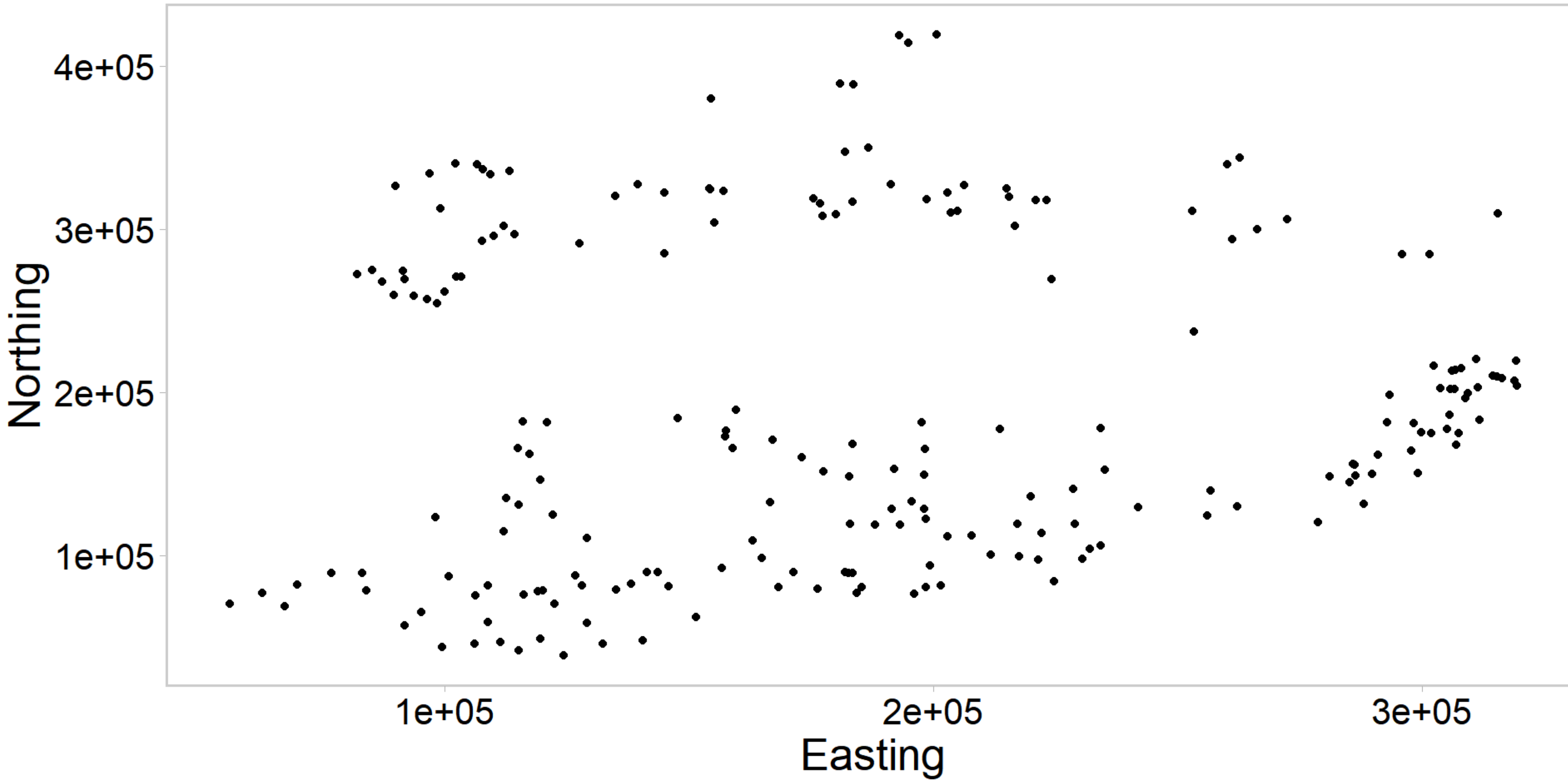
Zuur et al. 2017

Visualize experimental design: spatial dependency

```
library(tidyverse)
library(ggqfc)
data <- read.table(file = "data/ph_dat.txt", header = TRUE, dec = "
data %>%
  ggplot(aes(x = Easting, y = Northing)) +
  geom_point(shape = 16) + theme_qfc()
```

Zuur et al. 2017

Visualize experimental design: spatial dependency



Zuur et al. 2017

Running an analysis

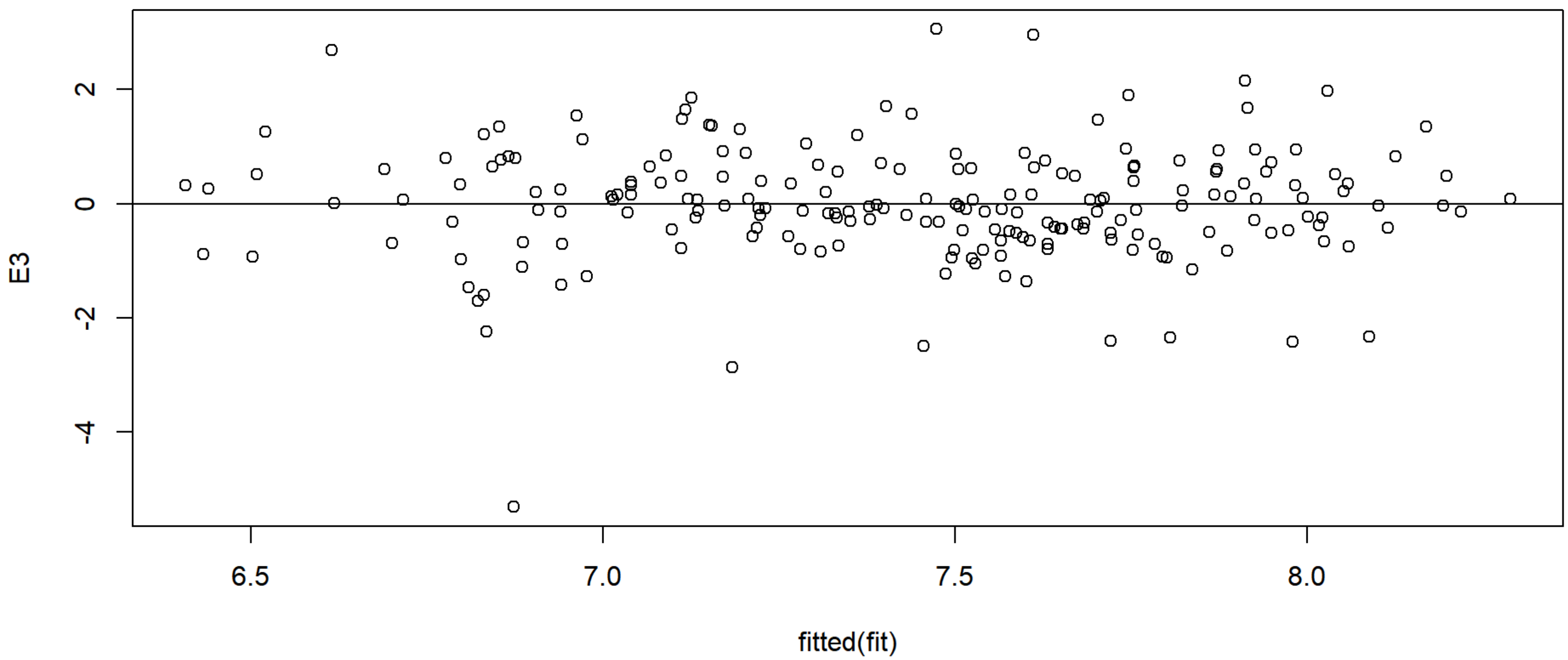
```r
data$fForested <- factor(data$Forested,
                         levels = c(1, 2),
                         labels = c("Yes", "No"))
data$LOGAltitude <- log10(data$Altitude)

# Model selected using some stepwise AIC approach
fit <- lm(pH ~ LOGAltitude + SDI + fForested + LOGAltitude:fForeste
          data = data)
E3 <- rstandard(fit) # could also use resid(), but this is better w
```

Zuur et al. 2017

Validating the model

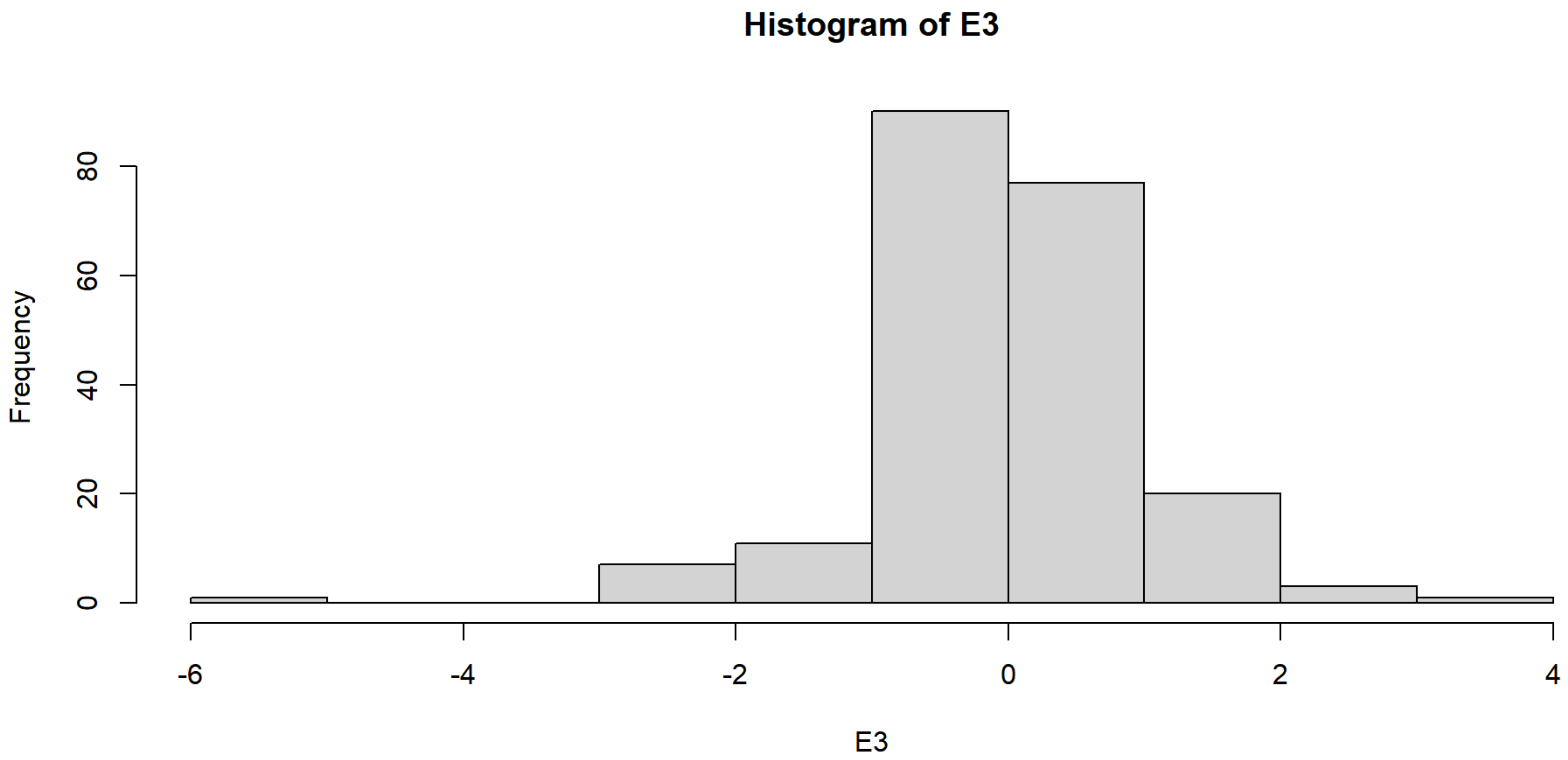
```r
plot(x = fitted(fit), y = E3)
abline(h = 0, v = 0)
```



Zuur et al. 2017

Validating the model

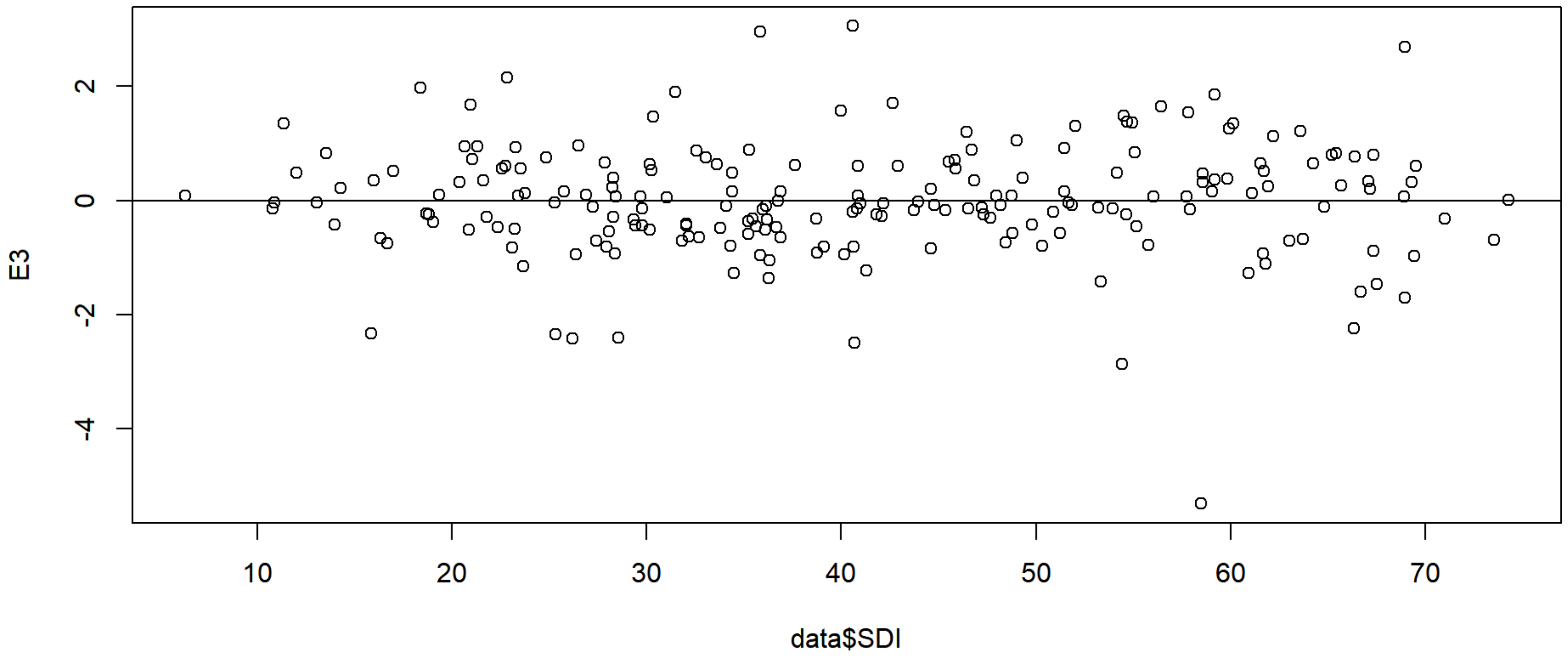
```r
hist(E3)
```



Zuur et al. 2017

Validating the model

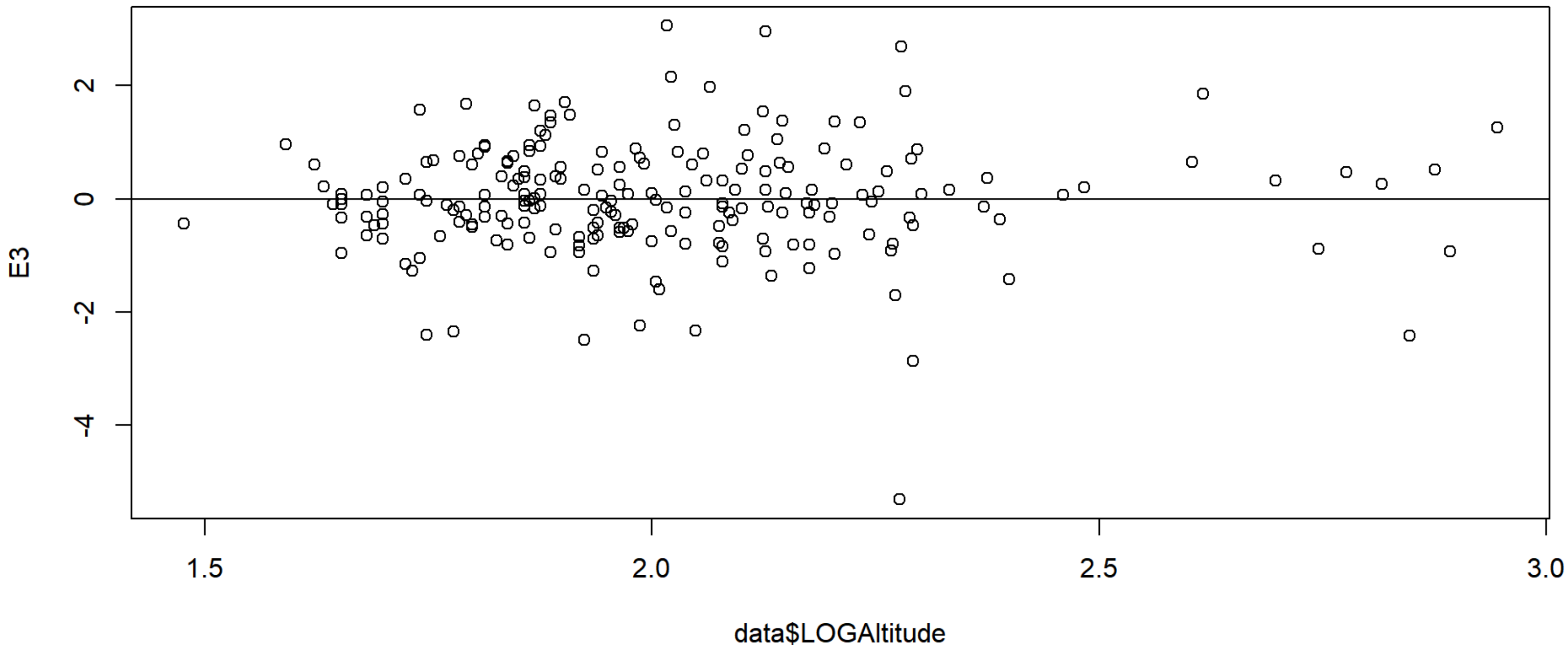
```r
# Independence due to model misfit
plot(x = data$SDI, y = E3) # residuals vs. predictors
abline(h = 0, v = 0)
```



Zuur et al. 2017

Validating the model

```r
# Independence due to model misfit
plot(x = data$LOGAltitude, y = E3) # residuals vs. predictors
abline(h = 0, v = 0)
```



Zuur et al. 2017

Can we find spatial patterns?

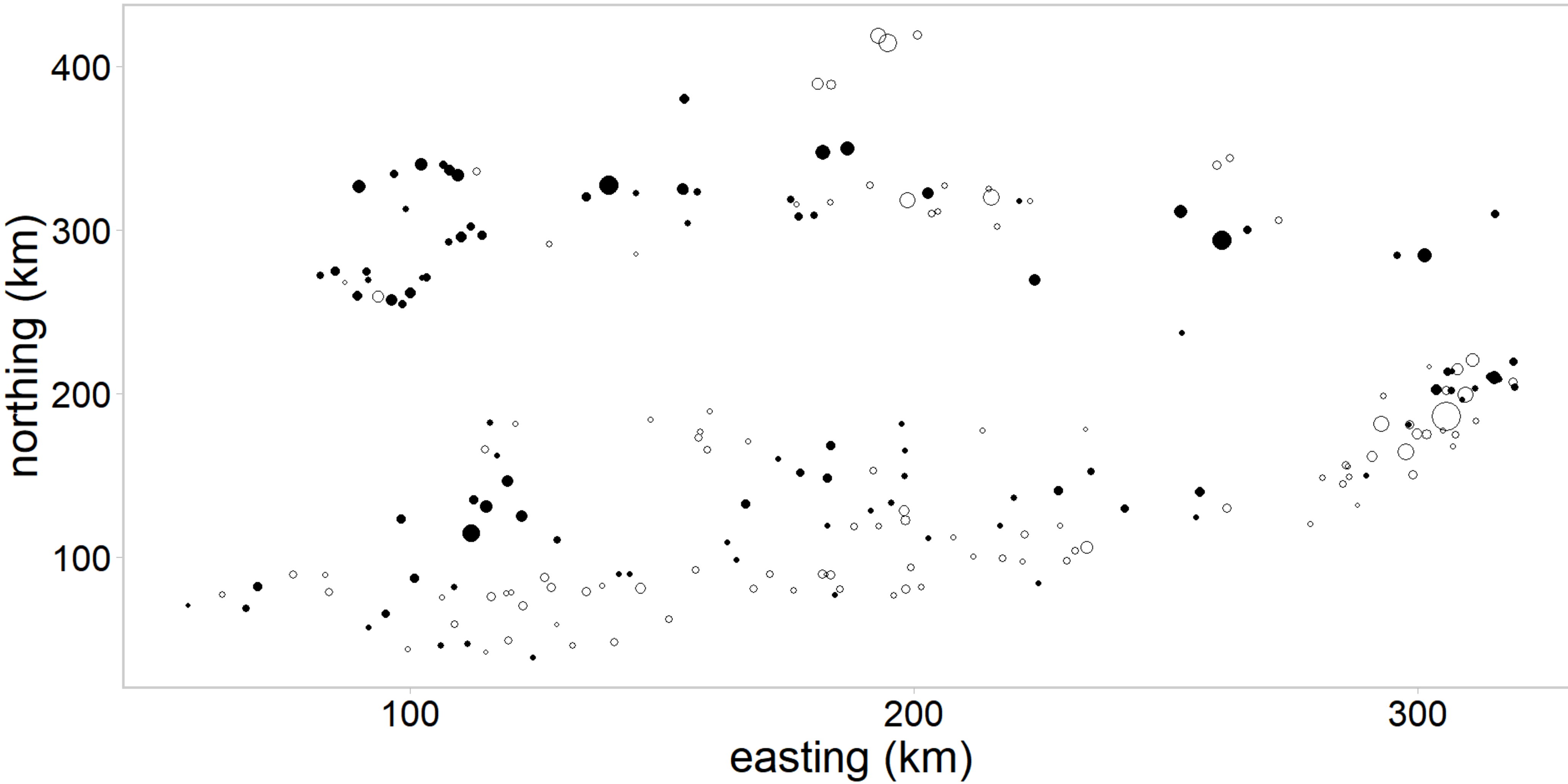
```
# Can we find spatial patterns in the residuals?
data$E3 <- resid(fit)
data$my_cex <- 3 * abs(data$E3) / max(data$E3) + 0.75
data$sign <- as.numeric(data$E3 >=0) + 1
data$my_pch <- c(1, 16)[data$sign]

# Convert Easting / Northing to km
data$easting_km <- data$Easting/1000
data$northing_km <- data$Northing/1000

data %>%
  ggplot(aes(x = easting_km, y = northing_km)) +
  geom_point(size = data$my_cex, shape = data$my_pch) +
  ylab("northing (km)") + xlab("easting (km)")
```

Zuur et al. 2017

Can we find spatial patterns?



Zuur et al. 2017

A point about the types of spatial data

- Point process vs. areal data

Zuur et al. 2017

Introduction to spatial random effects

Introduction to spatial random effects

- Spatial effects can be incorporated many ways
  - Predictors of model's mean response
  - Models of covariance (e.g., kriging)
    - splines
    - hierarchical intercepts indexed by region
    - Gaussian Processes, etc.
- Recent extensions include modeling spatial deviations as random effects using Gaussian Random Fields

Cressie and Wikle 2011; Anderson and Ward 2019

Gaussian Random Fields (GRFs)

- GRFs represent a two-dimensional version of Gaussian processes and define the expected value, variance, and covariance of random draws from a Multivariate Normal Distribution (MVN)
- Gaussian processes in this context are analogous to kriging
- Simply put: interpolation technique that allows us to model the covariance among sites according to some function

Cressie and Wikle 2011; Anderson and Ward 2019

Spatial Gaussian Processes

- We have a few options with the kernel function
  - Exponential, squared-exponential, matern, etc.
- We will start with an exponential kernel function

The exponential covariance function

$$\epsilon_s \sim \mathrm{MVN}(0, \Sigma)$$

where Σ is a covariance matrix with:

$$\Sigma_{i,j} = \sigma^2_{gp} \exp(-\mathrm{distances}_{i,j}/\theta_{gp}), \text{ if } i \neq j$$

$$\Sigma_{i,j} = \sigma^2_{gp}, \text{ if } i = j$$

Cressie and Wikle 2011; Anderson and Ward 2019

The exponential covariance function

- is a spatially explicit random effect
- controls how quickly the correlation decays between locations
- is the marginal variability of the spatial function at all locations
- is a matrix describing Euclidean distance between sample locations

Cressie and Wikle 2011; Anderson and Ward 2019

An spatially explicit model

- pH measured at a n sites
- Want to estimate average pH and its uncertainty

- note the s

To the R and Stan code to implement this model

Recap

- Talked about them importance of space in applied ecological settings
- Discussed a bunch of ways to recognize dependency in applied datasets
- Showed one way (GRFs and GPs) to incorporate spatial effects into hierarchical models

Recap

References

- Anderson and Ward. 2019. Black swans in space: modeling spatiotemporal processes with extremes. Ecology.
- Cressie and Wikle 2011. Statistics for spatio-temporal data.
- Fischer 1935. The design of experiments.
- Tillman et al. 1994. Space: The final frontier for ecological theory. Ecology.
- Royle and Dorazio 2008. Hierarchical modeling and inference in ecology
- Tobler, W. R. 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. Economic Geography. 46: 234–240. doi:10.2307/143141. ISSN 0013-0095. JSTOR 143141. Retrieved 2023-10-02.
- Walters 1986. Adaptive Management.
- Zuur and Ieno 2016. A protocol for conducting and presenting results of regression-type analyses. Methods in Ecology and Evolution. 7: 636–664.
- Zuur, A.F., Ieno, E.N., and Saveliev, A.A. 2017. Spatial, temporal and spatial–temporal ecological data analysis with R-INLA. Highland Statistics Ltd.